

Caustics for Inner and Outer Billiards

Eugene Gutkin^{1, *}, Anatole Katok^{2, **}

¹ Mathematics Department, USC, Los Angeles, CA 90089-1113, USA

² Mathematics Department, Pennsylvania State University, University Park, PA 16802, USA

Received: 1 May 1994/in revised form: 15 August 1994

Abstract: With a plane closed convex curve, T , we associate two area preserving twist maps: the (classical) inner billiard in T and the outer billiard in the exterior of T . The invariant circles of these twist maps correspond to certain plane curves: the inner and the outer caustics of T . We investigate how the shape of T determines the possible location of caustics, establish the existence of open regions which are free of caustics, and estimate from below the size of these regions in terms of the geometry of T .

Introduction

A closed convex curve, $T \subset \mathbf{R}^2$, in the Euclidean plane defines two natural dynamical systems: the classical billiard ball map, ϕ_T (inside T), and the “outer billiard,” ψ_T , in $\mathbf{R}^2 \setminus \text{int}(T)$. The phase space, $\Phi = \Phi_T$, of the inner billiard is the space of directed straight lines (rays) intersecting T . The phase space, Ψ_T , of the outer billiard map is the set of points in the exterior of T . With a natural choice of coordinates in the phase space, both billiards are area preserving twist maps.

Invariant circles for these twist maps correspond to certain geometric objects. In the case of ϕ_T these are the caustics of T (famous in the geometric optics). A caustic, $\gamma \subset \text{int}(T)$, corresponding to an arbitrary invariant circle Γ may have a complicated structure. For instance, the jumps of the tangent direction for Γ correspond to the discontinuities of γ . In particular, these caustics are not convex. There are other types of nonconvex caustics. For instance, the equator of Φ_T , for a table T of constant width, is an invariant circle. The corresponding caustic is not convex (unless T is circular). In Sect. 1 where we deal with the inner billiard map, we restrict our setting to the convex caustics. The present techniques do not work without this assumption.

If a caustic γ is convex, then it is simply the envelope of the corresponding family, Γ , of rays. For sufficiently smooth tables T a necessary and sufficient

* Partially supported by NSF. email: (egutkin@math.usc.edu)

** Partially supported by NSF Grant DMS 9017995.

condition for the existence of such caustics is the positivity of the curvature of T . This is a consequence of two well known results. One is a theorem of Lazutkin [15] (a version of KAM) that ensures the existence of smooth convex caustics near T . The other one is a theorem of Mather [16] that excludes all invariant circles if the curvature vanishes at least at one point of T .

The situation for outer billiards is somewhat different. In this case the invariant circles directly correspond to the simple closed Lipschitz curves, $\Gamma \subset \mathbf{R}^2 \setminus \text{int}(T)$, encircling the table. In general, Γ is neither smooth nor convex. However, for a sufficiently smooth table T , with a strictly positive curvature, an application of KAM ensures the existence of smooth convex invariant circles Γ arbitrarily close to T and arbitrarily far away from it [18, 6]. To emphasize the analogy between the inner and the outer billiard maps, we call the invariant circles for ψ_T the outer caustics.

The objective of this work is twofold. First, for the inner billiard map we establish a quantitative version of Mather's theorem [16], compatible with Lazutkin's results [15]. Namely, we prove that if the minimal curvature of the billiard table T (appropriately normalized) is sufficiently small, then the convex caustics are located only near the boundary of the table. In particular, we estimate from below the area inside T which is "free of convex caustics," see Theorems 1.3 and 1.4. To interpret our results, let us consider the following "mental experiment." There is a one-parameter family, T_t , of convex billiard tables, sufficiently smooth, and with a strictly positive curvature. We can think that our billiard table, T , is slowly changing with time. Suppose that under the deformation the minimal curvature of T is gradually decreasing to zero, while the global shape of the table remains essentially unchanged. Then, by the results of Sect. 1, the convex caustics are gradually pushed out to the boundary (Theorems 1.3, 1.4).

Second, we find a counterpart of Mather's theorem for outer billiards (Theorem 2.1), and establish a quantitative version of it. We prove that if the minimal radius of curvature of the outer billiard table T is sufficiently small (with respect to the size of T), then the outer caustics are confined to the union of two disjoint annular regions in Ψ . One of the regions, Ψ_1 , is a topological annulus surrounding T . The other annular region, Ψ_2 , is infinite, and is bounded away from Ψ_1 . We estimate from below the area of the annular region, Ψ_f , between Ψ_1 and Ψ_2 , which is "free of outer caustics" (Theorem 2.2 and Corollary 2.8). This is a counterpart, for the outer billiard map, of Theorem 1.4.

Imagine a deformation of the outer billiard table, analogous to the "mental experiment" described above. Under the deformation the minimal radius of curvature of T goes to zero, while the general shape is unchanged. Then, by our results, the region Ψ_1 shrinks to the table, while Ψ_2 stays bounded away from T . Thus, when $\rho_{\min}(T) = 0$, the outer caustics are bounded away from T (Theorem 2.1.). The annulus Ψ_f , free of outer caustics, surrounds the outer table. This is a counterpart of Mather's theorem [16] for outer billiards. The result is sharp, in the sense that there are examples of outer billiard tables T , $\rho_{\min}(T) = 0$, with outer caustics (see examples in Sect. 3).

The annulus Ψ_f is contained in a Birkhoff region of instability [3, 13], \mathcal{A} , $\partial\mathcal{A} = \Gamma_1 \cup \Gamma_2$. The complement of \mathcal{A} in Ψ is the union of two invariant annuli, \mathcal{A}_1 and \mathcal{A}_2 . The annulus \mathcal{A}_1 is between T and Γ_1 . It contains the outer caustics of the first kind. The annulus \mathcal{A}_2 is the exterior of Γ_2 . It contains the outer caustics of the second kind. Thus Γ_1 and Γ_2 are the "last outer caustics of their kind." A region of instability contains orbits with complicated asymptotic behavior

(Mather [17], Le Calvez [5]). In particular, in $\text{int}(\mathcal{A})$ there are orbits homoclinic to Γ_1, Γ_2 , as time goes to $\pm\infty$ respectively (Theorem 2.3).

The results on outer caustics are presented in Sect. 2. Note that in Sect. 2 we don't assume that the caustics are convex. In fact, we don't impose any additional restrictions on outer caustics. This is why the proofs in Sect. 2, especially in Subjects. 2.6, 2.7, are longer and more technical than the proofs in Sect. 1.

We point out that the radius of curvature for an outer billiard table plays in the preceding discussion a role analogous to that of the curvature for an inner billiard table. This is a manifestation of a partial duality between the two dynamical systems. The origin of this phenomenon lies in the projective duality between points and lines in the Euclidean plane. This correspondence shows in the similarity of the basic constructions for the two billiard maps in question. For example, the string construction for the inner billiard is analogous to the area construction for the outer billiard (see Sects. 1.3 and 2.3). The mirror equation for the inner billiard has a counterpart for the outer billiard (see Eqs. (1.5) and (2.11)).

The last section of the paper is devoted to an informal discussion, comments and examples. As we have noted above, the techniques of Sect. 1 are suitable for obtaining a priori estimates on convex (inner) caustics only. In order to extend our results to the general case, we need to improve our understanding of the geometry of nonconvex billiard caustics. There is virtually nothing on this subject in the literature. In Sect. 3 we discuss a few examples of billiard tables with special nonconvex inner caustics. In one of the examples we construct a one-parameter family of billiard tables having the astroid as a caustic. Also in Sect. 3 we give examples of outer billiard tables with a vanishing radius of curvature (or a corner), and having outer caustics. This illustrates the sharpness of Theorem 2.1.

We have reported and discussed the results of this paper at various meetings and workshops in 1992–93 [9]. When the present work was finished, Phil Boyland brought to our attention that he has obtained a version of Theorem 2.1 in 1988, and has announced this and other results in his lectures on outer billiards in 1989. He also informed us that he was writing up his results. The manuscript has since appeared [23]. (Added in proof).

1. Inner Caustics

1.1. Preliminaries and Notation. A billiard table, $[T]$, is a domain in \mathbf{R}^2 bounded by a convex closed C^1 curve, T , which is C^2 except for finitely many (singular) points, $s_i, 1 \leq i \leq N$. We denote by $\Phi = \Phi_T$ the set of unit tangent vectors in \mathbf{R}^2 with footpoints on T , directed inward, and endow Φ with the natural topology. Fixing an initial point in T (and orienting T positively), we parametrize T by the arclength, $0 \leq s \leq |T|$. Let $v \in \Phi$ be a vector with a footpoint s , and let θ be the angle between v and the positive tangent direction to T at s (Fig. 1.1). The parametrization $v = v(s, \theta)$ identifies Φ with the cylinder $\{(s, \theta) : 0 \leq s < |T|, 0 \leq \theta \leq \pi\}$. The curvature, $\kappa(s)$ is defined for $s \neq s_i, 1 \leq i \leq N$, and we don't impose any restrictions on $\kappa(s)$ as $s \rightarrow s_i$.

The billiard flow in $|T|$ is defined as the free motion of a point mass with the unit speed inside $|T|$, with the elastic reflections from the boundary. The set Φ is a natural cross-section of the billiard flow, and the first return transformation, $\phi : \Phi \rightarrow \Phi, \phi(s, \theta) = (s_1, \theta_1)$, is called *the billiard ball map* (Fig. 1.1).

Let $l(s, s')$ be the Euclidean distance between the corresponding points on T . The map ϕ is differentiable at (s, θ) if T is C^2 at s and s_1 . The differential is

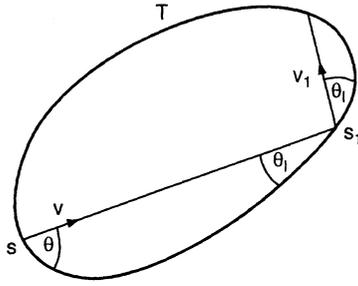


Fig. 1.1. The billiard ball map.

given by

$$\begin{aligned} \frac{\partial s_1}{\partial s} &= \frac{\kappa(s)l(s, s_1) - \sin \theta}{\sin \theta_1}, & \frac{\partial s_1}{\partial \theta} &= \frac{l(s, s_1)}{\sin \theta_1}, \\ \frac{\partial \theta_1}{\partial s} &= \frac{\kappa(s)\kappa(s_1)l(s, s_1) - \kappa(s)\sin \theta_1 - \kappa(s_1)\sin \theta}{\sin \theta_1}, \\ \frac{\partial \theta_1}{\partial \theta} &= \frac{\kappa(s_1)l(s, s_1) - \sin \theta_1}{\sin \theta_1}. \end{aligned} \tag{1.1}$$

This formula is a special case of Eq. (4.10) in [14]. The proof is elementary, and we leave it to the reader.

Denote by $s' - s$ the arclength of T from s to s' (in the positive direction on T). Set

$$h(s, s') = (s' - s) - l(s, s'), \quad t = 1 - \cos \theta. \tag{1.2}$$

Then $ds dt = ds_1 dt_1$, and

$$dh(s, s_1) = t_1 ds_1 - t ds, \tag{1.3}$$

i.e., ϕ is an area preserving *twist map*, and h is a *generating function* (see, e.g., [5, 6, 11, 13]).

We identify Φ with the set of *rays*, (i.e., oriented straight lines) intersecting T , and let X be the *space of rays* in \mathbf{R}^2 . Endowed with the natural topology, X is diffeomorphic to the infinite cylinder, $X \simeq S^1 \times \mathbf{R}$, and $\Phi \subset X$ is a bounded domain. Denote by TX the tangent bundle of X , and let $T_P X$ be the *projectivized tangent bundle*. Topologically, $T_P X \simeq S^1 \times S^1 \times \mathbf{R}$. We will make use of a “physical” interpretation of $T_P X$, as follows. A *point at infinity* of \mathbf{R}^2 is a family of rays with the same direction. The set of points at infinity is the *absolute* (a topological circle), and let $\hat{\mathbf{R}}^2$ be the union of \mathbf{R}^2 and the absolute (with the natural topology). Note that $\hat{\mathbf{R}}^2$ is homeomorphic to the disc, not the projective plane. Any ray, $l \in X$, defines a unique point at infinity, $\infty(l)$, and we complete l with $\infty(l)$. Any pair (l, A) , $A \in l$, (including the pair $(l, \infty(l))$) defines an *infinitesimal light beam*, centered at l , with the *focusing point* A . If A is a finite point, the light beam consists of the rays l' , “very close to” l , and $l' \cap l = A$. If $A = \infty(l)$, the light beam consists of rays l' , very close to and parallel to l .

With any infinitesimal light beam (l, A) we associate a differentiable curve, $\gamma = \gamma_{(l, A)} = \{l(t) : -\varepsilon < t < \varepsilon\}$, on X , where $l(0) = l, l(t) \cap l = A$. The curve γ is

not uniquely defined, but its tangent line at $t = 0$ does not depend on the choices involved, and defines a point, $\tau = \tau(l, A) \in T_P X$. This is a one-to-one correspondence between $T_P X$ and the set of infinitesimal light beams. Restricting this isomorphism to $T_P \Phi \subset T_P X$, we obtain a description of points of $T_P \Phi$ as the infinitesimal light beams $(l, A), l \cap T \neq \emptyset$. We set $f(l, A) = A \in \hat{\mathbf{R}}^2$, the *focusing point* of the beam. The mapping $f : T_P X \rightarrow \hat{\mathbf{R}}^2 \simeq D$, as well as the restriction, $f : T_P \Phi \rightarrow \hat{\mathbf{R}}^2$, is onto (but neither map is a fibration).

We say that a curve $\Gamma \subset \Phi$ is an *invariant circle* if Γ is isotopic to a boundary component of Φ and $\phi(\Gamma) = \Gamma$. Both boundary components of Φ are trivial invariant circles, and we disregard them in what follows. If T is C^2 then ϕ is a C^1 area preserving twist map, and, by Birkhoff's theorem ([3], see also [11, 13]), any invariant circle is the graph of a Lipschitz function, i.e., $\Gamma = \{(s, \theta(s)) : 0 \leq s \leq |T|\}$. We identify Γ with a family of rays, $\Gamma = \{l_\Gamma(s) : 0 \leq s \leq |T|\}$, in \mathbf{R}^2 , and let $\langle \Gamma \rangle \subset \mathbf{R}^2$ be the intersection of the closed (left) half-planes of $\{l_\Gamma(s) : 0 \leq s \leq |T|\}$. By definition, $\langle \Gamma \rangle$ is a convex closed set, and $\langle \Gamma \rangle \subset [T]$.

Definition 1.1. *Let the notation be as above, and let $\Gamma \subset \Phi_T$ be an invariant circle. We say that Γ is a convex invariant circle if every ray $l_\Gamma(s), 0 \leq s \leq |T|$, is a supporting line of $\langle \Gamma \rangle$. The boundary, $\gamma = \partial(\langle \Gamma \rangle) \subset [T]$, is the caustic corresponding to the invariant circle Γ .*

Remarks. Our setting is closely related to the billiard flow in $[T]$, which is a Hamiltonian system [1, 13]. Invariant circles, $\Gamma \subset \Phi$, correspond to the invariant tori in the phase space of the flow. A caustic, $\gamma = \partial(\langle \Gamma \rangle)$, is the “trace” of Γ in the configuration space. The extremal points, $A \in \gamma$, are the focusing points of the infinitesimal light beams, $(l, A) \in T_P \Gamma$.

In this work we restrict our attention to convex invariant circles. Their caustics are, by definition, convex. This property fails for general invariant circles (see Sect. 3 below).

1.2. Mirror Equation of the Geometric Optics. Let T be an oriented plane curve without selfintersections (not necessarily closed). Assume that T is C^2 , with the curvature $\kappa(M), M \in T$, and denote by Φ the space of rays intersecting T . Let $M \in T, A \in \mathbf{R}^2$. Set $d(A) = \pm|AM|$, with the plus sign if A belongs to the left half-plane determined by the positive tangent to T at M , and the negative sign otherwise. Let $(l, A) \in T_P \Phi, M = l \cap T$. The signed distance $d(A) = \pm|AM|$ is a function on $T_P \Phi$. Denote by $d\phi$ the differential of ϕ , and let $d_P \phi$ be the projectivization, $d_P \phi : T_P \Phi \rightarrow T_P \Phi$. We will use a well known formula from the geometric optics (the “mirror equation,” see, e.g., [21], Lemma 1).

Proposition 1.1. *In the notation above, let $(l, A) \in T_P \Phi$, and let $(m, B) = d_P \phi(l, A), M = l \cap m$. Let θ be the angle between T and l , and set $a = d(A), b = d(B)$ (Fig. 1.2). Then*

$$\frac{1}{a} + \frac{1}{b} = 2 \frac{\kappa(M)}{\sin \theta}. \tag{1.4}$$

If $A \neq B$ are points in \mathbf{R}^2 , we use notation AB for the interval directed from A to B , and, if there is no danger of confusion, for the corresponding oriented line. We denote by $|AB|$ the length of the interval AB .

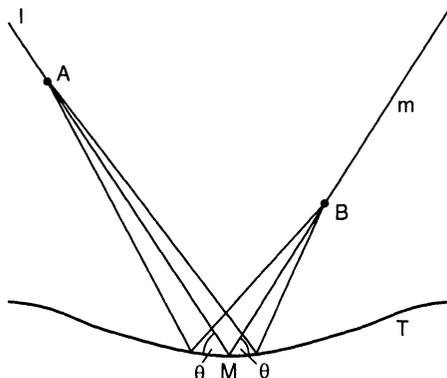


Fig. 1.2. Mirror equation of the geometric optics.

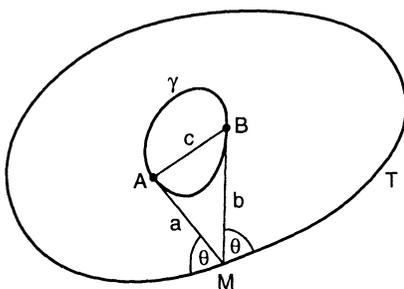


Fig. 1.3. Notation for Corollary 1.1, Lemma 1.1, and Lemma 1.2.

Corollary 1.1. *Let $[T]$ be a billiard table with a convex caustic γ . Let $A, B \in \gamma$ and $M \in T$ be as shown in Fig. 1.3. Let $\theta = \theta(\gamma, M)$ be the corresponding angle. If M is a C^2 point of T , then*

$$\frac{1}{|AM|} + \frac{1}{|BM|} = 2 \frac{\kappa(M)}{\sin \theta}. \tag{1.5}$$

Proof. This is a special case of Eq. (1.4) with $a = |AM|$, $b = |BM|$. \square

Equation (1.5) immediately implies the following.

Theorem 1.1. *A convex, C^2 billiard table with a point of zero curvature has no convex caustics.*

Since convex caustics correspond to convex invariant circles, Theorem 1.1 is a weakened version of a result of J. Mather’s [16]. Mather’s proof of the nonexistence of invariant circles [16] uses the differential of the billiard map (see Eq. (1.1)), and the orientation preserving homeomorphism $\phi_\Gamma : T \rightarrow T$, associated with an invariant circle Γ . We can deduce Mather’s theorem from Eq. (1.4). Since the proof would require a discussion of general (nonconvex) caustics, we do not give it here. We refer the reader to Wojtkowski [21], page 397, for a geometric optics approach to Mather’s theorem. In the rest of this section we will obtain a quantitative improvement of Theorem 1.1.

1.3. Lazutkin Parameter of a Caustic. The String Construction. Let $\gamma \subset [T]$ be a convex caustic. For any $M \in T$ let MA, MB be the supporting rays to γ , and let $\theta = \theta(M) \leq \pi/2$ be the corresponding angle (see Fig. 1.3). The curve γ is rectifiable, and we denote by $|\gamma_A^B|$ the arclength of γ from A to B (in the positive direction on γ).

Lemma 1.1. *The quantity*

$$L(\gamma, M) = |AM| + |BM| - |\gamma_A^B| \tag{1.6}$$

does not depend on M .

Proof. Denote by $d\sigma$ the arclength element on γ . A straightforward computation gives

$$d(|MB|) = -\cos \theta ds(M) + d\sigma(B), \quad d(|MA|) = \cos \theta ds(M) - d\sigma(A).$$

This, and the equation

$$d(|\gamma_A^B|) = d\sigma(B) - d\sigma(A)$$

imply the claim. \square

By the preceding lemma, the quantity $L(\gamma, M) = L(\gamma)$, defined by Eq. (1.6), depends only on the (convex) caustic γ . Lazutkin [15] used $L(\gamma)$ to study the caustics near T (these are necessarily convex). We call $L(\gamma)$ the Lazutkin parameter of the caustic γ .

We set $S(\gamma) = L(\gamma) + |\gamma|$. The quantity $S(\gamma)$ has a ‘‘physical’’ interpretation. Imagine that the curve γ is made from a thin smooth wire, and that we have a closed string of length $S(\gamma)$, made from unstretchable material. We wrap the string around the wire, and pull it tight at a point, M . When we drag the string around γ , preserving the conditions above, the point M traces the table T . This is the *string construction*, and $S(\gamma)$ is the *string parameter* of the caustic γ . For a given convex closed curve, γ , and any $S > |\gamma|$, the string construction produces a unique table $T = T(S)$ such that γ is a caustic in T , and $S(\gamma) = S$. This is well known in the literature, see, e.g., [18, 12], or [19], where it is attributed to Minasian. In fact, the string construction goes back to the antiquity, at least in the special case when γ is an interval, $[AB]$. It produces the family of (confocal) ellipses with foci A and B . When $B = A$, we obtain the circles centered at A .

1.4. Estimates from Above on the Lazutkin Parameter. We fix a billiard table $[T]$, and use the preceding notation.

Lemma 1.2. *Let γ be a convex caustic, and let $L = L(\gamma)$ be the Lazutkin parameter. Let $M \in T$, and let $\theta(M)$ be the corresponding angle (Fig. 1.3). Then*

$$L < 2 \frac{\sin^3 \theta(M)}{\kappa(M)}. \tag{1.7}$$

Proof. Denote by a, b, c the sides of the triangle ABM . Since $|\gamma_A^B| > c$, by Lemma 1.1, $L < a + b - c$. By cosine theorem

$$c = (a + b) \left[1 - \frac{4abs \sin^2 \theta}{(a + b)^2} \right]^{1/2}.$$

Using the inequality (valid for $0 < x < 1$) $1 - x < (1 - x)^{1/2}$ with $x = 4ab \sin^2 \theta / (a + b)^2$, we obtain

$$L < \frac{4ab \sin^2 \theta}{a + b} = 4 \left(\frac{1}{a} + \frac{1}{b} \right)^{-1} \sin^2 \theta. \quad (1.8)$$

The claim follows, by Corollary 1.1. \square

For any $A \in \mathbf{R}^2$, let $\delta(A, T)$ be the distance from A to T . For any continuous curve $\gamma \subset [T]$, let $\delta_{\min}(\gamma) = \min_{A \in \gamma} \delta(A, T)$ and $\delta_{\max}(\gamma) = \max_{A \in \gamma} \delta(A, T)$ be the distance and the deviation of γ from T . Set

$$\kappa = \inf_{M \in T} \kappa(M), \quad K = \sup_{M \in T} \kappa(M).$$

Note that if T is C^2 , then $\kappa = \kappa_{\min}(T)$, $K = \kappa_{\max}(T)$. Denote by $d = d(T)$ the diameter of T . For $M \in T$ denote by $w(M)$ the width of the strip formed by two parallel supporting lines of T , one of which contains M . Set $w = \min_M w(M)$, the width of T .

Proposition 1.2. *For any convex caustic γ its Lazutkin parameter, $L = L(\gamma)$, satisfies*

$$L \leq 2d^3 \kappa^2, \quad L \leq 2/K. \quad (1.9)$$

Proof. Rewrite Eq. (1.5) as

$$\sin \theta(M) = 2 \left(\frac{1}{a} + \frac{1}{b} \right)^{-1} \kappa(M).$$

From this, Eq. (1.8), and the well known inequality

$$2 \left(\frac{1}{a} + \frac{1}{b} \right)^{-1} \leq \frac{a + b}{2},$$

we obtain

$$L < 2 \left(\frac{a + b}{2} \right)^3 \kappa^2(M).$$

Since $a + b \leq 2d$, we have $L < 2d^3 \kappa^2(M)$ which implies the first inequality in Eq. (1.9). The second inequality is immediate from Eq. (1.7). \square

The following is a variation of Theorem 1.1.

Theorem 1.2. *Let $[T]$ be a convex billiard table satisfying the standing assumptions: T is C^1 and piecewise C^2 . If the curvature $\kappa(M)$ on T is not bounded away either from zero or from infinity, then T has no convex caustics. Equivalently, the billiard ball map, $\phi: \Phi \rightarrow \Phi$, has no convex invariant circles.*

Proof. The equations $\kappa = 0$ or $K = \infty$ imply, by Proposition 1.2, that $L(\gamma) \leq 0$ for any convex caustic γ . But the Lazutkin parameter of a caustic is positive. \square

1.5. Estimates from below on the Lazutkin Parameter

Proposition 1.3. *Let $\gamma \subset [T]$ be a convex caustic with a Lazutkin parameter $L(\gamma)$. Then*

$$\delta_{\max}(\gamma) < \sqrt{dL(\gamma)}^{1/2}.$$

Proof. Set $L = L(\gamma)$. Let $M \in T$, and let $A, B \in \gamma$ be the corresponding points of support (see Fig. 1.4.). The line A_1B_1 is tangent to γ at $C \in \gamma$, and is parallel to the tangent line to T at M . The height $|MH|$ of the isocetes triangle A_1B_1M is equal to the width of the strip formed by the two tangent lines. Since γ is convex, $|\gamma_A^B| < |AA_1| + |A_1B_1| + |BB_1|$, hence

$$L > |A_1M| + |B_1M| - |A_1B_1|.$$

Expressing the sides of the triangle A_1MB_1 via $|HM|$ and the angle $A_1MB_1 = \pi - 2\theta$, we obtain

$$|HM| < \frac{L \sin \theta}{2(1 - \cos \theta)} = \frac{1}{2}L \cot \frac{\theta}{2}.$$

Since $|HM| = |A_1M| \sin \theta < d \sin \theta$, we have

$$|HM| < 2d \frac{\tan \theta/2}{1 + \tan^2 \theta/2}.$$

The graphs $\{(t, L/2t) : 0 < t < 1\}$ and $\{(t, 2dt/(1 + t^2)) : 0 < t < 1\}$ intersect at $t_0 = (L/(4d - L))^{1/2}$. Note that $t_0 < 1$, since $L(\gamma) < 2d$ for any γ . Substituting this into the inequalities above, with $t = \tan \theta/2$, we obtain

$$|HM| < \frac{1}{2}L^{1/2}(4d - L)^{1/2} < \sqrt{dL}.$$

Since $\delta_{\max}(\gamma) = \max_M |HM|$, this implies the claim. \square

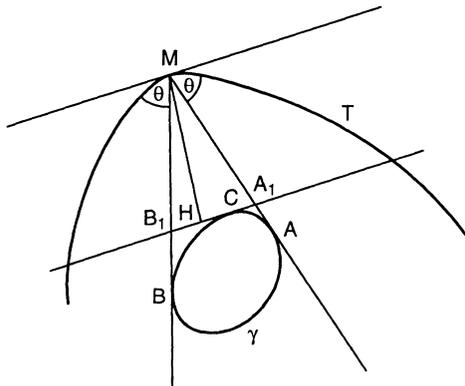


Fig. 1.4. Notation for the proof of Proposition 1.3.

Corollary 1.2. *The deviation of any convex caustic from T satisfies*

$$\delta_{\max}(\gamma) < \min\{\sqrt{2}\kappa d^2, (2d/K)^{1/2}\}.$$

Proof. Immediate from Propositions 1.2, 1.3. \square

1.6. *Regions free of Convex Caustics*

Definition 1.2. *We say that a region $X \subset [T]$ is free of convex caustic, if $\gamma \cap \text{int}(X) = \emptyset$ for any convex caustic γ .*

We denote by $|X|$ the Lebesgue area of a region, thus $|[T]|$ is the area of the billiard table. Let $r_{in} = r_{in}(T)$ be the radius of a maximal inscribed disc (inradius).

Proposition 1.4. *If a billiard table satisfies the condition*

$$r_{in} \geq \sqrt{2}\kappa d^2, \tag{1.10}$$

then $[T]$ contains a disc, X , free of convex caustics. Its radius, r_f , satisfies $r_f > r_{in} - \sqrt{2}\kappa d^2$.

Proof. Let O_{in} be the center of a maximal inscribed disc, and let \tilde{X} be the closed disc of radius $r_{in} - \sqrt{2}\kappa d^2$ centered at O_{in} . Assume that convex caustics can pass arbitrarily close to \tilde{X} . Since the set of convex caustics is closed, there is one, γ , passing through \tilde{X} . By triangle inequality and Corollary 1.2, $d(O_{in}, T) < r_{in}(T)$, which is a contradiction. Thus, \tilde{X} is contained in a bigger disc, X , free of convex caustics. \square

Corollary 1.3. *If a billiard table satisfies*

$$\kappa K \leq (\sqrt{2}d^2)^{-1}, \tag{1.11}$$

then there is a disc, $X \subset [T]$, free of convex caustics, and $r_f > K^{-1} - \sqrt{2}\kappa d^2$.

Proof. Follows from Proposition 1.4, since $r_{in} \geq K^{-1}$. \square

Corollary 1.4. *Suppose the billiard table satisfies*

$$\kappa \leq \frac{1}{3\sqrt{2}} \frac{w}{d^2}. \tag{1.12}$$

Then there is a disc, $X \subset [T]$, free of convex caustics, and

$$r_f > w/3 - \sqrt{2}\kappa d^2. \tag{1.13}$$

Proof. By a theorem of Blaschke ([4], see also [22]), $r_{in} \geq w/3$. Hence, Eq.(1.10) is satisfied, and the claim follows from Proposition 1.4. \square

Theorem 1.3. *Let T be a billiard table satisfying condition (1.11). Then $[T]$ contains a convex region, X , free of convex caustics, and*

$$\frac{|X|}{|[T]|} \geq 1 - \frac{\sqrt{2}\kappa|T|d^2}{|[T]|}. \tag{1.14}$$

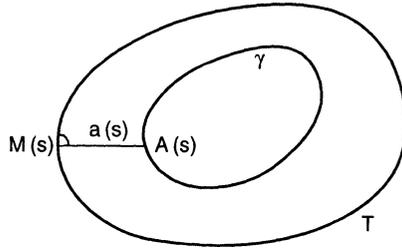


Fig. 1.5. Notation for the proof of Theorem 1.3.

Proof. In view of Theorem 1.2, we can assume that $\kappa > 0$. For any $a \geq 0$ set $X_a = \{A \in [T] : \delta(A, T) \geq a\}$. Then X_a is a convex region (empty for large a).

For a closed convex curve, $\gamma \subset [T]$, and $s \in T$, let $R(s)$ be the halfline perpendicular to T at s , and let $a(s)$ be the distance from s to the nearest point of $R(s) \cap \gamma$ (Fig. 1.5). If $a(s) \leq \rho_{\min}(T)$ for all s , then the area $|[T] \setminus [\gamma]|$, between γ and T , is given by

$$|[T] \setminus [\gamma]| = \int_0^{|T|} a(s) \left(1 - \frac{1}{2} a(s) \kappa(s)\right) ds. \tag{1.15}$$

Equation (1.15) is proved by approximating T near s by the circle of curvature.

Let $a \leq 1/K$, and set $\gamma = \partial X_a$. Equation (1.15) implies

$$|X_a| > |[T]| - a|T|. \tag{1.16}$$

Set $a = \sqrt{2}\kappa d^2$. By Corollary 1.2, Eqs.(1.11) and (1.16), the region $X = X_a$ is free of convex caustics and satisfies Eq.(1.14). \square

The estimate on the area free of convex caustics becomes especially attractive for C^2 billiard tables.

Theorem 1.4. *Let T be a C^2 billiard table with*

$$\sqrt{2}d^2 \kappa_{\min}(T) \kappa_{\max}(T) \leq 1. \tag{1.17}$$

Then $[T]$ contains a convex region, X , free of convex caustics, and the area of X is bounded below by Eq. (1.14).

Proof. Since T is C^2 , we have $\kappa = \kappa_{\min}(T)$, $K = \kappa_{\max}(T)$, and the preceding theorem applies. \square

2. Outer Caustics

2.1. Outer Billiard Map. Let $\rho(\alpha), 0 \leq \alpha \leq 2\pi$, be a nonnegative periodic function, satisfying the following conditions. First, $\rho(\cdot)$ is continuous, except for a finite number of points, and if ρ is discontinuous at α , then the one-sided limits $\rho_{\pm}(\alpha) = \lim \rho(\alpha')$, as $\alpha' \rightarrow \alpha \pm 0$, exist (we allow $\rho_{\pm}(\alpha) = \infty$). Second, $\int_0^{2\pi} \rho(\alpha) d\alpha < \infty$. Third

$$\int_0^{2\pi} \rho(\alpha) \cos \alpha d\alpha = \int_0^{2\pi} \rho(\alpha) \sin \alpha d\alpha = 0. \tag{2.1}$$

Finally, we assume that the set $\{\rho(\alpha) = 0\}$ is the union of a finite number of closed intervals and points.

The curve $T = \{(x(\alpha), y(\alpha)) : 0 \leq \alpha \leq 2\pi\}$, with the radius of curvature $\rho(\alpha)$, is given by the differential equations

$$x'(\alpha) = \rho(\alpha) \cos \alpha, \quad y'(\alpha) = \rho(\alpha) \sin \alpha. \tag{2.2}$$

By our assumptions, T is a *closed strictly convex curve*, of length $|T| = \int_0^{2\pi} \rho(\alpha) d\alpha$.

An interval $[\alpha_-, \alpha_+]$ of zeros of ρ produces a *corner point*, $M \in T$, with the two directed *tangent lines*, $l_{\pm}(M)$, with the angles α_{\pm} , and a cone of *supporting rays* $l(\alpha) : \alpha_- \leq \alpha \leq \alpha_+$. Thus, T has a finite number of corner points, and T is C^1 everywhere else. Let α be an isolated zero of ρ , or let ρ be discontinuous at α , with $\rho_-(\alpha) = 0$ or $\rho_+(\alpha) = 0$. Then T is C^1 but not C^2 at the corresponding point, M , and the *curvature*, $\kappa(M') = 1/\rho(M')$, goes to infinity, as M' approaches M from at least one side. Thus, T is piecewise C^1 and piecewise C^2 , with a finite number of points of *infinite curvature*. The one-sided curvatures, $\kappa_{\pm}(M) \leq \infty$, are defined everywhere. To simplify the exposition, we assume in what follows that the function $\rho(M)$ is bounded.

Let T be as above, let $[T] \subset \mathbf{R}^2$ be the corresponding convex domain, and set $\Psi = \mathbf{R}^2 \setminus [T]$. We define a mapping $\psi : \Psi \rightarrow \Psi$. For $A \in \Psi$, let l_+, l_- be the two *supporting lines* for T through A . Let $M_{\pm} = l_{\pm} \cap T$ be the corresponding points of support. Let $A_1 \in l_+(A_{-1} \in l_-)$ be the unique point such that $|AM_+| = |A_1M_+|$ ($|AM_-| = |A_{-1}M_-|$) (Fig. 2.1). We set $\psi(A) = A_1$. By our assumptions on T , $\psi = \psi_T$ is a homeomorphism, and $\psi^{-1}(A) = A_{-1}$.

Definition 2.1. *The homeomorphism $\psi : \Psi \rightarrow \Psi$ is the outer billiard map (corresponding to the outer billiard table T), and Ψ is the phase space of the outer billiard.*

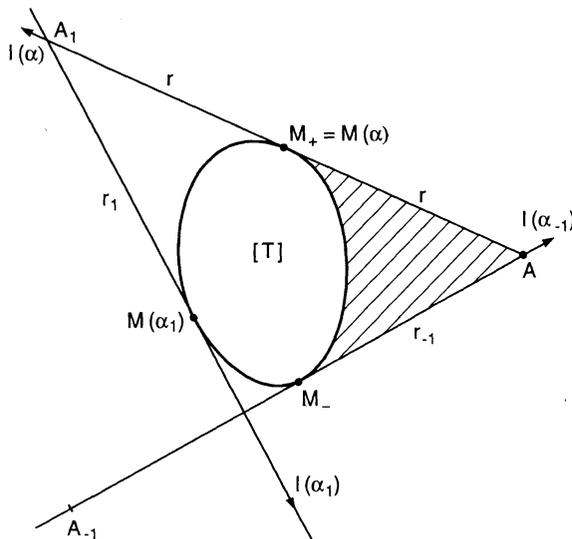


Fig. 2.1. The outer billiard map.

Remark. If T is not strictly convex, the outer billiard map may have discontinuities. Outer billiards about convex polygons were investigated in [10].

We choose a reference direction in \mathbf{R}^2 , and assign to any ray l the angle, $\alpha(l)$, it makes with the reference direction. The function $l \rightarrow \alpha(l)$ takes its values in the *circle of directions*, $S^1 = \{0 \leq \alpha \leq 2\pi\}$. For every direction α , there is a unique supporting ray l for T (on the right of T), with $\alpha(l) = \alpha$. This is a homeomorphism between S^1 and the set of supporting rays for T . Let $M(\alpha) \in T$ be the supporting point of $l(\alpha)$ (Fig. 2.1). This establishes a parametrization, $\alpha \rightarrow M(\alpha)$, of T by S^1 . The mapping $\alpha \rightarrow M(\alpha)$ is one-to-one if (and only if) T has no corner points, i.e., T is C^1 . We will also use the parametrization of T by the arclength. Since

$$d\alpha/ds = \kappa(M), \quad ds/d\alpha = \rho(M), \tag{2.3}$$

the two parametrizations are equivalent if $0 < \kappa(\cdot) < \infty$.

For $A \in \Psi$, let $l(\alpha)$ be the corresponding supporting ray, let $M = M(\alpha) \in T$ be the point of support, and set $r = |MA|$ (Fig. 2.1). This defines a homeomorphism $A \rightarrow (\alpha, r)$ of Ψ onto the semiinfinite cylinder, $S^1 \times \mathbf{R}_+$. We set $\psi(\alpha, r) = (\alpha_1, r_1)$.

Lemma 2.1. *Let the notation be as above (Fig. 2.1). Suppose that M, M_1 are regular points, and set $\rho = \rho(M), \rho_1 = \rho(M_1)$. Then ψ is C^1 at (α, r) :*

$$\begin{aligned} \frac{\partial \alpha_1}{\partial \alpha} &= \frac{\rho \sin(\alpha_1 - \alpha) - r \cos(\alpha_1 - \alpha)}{r_1}, \\ \frac{\partial \alpha_1}{\partial r} &= \frac{\sin(\alpha_1 - \alpha)}{r_1}, \\ \frac{\partial r_1}{\partial \alpha} &= \frac{(\rho \rho_1 - r r_1) \sin(\alpha_1 - \alpha) - (\rho r_1 + \rho_1 r) \cos(\alpha_1 - \alpha)}{r_1}, \\ \frac{\partial r_1}{\partial r} &= \frac{\rho_1 \sin(\alpha_1 - \alpha) - r_1 \cos(\alpha_1 - \alpha)}{r_1}. \end{aligned} \tag{2.4}$$

Proof. Straightforward computation. \square

Equations (2.4) extend to the case when M, M_1 are not necessarily regular. Denote by ρ^\pm, ρ_1^\pm the one sided radii of curvature at M, M_1 respectively. The tangent space, \mathbf{R}_A^2 , to Ψ at A splits into a finite number of cones (at most four). The differential, $d\psi|_A$, exists in the interior of each cone, and is given by Eqs.(2.4), with $\rho = \rho^\pm, \rho_1 = \rho_1^\pm$, depending on the cone. Across a ray in \mathbf{R}_A^2 , separating a pair of these cones, $d\psi|_A$ may have a jump discontinuity. We leave the details to the reader. The following is immediate from Eqs.(2.4) and these remarks.

Corollary 2.1. *The differential of the outer billiard map is bounded everywhere, and can have only jump discontinuities. If the radius of curvature is a continuous function, then the map is C^1 .*

Remarks. 1) By preceding corollary, T may not be C^2 (not even C^1), and have a C^1 outer billiard map. 2) Comparing Eqs.(2.4) with Eq.(1.1) we see that the radius of curvature plays for outer billiards the role of curvature for the inner billiards.

For any pair $\alpha - \pi < \alpha_{-1} < \alpha$, of directions, we denote by $H(\alpha_{-1}, \alpha)$ the area of the “triangle” formed by the supporting lines $l(\alpha), l(\alpha_{-1})$ and T (Fig. 2.1).

Proposition 2.1. *The function $H(\alpha_{-1}, \alpha)$ is C^1 , and*

$$dH = \frac{1}{2}r^2 d\alpha - \frac{1}{2}r_{-1}^2 d\alpha_{-1}. \tag{2.5}$$

For any cartesian coordinates ξ, η in \mathbf{R}^2 , we have $d\xi d\eta = r dr d\alpha$. The outer billiard map preserves the form $r dr d\alpha$.

Proof. Equation (2.5) and the identity $d\xi d\eta = r dr d\alpha$ are proved by a straightforward calculation. By Eqs. (2.4), $r_1 dr_1 d\alpha_1 = r dr d\alpha$. \square

Corollary 2.2. *Set $x = \alpha, y = \frac{1}{2}r^2$. The map $\psi(x, y) = (x_1, y_1)$ is an area preserving monotone twist map, and $H(x_{-1}, x)$ is a generating function.*

Proof. Immediate from Eqs. (2.4) and Proposition 2.1. \square

Remark. Denote by \mathcal{A} the group of affine orientation preserving automorphisms of \mathbf{R}^2 , and let $\mathcal{E} \subset \mathcal{A}$ be the subgroup of conformal transformations. If $T, T_1 = g(T), g \in \mathcal{A}$, are two outer billiard tables, then g conjugates the outer billiard maps: $g\psi = \psi_1 g$. Thus the outer billiard is *equivariant* with respect to \mathcal{A} , $\dim \mathcal{A} = 6$, as opposed to the inner billiard which is equivariant only with respect to \mathcal{E} , $\dim \mathcal{E} = 4$.

We fix an outer billiard table T , and study the *invariant circles* of the corresponding map $\psi : \Psi \rightarrow \Psi$. To emphasize the analogy with the inner billiard, we call them the *outer caustics* of T . Thus, $\Gamma \subset \Psi$ is an outer caustic, if Γ is isotopic to T , and $\psi(\Gamma) = \Gamma$. To simplify the statements, we assume $\Gamma \neq T$, the trivial outer caustic.

With any Lipschitz curve, $\Gamma \subset \mathbf{R}^2$, we associate its set of *tangent cones*, $\{C_P \subset \mathbf{R}_P^2 : P \in \Gamma\}$. The lines $l \subset C_P$ are the (generalized) *tangent lines* of Γ . For almost all $P \in \Gamma$ the cone C_P consists of a unique tangent line, $C_P = l_P$.

Lemma 2.2. *Let T be an outer billiard table such that the outer billiard map is C^1 . Let $\Gamma \subset \Psi$ be an outer caustic. Then the tangent lines of Γ don't intersect T .*

Proof. Since Γ is an invariant circle for a C^1 twist map, by Birkhoff's theorem ([3], see [11] or [13] for a modern proof), Γ is a Lipschitz graph, $r = r(\alpha), 0 \leq \alpha \leq 2\pi$.

Assume that there is a tangent line to Γ intersecting T . By a straightforward topological argument, the set of lines which are tangent to both Γ and T is nonempty. Let $l = l(\alpha_0)$ be such, let $P_0 \in \Gamma, M_0 \in T$ be the points of tangency, and set $|P_0 M_0| = r_0 > 0$. Then the ratio $|r(\alpha) - r_0|/|\alpha - \alpha_0|$ is not bounded as $\alpha \rightarrow \alpha_0$, which contradicts Birkhoff's theorem. \square

2.2. Counterpart of the Mirror Equation. We use the notation above. Let $A \in \Psi, B = \psi(A)$, and let M be the point of support (Fig. 2.2). Set $r = |AM| = |BM|$. Choose Euclidean coordinates (x, y) in \mathbf{R}^2 so that AB is the positive x -axis. We compute $d\psi : \mathbf{R}_A^2 \rightarrow \mathbf{R}_B^2$ in the basis $\partial/\partial x, \partial/\partial y$.

Lemma 2.3. *Let T be a regular outer table. i) Let $M \in T$ be a point of continuity for the radius of curvature of T . Then*

$$d\psi = \begin{pmatrix} -1 & -2\rho(M)/r \\ 0 & -1 \end{pmatrix}. \tag{2.6}$$

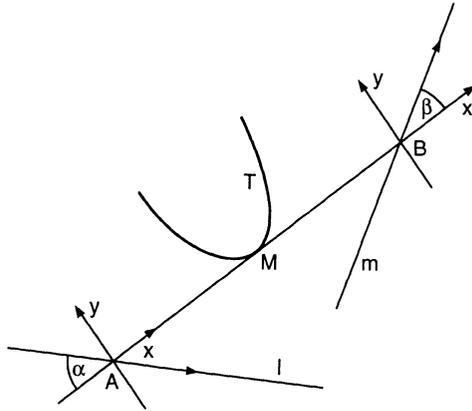


Fig. 2.2. Notation for Lemma 2.3, Proposition 2.2.

ii) Let $M \in T$ be a C^1 point, which is a jump point for the radius of curvature, and let $\rho_-(M) \neq \rho_+(M)$ be the one-sided radii of curvature. Then $d\psi$ has a jump across the x -axis in \mathbf{R}_A^2 . Denote by $d\psi_-(d\psi_+)$ the differential restricted to the halfplane $\mathbf{R}_A^{2-} = \{v = (v_x, v_y) : v_y > 0\}$ ($\mathbf{R}_A^{2+} = \{v = (v_x, v_y) : v_y < 0\}$). Then

$$d\psi_- = \begin{pmatrix} -1 & -2\rho_-(M)/r \\ 0 & -1 \end{pmatrix}, \quad d\psi_+ = \begin{pmatrix} -1 & -2\rho_+(M)/r \\ 0 & -1 \end{pmatrix}. \tag{2.7}$$

iii) Let M be a corner point of T , with the tangent lines $l_{\pm}(M)$, and let $V_M \subset \Psi$ be the cone formed by their tails, $R_{\pm}(M)$. Let $\rho_{\pm}(M)$ be the one-sided radii of curvature at M . For points $A \in \text{int}(V_M)$,

$$d\psi = \begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix}. \tag{2.8}$$

For points $A \in R_-(M)$, the differential $d\psi|_A$ can have a jump across the x -axis. The two matrices $d\psi_{\pm}$, defined as in ii), are given by

$$d\psi_- = \begin{pmatrix} -1 & -2\rho_-(M)/r \\ 0 & -1 \end{pmatrix}, \quad d\psi_+ = \begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix}. \tag{2.9}$$

Analogously, there can be a jump in $d\psi|_A$ for $A \in R_+(M)$, and $d\psi_{\pm}$ are given by

$$d\psi_- = \begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix}, \quad d\psi_+ = \begin{pmatrix} -1 & -2\rho_+(M)/r \\ 0 & -1 \end{pmatrix}. \tag{2.10}$$

Proof. Straightforward calculation. \square

Corollary 2.3. *The outer billiard map about a table T is C^1 if and only if the radius of curvature is a continuous function.*

Proof. Immediate from the preceding lemma and Corollary 2.1. \square

In what follows we assume that T is an outer table with a continuous radius of curvature.

Proposition 2.2. *Let $\psi : \Psi \rightarrow \Psi$ be the outer billiard map about T . Let $A, B = \psi(A) \in \Psi$, and let $M \in T$ be the point of support (Fig. 2.2). Set $r = |AM| = |BM|$. Let l be a line in \mathbf{R}_A^2 , and let $m = d\psi(l) \subset \mathbf{R}_B^2$. Denote by $\alpha(\beta)$ the signed angle between MA (MB) and $l(m)$. If M is not a corner point of T , then*

$$\cot \alpha + \cot \beta = 2 \frac{\rho(M)}{r}. \tag{2.11}$$

If M is a corner point, then $\cot \alpha + \cot \beta = 0$.

Proof. We consider the three cases of Lemma 2.3. In the first case we choose a vector $u = (x, y) \in l \subset \mathbf{R}_A^2$, and set $v = d\psi(u) \in m$. By Lemma 2.3i, $v = (-x - 2\rho(M)y/r, -y)$. Then $\cot \alpha = -x/y$, $\cot \beta = [x + 2\rho(M)y/r]/y$, and Eq. (2.11) follows. In the remaining two cases, the lines l, m are parallel, hence $\alpha + \beta = 0$. \square

Equation (2.11) is a counterpart of the “mirror equation” of the geometric optics, Eq. (1.4).

2.3. Area Parameter and Area Construction. Let T be an outer table, and let $\Gamma \subset \Psi$ be a Jordan curve such that any supporting ray, $l(\alpha)$, of T intersects Γ at two points. Denote them by $A(\alpha), B(\alpha)$, and let $S(\alpha)$ be the area enclosed between Γ and the line $l(\alpha)$ (see Fig. 2.3). Set $a(\alpha) = |MA(\alpha)|$, $b(\alpha) = |MB(\alpha)|$. The following is proved by a direct computation.

Lemma 2.4. *Let the notation be as above, and assume that the functions $a(\alpha), b(\alpha)$ are continuous. Then $S(\alpha)$ is a C^1 function, and*

$$dS(\alpha) = \frac{1}{2}(b(\alpha)^2 - a(\alpha)^2) d\alpha. \tag{2.12}$$

Corollary 2.4. *Let $\Gamma \subset \Psi$ be an outer caustic for a table T . The area enclosed between Γ and a supporting line, $l(\alpha)$, does not depend on α .*

Proof. By Birkhoff’s theorem, Γ is the graph of a Lipschitz function, $r(\alpha)$. Since $a(\alpha) = b(\alpha) = r(\alpha)$, by Eq. (2.12), $S(\alpha) = \text{const}$. \square

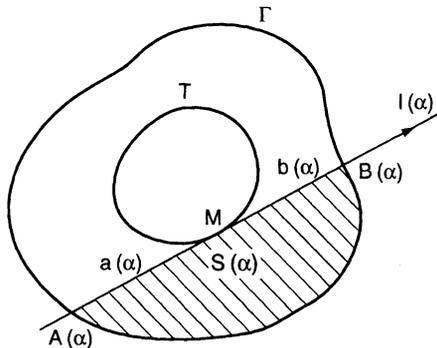


Fig. 2.3. Area parameter of an outer caustic.

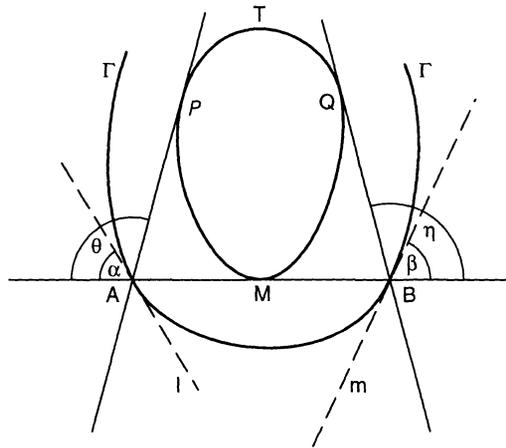


Fig. 2.4. Notation for Lemma 2.5.

Definition 2.2. Let T be an outer table, and let Γ be an outer caustic for T . The area S cut off of Γ by any supporting line of T depends only on Γ . We say that $S = S(\Gamma)$ is the area parameter of Γ .

Let $\Gamma \subset \mathbf{R}^2$ be a Jordan curve, and $S > 0$. Suppose that for any direction α there is a unique ray, $l(\alpha)$, such that the area enclosed between $l(\alpha)$ and Γ is equal to S . Assume that $l(\alpha)$ intersects Γ in two points, $A(\alpha)$ and $B(\alpha)$ (Fig. 2.3). Denote by $T = T(\Gamma, S)$ the locus of the midpoints, $M(\alpha)$, of the intervals $[A(\alpha), B(\alpha)]$, $0 \leq \alpha \leq 2\pi$.

Proposition 2.3. If T is convex, then T is an outer billiard table, Γ is an outer caustic for T , and $S = S(\Gamma)$.

We will not need this result, and we leave its proof to the reader. We say that $T = T(\Gamma, S)$ is obtained by the *area construction*. The area construction is a counterpart of the string construction, and the area parameter is an analog of the Lazutkin parameter for a caustic (see Sect. 1.3). For a given Jordan curve Γ , and $S > 0$, the curve $T = T(\Gamma, S)$ may have singularities, even if Γ is convex (see [7] and Sect. 3 below). We will use Corollary 2.4 in a slightly different form.

Corollary 2.5. If Γ is an outer caustic for T , and S is the area parameter of Γ , then $T = T(\Gamma, S)$.

2.4. Counterpart of Mather's Theorem for Outer Billiards. We fix a regular outer table T . By an *annulus* we will mean a region, $\mathcal{A} \subset \Psi$, homeomorphic to the standard annulus and homotopic to T . The boundary, $\partial\mathcal{A}$, of a compact annulus consists of two disjoint Jordan curves.

Definition 2.3. We say that a region $X \subset \Psi$ is free of outer caustics if $\text{int}(X) \cap \Gamma = \emptyset$ for any outer caustic Γ .

Lemma 2.5. Let Γ be an outer caustic, $A \in \Gamma$, $B = \psi(A)$, $M = AB \cap T$ (see Fig. 2.4). Let AP (BQ) be the other supporting line from A (B). Let θ (η) be

the angle between AB and AP (BQ), and set $r = |AM| = |BM|$. Then

$$\cot \theta + \cot \eta < 2\rho(M)/r. \tag{2.13}$$

Proof. Assume first that A is a C^1 point of Γ , and let l, m be the tangent lines to Γ at A, B respectively. Let α, β be the angles between AB and l, m (Fig. 2.4).

By Lemma 2.2, the lines l, m don't intersect T . Hence, $\alpha < \theta, \beta < \eta$, and Eq. (2.11) implies the claim. Since Γ is Lipschitz, C^1 points are dense, and, by continuity, Eq. (2.13) holds everywhere. \square

Inequality (2.13) will play a crucial role in obtaining a priori estimates for outer caustics. Before proceeding with this, we establish an analog of Mather's theorem [16] for outer billiards. We say that a region, $X \subset \Psi$, is *invariant* if $\psi(X) = X$. Recall that a *Birkhoff region of instability* (for a general twist map) is an invariant region, \mathcal{A} , which is an annulus whose boundary consists of invariant circles, and these are the only invariant circles in \mathcal{A} . We are interested in the Birkhoff regions of instability for the outer billiard map. If \mathcal{A} is compact, then $\partial\mathcal{A} = \Gamma_0 \cup \Gamma_1$, where $\Gamma_0 = T$ is possible. In the noncompact case, $\partial\mathcal{A} = \Gamma_0$.

Theorem 2.1. *Suppose that the radius of curvature of an outer table T has a zero. Then there is a Birkhoff region of instability, $\mathcal{A} \subset \Psi$, such that $\Gamma_0 = T$.*

Proof. By assumption, T has either a corner point or a C^1 point, M , with $\rho(M) = 0$. Let λ be a supporting line to T at M , and let A, B be the intersection points of λ and an outer caustic Γ (Fig. 2.5). By Proposition 2.2, the tangent lines l, m to Γ at A, B are parallel, and the distance between them satisfies $p > w = w(T)$, the *width* of T . Hence $|AM| = |BM| > w/2$.

By Birkhoff's theorem, Γ is the graph of a function, $r_\Gamma(\alpha)$. We can assume without loss of generality that $|AM| = r_\Gamma(0)$. Thus $r_\Gamma(0) > w/2$ for any outer caustic Γ . Therefore there exists $a > 0$ such that $\min_\alpha r_\Gamma(\alpha) \geq a$ for all outer caustics Γ . Otherwise, there would be an outer caustic, $\Gamma \neq T$, intersecting T , which is clearly impossible (see [12] for a related argument).

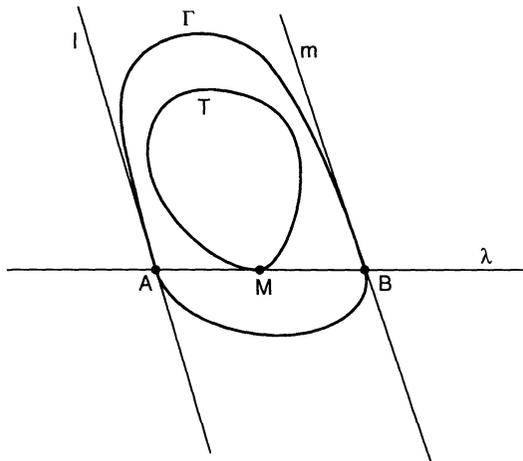


Fig. 2.5. Notation for the proof of Theorem 2.1.

We have shown that the annulus $\mathcal{A} = \{(\alpha, r) : r < a\}$ is free of outer caustics. By a well known argument (see [11, 16] or [13]), there is a Birkhoff region of instability, \mathcal{A} , containing \mathcal{A} . \square

2.5. “Icecream cone” Construction. We fix an outer billiard table T , and denote by ρ_{\min}, ρ_{\max} the minimum and the maximum of the radius of curvature. Let $l = l(\alpha)$ be a supporting ray for T , and let $M \in T$ be the point of support. For $r > 0$ let $A, B \in l$ be the two points with $|AM| = |BM| = r$. Denote by $\theta = \theta(r)$ ($\eta = \eta(r)$) the angle between l and the other supporting line, $AP(BQ)$, from $A(B)$ to T (Fig. 2.6). Both $\theta(r), \eta(r)$ are positive monotonically increasing functions of r , and $\theta(0) = \eta(0) = 0$. Let $\bar{r} = \bar{r}(\alpha)$ be such that $\theta(\bar{r}) + \eta(\bar{r}) = \pi$, and set $\theta(\bar{r}) = \bar{\theta}, \eta(\bar{r}) = \bar{\eta}, A(\bar{r}) = \bar{A}, B(\bar{r}) = \bar{B}, P(\bar{r}) = \bar{P}, Q(\bar{r}) = \bar{Q}$. For any r in the interval $(0, \bar{r})$, the supporting lines PA and QB intersect at C , forming the triangle ABC . The figure formed by the intervals CP, CQ , and the region $[T]$, resembles an icecream cone, hence the name.

Lemma 2.6. *Let the notation be as above, and let $d = d(T)$ be the diameter of T . i) For $0 \leq \alpha \leq 2\pi$ and $0 < r < \bar{r}(\alpha)$, we have*

$$2(\bar{r} - r)/d < \cot \theta + \cot \eta. \tag{2.14}$$

ii) *Let Γ be an outer caustic intersecting l at a distance $r(\alpha) < \bar{r}(\alpha)$ from M . Then*

$$r^2(\alpha) - \bar{r}(\alpha)r(\alpha) + \rho(\alpha)d > 0. \tag{2.15}$$

Proof. i) We suppress α from notation. Let \bar{P}_1, \bar{Q}_1 be the projections of \bar{P}, \bar{Q} to l (Fig. 2.6). If X, Y, Z are points in \mathbf{R}^2 , we denote by $\angle XYZ$ the angle formed by XY and YZ . As r varies from 0 to \bar{r} , the points P and Q monotonically move on T from M to \bar{P} and \bar{Q} respectively. By convexity of T , $\theta \leq \angle \bar{A}A\bar{P}, \eta \leq \angle \bar{B}B\bar{Q}$. Hence

$$\cot \theta + \cot \eta \geq \cot(\angle \bar{A}A\bar{P}) + \cot(\angle \bar{B}B\bar{Q}).$$

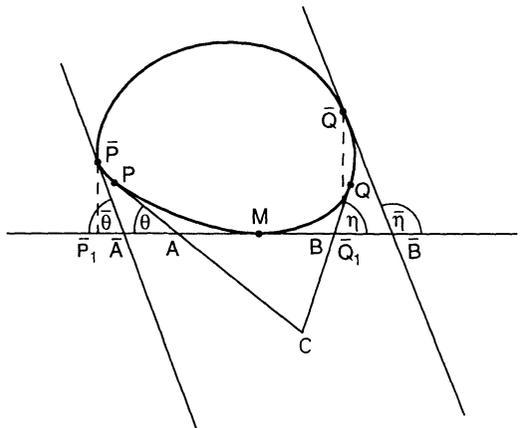


Fig. 2.6. “Icecream cone” construction.

From Fig. 2.6 we have

$$\begin{aligned}\cot(\bar{A}A\bar{P}) &= \frac{|\bar{P}_1\bar{A}| + |\bar{A}A|}{|\bar{P}\bar{P}_1|} = \cot\bar{\theta} + (\bar{r} - r)/|\bar{P}\bar{P}_1|, \\ \cot(\bar{B}B\bar{Q}) &= \frac{|\bar{B}\bar{B}| - |\bar{B}\bar{Q}_1|}{|\bar{Q}\bar{Q}_1|} = -\cot(\pi - \bar{\eta}) + (\bar{r} - r)/|\bar{Q}\bar{Q}_1|.\end{aligned}$$

From $\bar{\theta} + \bar{\eta} = \pi$, we obtain

$$\cot\theta + \cot\eta \geq (\bar{r} - r)(|\bar{P}\bar{P}_1|^{-1} + |\bar{Q}\bar{Q}_1|^{-1}).$$

Since $|\bar{P}\bar{P}_1|, |\bar{Q}\bar{Q}_1| < d$, this implies Eq. (2.14). ii) From Eqs. (2.13–2.14) we have

$$2(\bar{r} - r)/d < \cot\theta + \cot\eta < 2\rho(\alpha)/r, \quad (2.16)$$

which immediately implies the claim. \square

Remark. Equation (2.15) yields another proof of Theorem 2.1.

2.6. Bounds on the Area Parameter. We assume the preceding notation, and denote by $\Gamma = \{r(\alpha) : \alpha \in S^1\}$ an arbitrary outer caustic. For $A \in \Psi$, set $\delta(A) = d(A, T)$, the distance from A to T . Set

$$\begin{aligned}r_{\min}(\Gamma) &= \min_{\alpha} r(\alpha), & r_{\max}(\Gamma) &= \max_{\alpha} r(\alpha); \\ \delta_{\min}(\Gamma) &= \min_{A \in \Gamma} \delta(A), & \delta_{\max}(\Gamma) &= \max_{A \in \Gamma} \delta(A).\end{aligned}$$

Thus, $\delta_{\min}(\Gamma)$ is the *distance* and $\delta_{\max}(\Gamma)$ is the *deviation* of Γ from T . If we think of Ψ as a vertical cylinder, then r is the height of the point (α, r) . Thus, $r_{\min}(\Gamma)$ and $r_{\max}(\Gamma)$ are the smallest and the biggest heights for points in Γ .

Lemma 2.7. i) *Let $R > 0$ be such that the osculating disc of radius R at any $M \in T$ contains T . Then for any $A(\alpha, r) \in \Psi$,*

$$\delta(A) \geq \sqrt{R^2 + r^2} - R \geq \min\{R, r^2/3R\}.$$

ii) *For any outer caustic Γ ,*

$$\sqrt{\rho_{\max}^2 + r_{\min}^2(\Gamma)} - \rho_{\max} \leq \delta_{\min}(\Gamma) < r_{\min}(\Gamma), \quad (2.17)$$

$$\sqrt{\rho_{\max}^2 + r_{\max}^2(\Gamma)} - \rho_{\max} \leq \delta_{\max}(\Gamma) < r_{\max}(\Gamma) < \delta_{\max}(\Gamma) + d. \quad (2.18)$$

iii) *If $r_{\min}(\Gamma) < \sqrt{3}\rho_{\max}$, then $r_{\min}^2(\Gamma) \leq 3\rho_{\max}\delta_{\min}(\Gamma)$. If $r_{\max}(\Gamma) \geq \sqrt{3}\rho_{\max}$, then $\rho_{\max} \leq \delta_{\max}(\Gamma)$.*

Proof. i) Let l be the supporting ray through A , let $M \in T$ be the point of support, and let $D = D(M, R)$ be the osculating disc. Then A is outside D , and $\delta(A) \geq \sqrt{R^2 + r^2} - R$, the distance from A to D . The other inequality is elementary. ii) Taking $R = \rho_{\max}$ in i), we have for any $A(\alpha, r) \in \Gamma$: $\sqrt{\rho_{\max}^2 + r^2} - \rho_{\max} < \delta(A) < r$. Maximizing and minimizing δ and r , we obtain Eqs. (2.17), (2.18) respectively. iii) By i) and Eqs. (2.17), (2.18). \square

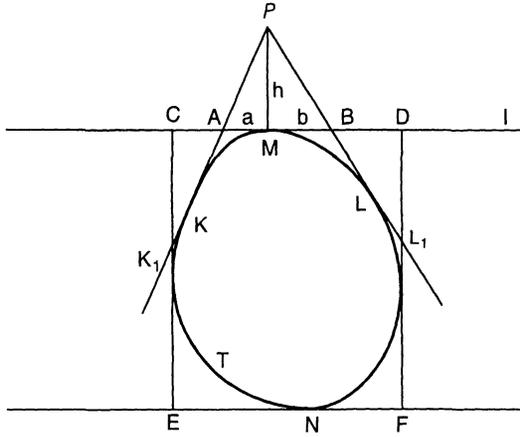


Fig. 2.7. Notation for Lemma 2.8.

Lemma 2.8. *Let $[T]$ be an arbitrary convex table, of diameter d and width w . Let $M \in T$, and let l be a supporting line for T at M . Let MP be orthogonal to l , and set $h = |MP|$ (Fig. 2.7). Let PK, PL be the supporting lines for T through P , and let $A, B \in l$ be the points of intersection. Then*

$$|AB| > hw/(h + d). \tag{2.19}$$

Proof. Let CE and DF be the supporting lines orthogonal to l , and let $K_1 = CE \cap PK$, $L_1 = DF \cap PL$ be the intersection points (Fig. 2.7). Set $|MA| = a$, $|MC| = u$, $|MB| = b$, $|MD| = v$. From the similar triangles AMP and ACK_1 , BMP and BDL_1 :

$$\frac{a}{h} = \frac{u - a}{|CK_1|}, \quad \frac{b}{h} = \frac{v - b}{|DL_1|}.$$

Since $|CK_1|, |DL_1| < d$, we have

$$\frac{a + b}{h} > \frac{(u + v) - (a + b)}{d} > \frac{w - (a + b)}{d},$$

which implies the claim. \square

Proposition 2.4. *Let Γ be an arbitrary outer caustic, and let $S(\Gamma)$ be the area parameter. Then*

$$S(\Gamma) > \frac{1}{2} \frac{\delta_{\max}^2(\Gamma)w}{\delta_{\max}(\Gamma) + d}, \tag{2.20}$$

$$S(\Gamma) > \frac{wr_{\max}^4(\Gamma)}{4(r_{\max}(\Gamma) + d)(r_{\max}^2(\Gamma) + 2\rho_{\max}^2)}. \tag{2.21}$$

Proof. We use notation of Fig. 2.8. For $M \in T$ and a supporting line $l = l(\alpha)$ at M , Let $P \in \Gamma$ be such that PM is perpendicular to l . Let PA, PB be the supporting lines for T from P . Since Γ is an outer caustic, it intersects the line PA (PB) at one more point, $A'(B')$. Therefore ΔABP is contained in the region between l and Γ , hence $S(\Gamma) > |\Delta ABP|$. Set $|PM| = h = h(\alpha)$. Since $\delta_{\max}(\Gamma) = \max_x h(\alpha)$,

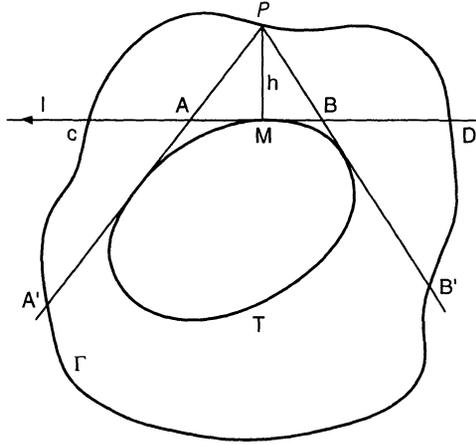


Fig. 2.8. Notation for the proof of Proposition 2.4.

Eq. (2.19) implies inequality (2.20). We set, until the end of the proof, $r_{\max}(\Gamma) = r$, $\rho_{\max} = \rho$. By Lemma 2.7,

$$\delta_{\max}(\Gamma) \geq \sqrt{r^2 + \rho^2} - \rho = \frac{r^2}{\sqrt{r^2 + \rho^2} + \rho}.$$

Substituting this into Eq. (2.20), and noting that $\delta_{\max} < r$, we obtain

$$S(\Gamma) > \frac{wr^4}{2(r^2 + 2\rho^2 + 2\rho\sqrt{r^2 + \rho^2})(r + d)}.$$

The inequality $\sqrt{r^2 + \rho^2} < \rho + r^2/2\rho$ implies Eq. (2.21). \square

We assume from now on that T satisfies

$$16d\rho_{\min}/w^2 < 1, \quad (2.22)$$

and let \mathcal{A} be the nonempty set of directions α such that

$$16d\rho(\alpha)/w^2 < 1. \quad (2.23)$$

Lemma 2.9. *Let $\bar{r}(\alpha)$ be as in Sect. 2.5. Let Γ be an outer caustic. For any $\alpha \in \mathcal{A}$ we have the dichotomy:*

$$2r(\alpha) < \bar{r}(\alpha) - \sqrt{\bar{r}^2(\alpha) - 4d\rho(\alpha)} = 2r_-(\alpha), \quad (2.24)$$

$$2r(\alpha) \geq \bar{r}(\alpha) + \sqrt{\bar{r}^2(\alpha) - 4d\rho(\alpha)} = 2r_+(\alpha), \quad (2.25)$$

Proof. If $r(\alpha) \geq \bar{r}(\alpha)$, Eq. (2.25) holds. For $r(\alpha) < \bar{r}(\alpha)$ we use Lemma 2.6, ii). By Eq. (2.23) and the inequality (it holds for all α) $2\bar{r}(\alpha) \geq w$, the roots, $r_{\pm}(\alpha)$, of the quadratic inequality (2.15) are real and distinct, hence the claim. \square

Definition 2.4. We choose a reference direction, $0 \in S^1$, so that $\rho(0) = \rho_{\min}$. We denote by G_1 (G_2) the set of outer caustics Γ such that Eq.(2.24) (Eq. (2.25)) holds, with $\alpha = 0$, and let G be the set of all outer caustics. By Lemma 2.9, $G = G_1 \cup G_2$, a disjoint union. We say that $\Gamma \in G$ is of the first (second) kind if $\Gamma \in G_1$ ($\Gamma \in G_2$).

Proposition 2.5. Let the notation be as above, and let Γ be an outer caustic. i) If $\Gamma \in G_1$, then

$$S(\Gamma) < \frac{64d^3}{w^3} \rho_{\min}^2. \tag{2.26}$$

ii) If $\Gamma \in G_2$, then

$$S(\Gamma) > \frac{w^3(w^2 - 32d\rho_{\min})}{8(w + 2d)(w^2 + 8\rho_{\max}^2)}. \tag{2.27}$$

Proof. i) We set $\alpha = 0$ and suppress it from notation. Using the inequality $1 - x < \sqrt{1 - x} < 1 - x/2$ with $x = 4d\rho_{\min}/\bar{r}^2$, and Eqs. (2.24–2.25), we obtain that the roots r_{\pm} satisfy

$$r_- < 2d\rho_{\min}/\bar{r} \leq 4d\rho_{\min}/w < w/4, \tag{2.28}$$

$$r_+ > [\bar{r}^2 - 2d\rho_{\min}]/\bar{r} > w[1 - 8d\rho_{\min}/w^2]/2 > w/4. \tag{2.29}$$

Let Γ be an outer caustic, and suppose $r(\alpha) < \bar{r}(\alpha)$ for some $\alpha \in S^1$. We use notation of Fig. 2.9. By $|\Delta ABC| = 2r^2(\alpha)/(\cot \theta + \cot \eta)$, and Eq. (2.16),

$$S < r^2(\alpha)d/[\bar{r}(\alpha) - r(\alpha)]. \tag{2.30}$$

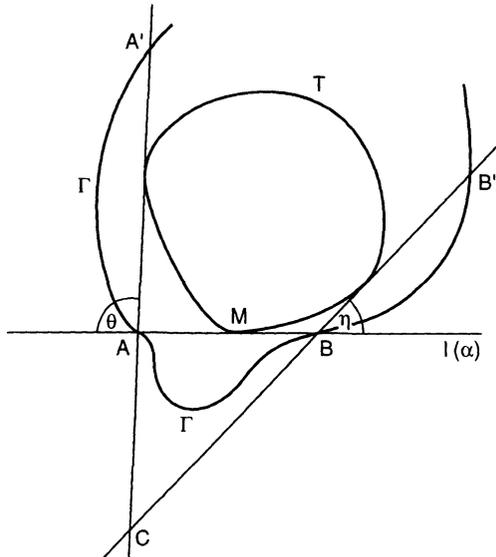


Fig. 2.9. Notation for the proof of Proposition 2.5.

By Eq. (2.28), $r = r(0) < w/4$. Since $\bar{r}(\alpha) \geq w/2$ for any α , the preceding inequality implies $S(\Gamma) < 4r^2d/w$. Using Lemma 2.9 and Eq. (2.28) again, we obtain Eq. (2.26). ii) For any $a, b > 0$ the function $f(x) = x^4/(x+a)(x^2+b^2)$ is monotonically increasing on $[0, \infty)$, and for any $x > \varepsilon > 0$.

$$f(x - \varepsilon) > \frac{x^4 - 4\varepsilon x^3}{(x + a)(x^2 + b^2)}.$$

(The proof is elementary, and we leave it to the reader). By Eqs. (2.21), (2.29),

$$S(\Gamma) > \frac{w}{4} f\left(\frac{w}{2} - \frac{4d\rho_{\min}}{w}\right),$$

which implies the claim. \square

2.7. Annulus Free of Outer Caustics and a Birkhoff Region of Instability

Lemma 2.10. *Suppose the outer table T satisfies*

$$32d\rho_{\min}/w^2 < 1. \quad (2.31)$$

Then for any outer caustic $\Gamma \in G_2$, we have, either $r_{\min}(\Gamma) \geq w/2$ or

$$\frac{r_{\min}(\Gamma)}{w} > \left[1 + \left(1 + \frac{16d(w+2d)(w^2+8\rho_{\max}^2)}{w^2(w^2-32d\rho_{\min})} \right)^{1/2} \right]^{-1}. \quad (2.32)$$

Proof. Suppose $r_{\min}(\Gamma) < w/2$, and set $r_{\min}(\Gamma) = w/2N, N > 1$. By Eq. (2.30), and $\bar{r}(\alpha) > w/2$ for any $\alpha \in S^1$, we have

$$S(\Gamma) < wd/2N(N-1).$$

By Proposition 2.5, $S(\Gamma)$ satisfies Eq. (2.27), hence

$$\frac{w^3(w^2-32d\rho_{\min})}{8(w+2d)(w^2+8\rho_{\max}^2)} < \frac{wd}{2N(N-1)}.$$

This implies, in view of Eq. (2.31),

$$N(N-1) < \frac{4d(w+2d)(w^2+8\rho_{\max}^2)}{w^2(w^2-32d\rho_{\min})}, \quad (2.33)$$

Using the identity $N(N-1) = (N-1/2)^2 - 1/4$, we rewrite this as

$$N < \frac{1}{2} \left[1 + \left(1 + \frac{16d(w+2d)(w^2+8\rho_{\max}^2)}{w^2(w^2-32d\rho_{\min})} \right)^{1/2} \right],$$

which implies Eq. (2.32). \square

Corollary 2.6. *Let the outer table satisfy*

$$64d\rho_{\min}/w^2 < 1. \quad (2.34)$$

Then for any $\Gamma \in G_2$,

$$r_{\min}(\Gamma) > \frac{w^3}{4(1 + 6\sqrt{2})d\rho_{\max}}. \tag{2.35}$$

Proof. If $r_{\min}(\Gamma) \geq w/2$, there is nothing to prove. We set $r_{\min}(\Gamma) = w/2N, N > 1$, and use Eq. (2.32). Using $(N - 1)^2 < N(N - 1)$, and that, by condition (2.34),

$$1 - 32d\rho_{\min}/w^2 > 1/2,$$

we obtain the estimate

$$N - 1 < \frac{2\sqrt{2}[d(w + 2d)(w^2 + 8\rho_{\max}^2)]^{1/2}}{w^2}.$$

From this, and $w \leq d \leq 2\rho_{\max}$, we get

$$N < 1 + 12\sqrt{2}d\rho_{\max}/w^2 < (2 + 12\sqrt{2})d\rho_{\max}/w^2,$$

which implies the claim. \square

Lemma 2.11. *Let Γ be any outer caustic. If*

$$S(\Gamma)/dw < 1/25, \tag{2.36}$$

then

$$r_{\max}^2(\Gamma) < 8\sqrt{d/w}\rho_{\max}\sqrt{S(\Gamma)}. \tag{2.37}$$

Proof. Set $S = S(\Gamma)$. By Eq. (2.20) (we leave details to the reader)

$$\delta_{\max}(\Gamma) < \frac{S}{w} + \sqrt{\frac{S^2}{w^2} + \frac{2Sd}{w}}.$$

To simplify notation, set $z = r_{\max}(\Gamma)$, $\rho = \rho_{\max}$, $\varepsilon^2 = S/4dw$, and let A be the right-hand side of the equation above. By Lemma 2.7, $\sqrt{z^2 + \rho^2} - \rho < A$. Using the inequality

$$\frac{z^2}{2\rho + z^2/\rho} < \sqrt{z^2 + \rho^2} - \rho$$

and that, by Eq. (2.36), $2\rho > A$, we obtain the bound

$$z^2 < 4\rho^2 A/(2\rho - A).$$

Since, by Eq. (2.36), $\varepsilon < 1/10$, the inequality

$$A = 4d\varepsilon^2 + 2\sqrt{2}d\varepsilon\sqrt{1 + 2\varepsilon^2} < 2\sqrt{2}d\varepsilon[1 + (1 + \sqrt{2})\varepsilon] < 4\sqrt{2}d\varepsilon[1 + (1 + \sqrt{2})\varepsilon]\rho$$

implies $A < \rho$ (in fact, $A < (.72)\rho$), thus, $z^2 < 4\rho A$. By the preceding inequality, $4\rho A < 8\sqrt{2}[1 + (1 + \sqrt{2})\varepsilon]d\rho\varepsilon < 16d\rho\varepsilon$. Substituting this into $z^2 < 4\rho A$, we obtain the claim. \square

Corollary 2.7. *Suppose the outer table T satisfies condition (2.34). Then for any $\Gamma \in G_1$ we have*

$$r_{\max}(\Gamma) < \sqrt{20}w^{-1}d\sqrt{\rho_{\min}\rho_{\max}}. \tag{2.38}$$

Proof. By Proposition 2.5, i)

$$S(\Gamma) < dw/64, \quad (2.39)$$

which is stronger than the assumption of the preceding lemma. The same proof, with the stronger bound (2.39) yields

$$r_{\max}^2(\Gamma) < \frac{5}{2} \sqrt{\frac{d}{w}} \rho_{\max} \sqrt{S(\Gamma)}, \quad (2.40)$$

which, by Eq. (2.26), implies the claim. \square

Now we can prove the main result of this section.

Theorem 2.2. *Let T be an outer table satisfying the condition*

$$\rho_{\min}^{1/2} \rho_{\max}^{3/2} < (8\sqrt{5}(1 + 6\sqrt{2}))^{-1} w^4/d^2. \quad (2.41)$$

Then there is a Birkhoff region of instability, $\mathcal{A} \subset \Psi$, $\partial\mathcal{A} = \Gamma_1 \cup \Gamma_2$, such that $\Gamma_1 \in G_1$, $\Gamma_2 \in G_2$. i) If $\rho_{\min} > 0$, then

$$|\mathcal{A}| \geq 20\pi \frac{d^2 \rho_{\min} \rho_{\max}}{w^2} \left[\frac{w^8}{320(1 + 6\sqrt{2})^2 d^4 \rho_{\min} \rho_{\max}^3} - 1 \right] > 0. \quad (2.42)$$

ii) *If $\rho_{\min} = 0$, then $\Gamma_1 = T$ (i.e., \mathcal{A} surrounds the table), and*

$$|\mathcal{A}| \geq \frac{\pi w^6}{16(1 + 6\sqrt{2})^2 d^2 \rho_{\max}^2}. \quad (2.43)$$

Proof. Condition (2.41) implies Eq. (2.34). Let $\Gamma_1 \in G_1$, $\Gamma_2 \in G_2$ be arbitrary outer caustics. By Corollaries 2.6, 2.7, and Eq. (2.41),

$$r_{\max}(\Gamma_1) < \frac{\sqrt{20}d\sqrt{\rho_{\min}\rho_{\max}}}{w} < \frac{w^3}{4(1 + 6\sqrt{2})d\rho_{\max}} < r_{\min}(\Gamma_2). \quad (2.44)$$

Set

$$a = \frac{\sqrt{20}d\sqrt{\rho_{\min}\rho_{\max}}}{w}, \quad b = \frac{w^3}{4(1 + 6\sqrt{2})d\rho_{\max}}.$$

The annulus $\tilde{\mathcal{A}} = \{(\alpha, r) : a \leq r \leq b\}$ is free of outer caustics, and let \mathcal{A} be the Birkhoff region of instability containing $\tilde{\mathcal{A}}$. Equations (2.42), (2.43) follow from $|\mathcal{A}| \geq |\tilde{\mathcal{A}}| = \pi(b^2 - a^2)$. \square

Corollary 2.8. *Let the outer table T satisfy*

$$\rho_{\min} \rho_{\max}^7 \leq 225(w/10)^8. \quad (2.45)$$

Then the conclusions of Theorem 2.2 hold.

Proof. The number $c = [8\sqrt{5}(6\sqrt{2} + 1)]^{-1}$ is between .006 and .0066, hence $c^2/16 > 225 \times 10^{-8}$. Since $d \leq 2\rho_{\max}$, the inequality (2.45) implies Eq. (2.41). \square

In the rest of this section we assume that T satisfies Eq. (2.41), and use the notation of Theorem 2.2. We assume, for simplicity of exposition, that \mathcal{A} is bounded.

(The possibility that \mathcal{A} is unbounded, i.e., $\partial\mathcal{A} = \Gamma_1$, remains open, see the discussion in Sect. 3). Then Γ_1 is the “last outer caustic of the 1st kind.” and Γ_2 is the “first outer caustic of the 2nd kind.” By results of Mather’s [17] and Le Calvez’ [5], the region \mathcal{A} contains orbits with complicated behavior. We formulate a simple special case, referring the reader to [17, 5] for complete statements.

Corollary 2.9. *Let the table T satisfy Eq. (2.41) or Eq. (2.45), and let the notation be as in Theorem 2.2. Then for any pair of indices $i, j \in \{1, 2\}$, there is an orbit $\mathcal{O} \subset \text{int}(\mathcal{A})$ which is α -asymptotic to a subset of Γ_i and ω -asymptotic to a subset of Γ_j .*

Remark. To illustrate the preceding results, consider the case $\rho_{\min}(T) = 0$. Recall that $\mathcal{A} \subset \Psi$ is the Birkhoff region of instability surrounding T . Then, by Corollary 2.9, there are orbits $\mathcal{O} = \{A_i : i \in \mathbf{Z}\} \subset \Psi$, with the following striking behavior. The “starting point”, $A_0 \in \mathcal{A}$, is far away from T , but $A_n = \psi^n(A_0)$ get arbitrarily close to T as $n \rightarrow +\infty$, or $n \rightarrow -\infty$, or both.

3. Examples and Discussion

3.1. Inner Billiard. Astroid as a Nonconvex Caustic. We use notation of Sect. 1. Let T be a billiard table. A C^1 invariant circle, $\Gamma \subset \Phi_T$, defines a caustic, $\gamma \subset \mathbf{R}^2$, as the envelope of the family of rays corresponding to Γ . We can show that, in fact, $\gamma \subset [T]$.

Let $\sigma : \Phi \rightarrow \Phi$, $\sigma(s, \theta) = (s, \pi - \theta)$, be the canonical involution (the time reversal symmetry). Since $\sigma\phi\sigma = \phi^{-1}$, the involution σ sends invariant circles into invariant circles. The pair, $\Gamma, \sigma(\Gamma)$, of invariant circles defines the same caustic, thus the correspondence $\Gamma \rightarrow \gamma$ is two-to-one.

Suppose Γ is piecewise C^1 . A corner point of Γ produces a jump discontinuity in the caustic. Let $A, B \in \gamma$ be the two focusing points in question. We include, by convention, the interval $[AB]$ into γ . For instance, let T be an ellipse with foci A, B . Let Γ be the family of rays passing through foci. Then Γ is an invariant circle. It is piecewise C^1 , and has two corners. The set of focusing points consists of A and B . At the corners of Γ the focusing point jumps from one of the foci to the other. By our convention, the caustic γ is the interval $[AB]$ traced twice, once in each direction.

If the billiard table is C^2 , then the map ϕ is C^1 , and Birkhoff’s theorem ensures only that the invariant circles are Lipschitz. We don’t know how to define the corresponding caustic if the invariant circle is only Lipschitz. On the other hand, we point out that nobody has yet exhibited an invariant circle that was not piecewise C^1 . Thus, in what follows, we restrict our attention to this class of invariant circles.

What closed plane curves γ can be caustics? If γ is convex, the string construction, see Sect. 3, provides a one-parameter family of tables, $T = T_r(\gamma)$, with the caustic γ . Note that the string construction, in general, increases the regularity by one. Thus, if γ is piecewise C^1 , the tables T will be only piecewise C^2 , in general. If γ is a convex polygon, then T is a finite union of elliptic arcs [12], and the curvature of T jumps at the junctions. In special cases, however, T may have greater regularity than piecewise C^2 . One such special case, of course, is $\gamma = [AB]$, an interval.

Which nonconvex curves can occur as caustics? R. Douady [6] mentions nonconvex caustics as a possibility (referring to his Fig. 3 on page IV-5 that shows a

billiard table with a special nonconvex caustic). To our knowledge, this is the only discussion of the subject in the literature. We point out that the string construction, as a means of producing billiard tables with a given caustic, can be extended to some nonconvex curves. By way of example, we discuss the case of *astroid*.

The astroid, $\gamma \subset \mathbf{R}^2$, is given by the equation

$$|x|^{2/3} + |y|^{2/3} = 1. \quad (3.1)$$

Denote by C_1, \dots, C_4 the four quadrants of the plane, and let $\gamma_1, \dots, \gamma_4$ be the corresponding arcs of γ (see Fig. 3.1). From every point, M , in the exterior of γ there are two tangent lines to γ . For instance, if $M \in C_1$, the tangency points, A, B , belong to γ_2 and γ_4 (Fig. 3.1). If M is on the coordinate axes, the tangent lines in question coincide, but the points of tangency remain distinct. They are the opposite cusps of γ .

Let $d\sigma$ be the arclength element on γ . For any $A, B \in \gamma$ we use notation $|\gamma_B^A|$ for the length of the arc of γ swept by a point moving along γ from B to A in the positive direction.

For $M \in C_1$ let $A = A(M) \in \gamma_2, B = B(M) \in \gamma_4$ be the points of tangency. Fix $L \geq -1$ and define the curve $T_1(L) \subset C_1$ by $T_1(L) = \{M \in C_1 : |AM| + |BM| - |\gamma_B^A| = L\}$. For $L > -1$ the curve $T_1(L)$ is real analytic, and belongs to the exterior of γ . The endpoints of $T_1(L)$ are on the coordinate axes. Analogously, we define the curves $T_i(L) \subset C_i, i = 2, 3, 4$. The four curves, $T_i(L)$, fit together, making a closed convex real analytic (for $L > -1$) curve $T(L)$, containing the astroid γ in its interior (Fig. 3.1). The curve $T(-1)$ passes through the cusps of γ , and has an infinite curvature there.

The description of $T(L)$ as the locus of points satisfying

$$|AM| + |BM| - |\gamma_B^A| = L \quad (3.2)$$

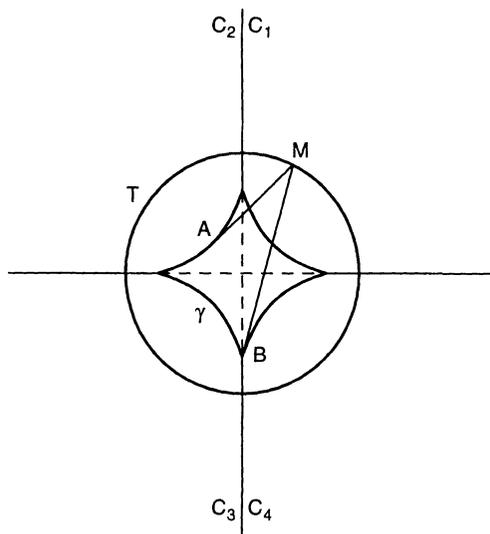


Fig. 3.1. Astroid as an inner caustic.

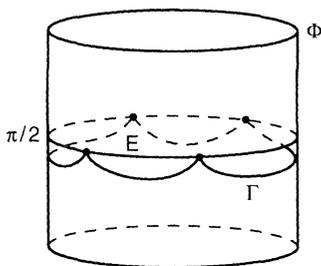


Fig. 3.2. Invariant circle corresponding to the astroid.

is completely analogous to the string construction, see Sect. 1.3. In fact, γ is a caustic for the table $T = T(L)$, and L plays the role of the Lazutkin parameter. More precisely, let Γ be the set of tangent rays to γ oriented so that the angle $\theta(M)$ (Fig. 3.1) satisfies $\theta(M) \leq \pi/2$. Then $\Gamma \subset \Phi_T$ is an invariant circle, and let $\Gamma' = \sigma(\Gamma)$ be its “mirror image.” The invariant circles Γ, Γ' have four points in common, the four corners corresponding to the four cusps of the astroid, see Fig. 3.2 (Γ' , which is the reflection of Γ about the equator, is not shown). These corner points belong to the equator, $E = \{\theta = \pi/2\} \subset \Phi$, and are periodic with period two. At the corners the focusing point (living on γ) jumps from a cusp of the astroid to the opposite one. Thus, by our convention, the caustic γ is the astroid together with the “diagonals” (Fig. 3.1). When the focusing point traces γ , it runs through each diagonal twice, once in each direction.

The example above demonstrates some of the differences between convex and nonconvex caustics. Note that while we are traversing Γ in the positive direction, the focusing point moves in the negative direction on γ . Since Γ contains periodic points of period two, the rotation number is equal to $1/2$, for any L .

3.2. *Billiard Tables of Constant Width and Related Examples.* Let T be a closed convex curve. For any direction α denote by $w(\alpha)$ the width of the strip formed by the tangent rays with the directions $\alpha - \pi/2$ and $\alpha + \pi/2$. The number $w(\alpha)$ is the width of T in direction α . The width and the diameter of T (see Sects. 1, 2) are given by $w = \min_{\alpha} w(\alpha), d = \max_{\alpha} w(\alpha)$.

Definition 3.1. (see, e.g., [22]). *If $w(\alpha) = \text{const.}$, we say that T is a curve of constant width.*

The circle is a trivial example of a curve of constant width. A popular example is the Releaux triangle [22]. It is formed by three circular arcs of the same radius, r , centered at the vertices of the equilateral triangle with the side r . The Releaux triangle is piecewise C^1 , it has three corners. A C^2 curve, T , has a constant width, w , if and only if its radius of curvature satisfies $\rho(\alpha) + \rho(\alpha + \pi) = w$. We will use the following characterization of curves of constant width [22]. Any chord of T , perpendicular to T at one end, is perpendicular to T at the other end.

Let Φ be the phase space of the billiard map, and let $E \subset \Phi$ be the equator. For any table T the time reversal involution is the reflection about E . The characterisation above means that a billiard table T has constant width if and only if the equator is an invariant circle (it consists of periodic points of period two).

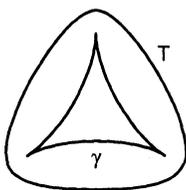


Fig. 3.3. Involute of a table of constant width.

Proposition 3.1. *Let T be a noncircular C^3 table of constant width. The caustic, $\varepsilon = \gamma(E)$, corresponding to the equator, is not convex.*

Proof. By preceding remarks, ε is the envelope of the normals to T , i.e., it is the involute of T (see, e.g., [2]). By our assumptions, the radius of curvature, $\rho(\cdot)$, of T is not constant. Hence ε has cusps, corresponding to the critical points of $\rho(\cdot)$. \square

Remark. For a typical table of constant width, the caustic ε consists of a finite number of concave arcs, joined by cusps, see Fig. 3.3. We also have examples of caustics that contain both convex and concave arcs. (They are joined by cusps. We can show that a caustic does not have inflection points.)

Proposition 3.1 shows that nonconvex caustics are not as rare as it seemed. Let T be a table of constant width, and let ε be the involute of T . Let $\Gamma \subset \Phi$ be an invariant circle near equator. Then the corresponding caustic, γ , is close to ε , hence it has cusps. This indicates that a typical table of constant width has infinitely many nonconvex caustics, with rotation numbers close to $1/2$.

The coordinates (s, θ) or (α, θ) on Φ have a geometric meaning, where $\theta, 0 \leq \theta \leq \pi$, is the height in Φ . The equator is a circle of constant height, $E = \{\theta = \pi/2\}$. It is natural to ask if other circles of constant height, $\Gamma_\theta \subset \Phi, \theta \neq 0, \pi/2, \pi$, can be invariant. It turns out that Γ_θ can be an invariant circle only if θ satisfies $\tan n\theta = n \tan \theta$, where $n > 1$ is an integer [8]. This equation has n solutions, $\theta_i(n)$ (including the two trivial ones), and there is a one-parameter family of real analytic billiard tables, $T_{(a,n)}, 0 < a < 1$, with the invariant circles $\Gamma_i^{(n)} = \{(s, \theta) : \theta = \theta_i(n)\}$ [8]. The caustics, $\gamma_i^{(n)}(a) \subset \mathbf{R}^2$, corresponding to these invariant circles are not convex, in general.

3.3. Examples of Outer Caustics. In this subsection T is an outer billiard table, $\Psi = \mathbf{R}^2 \setminus \text{int}(T)$ is the phase space of the outer billiard map, and $\Gamma \subset \Psi$ is an outer caustic. Then T is obtained from Γ by the area construction, with the area parameter $S = S(\Gamma)$, see Sect. 2.3. This observation allows to construct outer billiard tables with a prescribed outer caustic.

Let Γ be a Jordan curve enclosing the area $||[\Gamma]||$, and let $0 \leq S \leq ||[\Gamma]||$. We consider the family of rays, $\ell_\theta, 0 \leq \theta \leq 2\pi$, cutting the area S off of Γ . Let $T = T(\Gamma, S)$ be the envelope of the family. Assume that the region Ω_θ enclosed between the chord $C_\theta \subset \ell_\theta$ and Γ is connected for all θ (Fig. 3.4). Then T is the locus of the midpoints of $C_\theta, 0 \leq \theta \leq 2\pi$. Hence, if T is convex, it is an outer table, and Γ is an outer caustic for T .

When is $T = T(\Gamma, S)$ convex? Examples show that, in general, $T(\Gamma, S)$ is not convex for all S even when Γ itself is convex [7]. Note that if Γ is centrally symmetric and convex, then $T(\Gamma, ||[\Gamma]||/2) = O$, the centre of symmetry. Let α, β be the

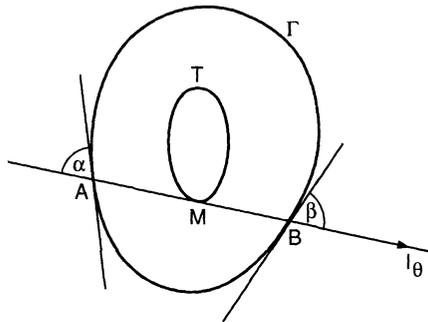


Fig. 3.4. Outer table with a prescribed outer caustic (area construction).

angles between Γ and C_θ at the intersection points A, B (Fig. 3.4). Proposition 2.2 yields a necessary condition for T to be convex: $\cot \alpha + \cot \beta \geq 0$. Under some technical assumptions, this condition is also sufficient. In particular, we have the following result.

Proposition 3.2. *Let Γ be a closed convex curve. There is a positive number, S_0 , depending on Γ such that the curve $T(\Gamma, S)$ is convex for any $S \leq S_0$.*

It is also clear from Proposition 2.2 that when $\alpha + \beta = \pi$, the corresponding point $M \in T$ has infinite curvature. Combining this with the preceding remark, we obtain the following.

Proposition 3.3. *Let Γ be a convex piecewise C^1 curve, and let S be such that the angles at the intersection satisfy $\alpha, \beta \leq \pi/2$ (Fig. 3.4). Then $T = T(\Gamma, S)$ is convex, and has at most a finite number of points of infinite curvature (corresponding to $\alpha = \beta = \pi/2$).*

Remark. It is by no means necessary for Γ to be convex in order that $T(\Gamma, S)$ be convex. For instance, let Γ be a small nonconvex C^1 perturbation of the unit circle. By choosing S small, but not too small, we can ensure that $T(\Gamma, S)$ is convex.

Let Γ be the unit square, and let $0 < S < 1/2$. The curve $T = T(\Gamma, S)$ is the union of four hyperbolic arcs (Fig. 3.5). Each diagonal of Γ is the axis of symmetry for two of the hyperbolas. In view of the obvious fourfold symmetry, it suffices to determine one of these hyperbolic arcs, e.g., HAE . The corresponding hyperbola is given by $xy = S/2$ in natural coordinates. The points A, E, H have coordinates $(\sqrt{S/2}, \sqrt{S/2}), (1/2, S), (S, 1/2)$ respectively. The angle at the vertex H is equal to $2 \arctan 1/2S$. In particular, T is convex, and has four corner points: E, F, G, H . The corner angle monotonically decreases from π (at $S = 0$) to $\pi/2$ (at $S = 1/2$). When $S = 1/2$, the curve T degenerates into the center of symmetry, O .

Let Γ be the parachute-shaped curve of Fig. 3.6 (the precise form of Γ is not important). We normalize Γ so that it encloses the unit disc centered at O (Fig. 3.6). The area in Γ below the diameter AB is equal to $\delta + \pi/2$, where $\delta > 0$ is the area below the dotted line in Fig. 3.6. Assume, for simplicity, that Γ is symmetric about the vertical axis through O . Proposition 3.3 and the elementary considerations show that for any $S, 0 \leq S \leq \pi/2$ the curve $T = T(\Gamma, S)$ is convex. For $S < \pi/2$, the table T is C^2 , but for $S = \pi/2$ it has a corner at O . The angle at O is $\pi - 2AOC$, i.e., it is determined by the length of the arc AC alone. In the “degenerate case”,

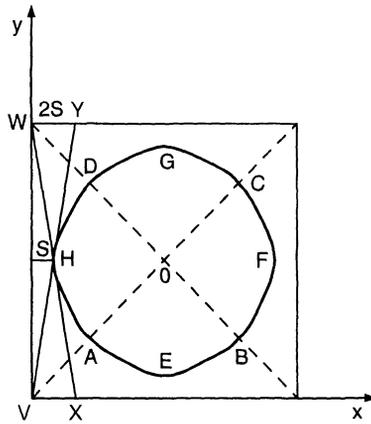


Fig. 3.5. Square as an outer caustic.

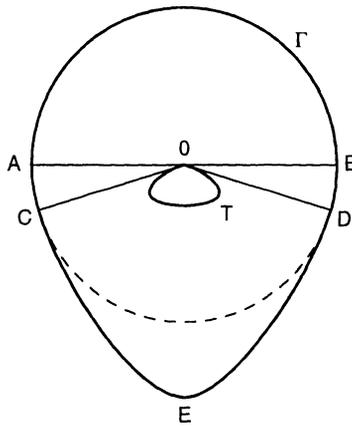


Fig. 3.6. "Parachute" as an outer caustic.

$A = C$, the point $O \in T$ is a C^1 point with $\rho(O) = 0$. Thus we obtain examples of C^1 outer tables T with $\rho_{\min}(T) = 0$, having outer caustics.

The examples above show that having a point of zero radius of curvature on an outer table does not preclude the existence of outer caustics sufficiently far away from the table. At present, there are no examples of outer tables where the nonexistence of outer caustics (at least at infinity) is proven. Outer tables made from two circular arcs seem to be good candidates [20].

References

1. Arnold, V.I.: Mathematical methods of classical mechanics. Berlin, Heidelberg, New York: Springer, 1987
2. Berger, M.: Geometry. Berlin, Heidelberg, New York: Springer, 1987
3. Birkhoff, G.D.: Collected Mathematical Papers, Vol. II. Providence, RI: Am. Math. Soc., 1950

4. Blaschke, W.: *Differentialgeometrie I, II*. New York: Chelsea, 1967
5. Calvez, P.Lc: Propriétés dynamiques des difféomorphismes de l'anneau et du tore. *Asterisque* **204**, Soc. Math. France, Paris, 1991
6. Douady, R.: Thèse de troisième cycle. Université Paris VII, 1982
7. Fuks, D.B., Tabachnikov, S.: Segments of equal areas. *Quantum* **2**, 26–31 (1992)
8. Gutkin, E.: Billiard tables of constant width and dynamical characterizations of the circle. Workshop on dynamics and related questions, Proceedings, PennState U., 1993
9. Gutkin, E., Katok, A.: A priori estimates on caustics for inner and outer billiards. Workshop on dynamics and related questions, Proceedings, U. Maryland, 1993, 20–27
10. Gutkin, E., Simanyi, N.: Polygonal dual billiards and necklace dynamics. *Commun. Math. Phys.* **143**, 431–450 (1992)
11. Herman, M. (with an Appendix by A. Fathi): Sur les courbes invariantes par les difféomorphismes de l'anneau. Vol. **1**, *Asterisque* 103–104, Soc. Math. France, Paris, 1983
12. Hubacher, A.: Instability of the boundary in the billiard ball problem. *Commun. Math. Phys.* **108**, 483–488 (1987)
13. Katok, A., Hasselblatt, B.: *Introduction to the modern theory of dynamical systems*. Cambridge: Cambridge University Press, 1995
14. Katok, A., Strelcyn, J.-M.: *Invariant Manifolds, Entropy and Billiards; Smooth Maps with Singularities*. *Lecture Notes Math.* **1222**, Berlin, Heidelberg, New York: Springer, 1986
15. Lazutkin, V.F.: The existence of caustics for a billiard problem in a convex domain. *Math. USSR Izvestija* **7**, 185–214 (1973)
16. Mather, J.: Glancing billiards. *Ergod. Theory and Dyn. Syst.* **2**, 397–403 (1982)
17. Mather, J.: Variational construction of orbits of twist difféomorphisms. *J. Amer. Math. Soc.* **4**, 207–263 (1991)
18. Moser, J.K.: *Stable and random motions in dynamical systems*. *Annals of Mathematics Studies* **77**, Princeton, NJ: Princeton University Press, 1973
19. Sinai, Ya.G.: *Introduction to ergodic theory*. Princeton, NJ: Princeton University Press, 1977
20. Tabachnikov, S., Monroe, I.: *Asymptotic dynamics of the dual billiard map*. Preprint, University of Arkansas, 1992
21. Wojtkowski, M.: Principles for the design of billiards with nonvanishing Lyapunov exponents. *Commun. Math. Phys.* **105**, 391–414 (1986)
22. Yaglom, I.M., Boltyanskii, V.G.: *Convex Figures*. New York, Holt, Rinehart and Winston, 1961

Added in Proof.

23. Boyland, P.: Dual billiards, twist maps, and impact oscillators. Preprint, Suny Stony Brook, 1994

Communicated by M. Herman

