Hyperbolic outer billiards: a first example

Daniel Genin

Department of Mathematics, Penn State University, University Park, PA 16802, USA

Received 29 November 2005, in final form 10 April 2006
Published 15 May 2006
Online at stacks.iop.org/Non/19/1403

Recommended by L Bunimovich

Abstract

We present the first example of a hyperbolic outer billiard. More precisely we construct a one-parameter family of examples which in some sense corresponds to the Bunimovich billiards.

Mathematics Subject Classification: 37D50, 37D40, 37D25

(Some figures in this article are in colour only in the electronic version)

1. Introduction

Since the introduction of outer billiards by Neumann in 1959 [14] and their popularization by Moser in [13] and [12] there have been several developments indicating that their dynamics in some respects parallels the dynamics of ordinary billiards [2, 6, 16, 17]. For example, it has been shown that the analogue of the Lazutkin’s theorem holds for outer billiards [4] and that the string construction in ordinary billiards has its parallel in the secant area construction of outer billiards [1]. In contrast, the rich theory of chaotic billiards so far has no counterpart, in significant part due to the lack of examples. The aim of the present paper is to make a first step towards eliminating this void, by providing a first example of a hyperbolic outer billiard.

More precisely we produce a one-parameter family of outer billiards which has a square invariant region of positive measure on which the Lyapunov exponents of the outer billiard map are non-zero almost everywhere. If one were to draw an analogy with ordinary billiards this family of examples most closely parallels the billiards of Bunimovich [3], in that the billiards are composed of arcs of hyperbolas which individually have ‘integrable’ dynamics joined by ‘neutral’ segments—corners. As in the case of Bunimovich billiards non-vanishing of the Lyapunov exponents guarantees positive entropy and other nice properties. Also as in the case of Bunimovich billiards ergodicity does not come for free and has to be proved separately. We expect, nevertheless, that the outer billiards in our family of examples are indeed ergodic in the invariant region.

Dynamics outside the invariant region appears to be non-hyperbolic. Numerical explorations indicate the presence of invariant curves and elliptic islands. So this family
of outer billiards, in addition, provides a nice example of coexistence of hyperbolic and elliptic behaviour.

The proof of the main result uses the cone field method introduced by Wojtkowski in [19]. After producing an invariant region we construct a measurable field of cones defined almost everywhere in the invariant region and then show that the cones are eventually strictly preserved. We begin the exposition with a brief introduction to outer billiards followed by a section on the secant area construction which allows us to obtain an invariant region, or to be exact, a table for a given invariant region; in the following section we define the cone field over the invariant region and then prove that it is eventually strictly preserved in the last section.

2. Outer billiards

We begin with a definition of the outer billiard map (see figure 1).

**Definition 2.1.** Let $\Gamma$ be an oriented strictly convex plane curve in $\mathbb{R}^2$ and $\Omega$ the domain enclosed by $\Gamma$. Suppose for the moment that $\Omega$ is strictly convex and let $D = \mathbb{R}^2 \setminus \Omega$. The outer billiard map is the continuous transformation $T : D \to D$ uniquely specified by

(i) the oriented segment $[p, T(p)]$ is tangent to $\Gamma$ at some $p_i$;
(ii) orientation of $\Gamma$ agrees with the orientation of the segment $[p, T(p)]$ at $p_i$;
(iii) $p_i$ bisects $[p, T(p)]$.

In analogy with inner billiards we will say that $p$ reflects at the boundary in $p_i$. It is easy to see that the resulting map is a continuous transformation of $D$. The condition of strict convexity can be relaxed to allow outer billiards with discontinuities. Flat segments of the boundary are equivalent to corners for inner billiards—the orbit of a point reflecting in a flat segment cannot be extended. Outer billiards about polygons are of this type and have been studied extensively in recent times as part of a more general programme for understanding the complexity arising in piece-wise isometries [8, 9, 11].

It is also easy to show that $T$ is an area preserving twist map, the invariant measure being simply the Lebesgue measure on $D$ [17].

Outer billiards have many remarkable properties (see for example [7,10,15–17]).
Hyperbolic outer billiards

3. Secant area construction

We begin by describing a construction that a given convex plane curve produces an outer billiard table for which the curve is an invariant one. This construction parallels the string construction of inner billiards which does the same for caustics.

Given a convex plane curve $\gamma$ and a parameter $a$ satisfying $0 < a < A$, where $A$ is the area enclosed by $\gamma$, we consider the family of lines $L$ that divide the region enclosed by $\gamma$ into parts with area $a$ and $A - a$. We have the following result.

**Lemma 3.1.** The envelope of $L$ is a closed curve $\Gamma_1$, which is convex if there are no cusps. Furthermore, $\gamma$ is an invariant curve of outer billiard about $\Gamma_1$.

The proof of this result can be found in [16].

Using the area construction we can construct a table with an arbitrary convex invariant region.

4. Table construction

We consider a one-parameter family of outer billiards obtained by the area construction from the unit square $\gamma$ (see figure 2). Fix the area parameter to be $2a$ and let $B$ be the (closed) billiard table, i.e. the closed region enclosed by $\Gamma_1$ and $D$ be the complement of $B$ in the open unit square. Let $\Gamma_1$ be oriented in the counter-clockwise direction.

**Lemma 4.1.** $\Gamma_1$ is composed of four arcs of hyperbolas, asymptotic to lines containing the sides of $\gamma$, meeting in corners opposite the centres of the sides of $\gamma$.

The proof of the above lemma is elementary and is left to the reader. Depending on which arc of $B$ contains $p_1$, it is convenient to use a different coordinate system on $D$. Each arc naturally corresponds to a corner, which we choose to be the origin when considering points reflecting in this arc. Choose the outgoing edge (considering $\gamma$ oriented clockwise) to be the $x$-axis and the incoming, the $y$-axis. We introduce corresponding coordinate functions $x(p)$ and $y(p)$. In these coordinates the hyperbolic arcs of $B$ from the above lemma are given by the equation $y(p)x(p) = a$ (see figure 3).
As a consequence of the construction $D$ is an invariant region for the outer billiard about $\Gamma$.

5. Cone field

We will define an invariant cone field on a subset $D_h$ of $D$ defined by $D_h = \{ p \in D | T^k(p) \text{ is not a corner of } B \text{ for some } k \in \mathbb{Z} \}$. That is, $D_h$ is the set of all points in $D$ whose orbits touch the interior of one of the hyperbolic arcs of $B$ at least once.

**Lemma 5.1.** The points in $D \setminus D_h$ are periodic.

**Proof.** A point in $D \setminus D_h$ reflects only in the corners of $B$. Hence so far as this is concerned the table is a square. But it is well known that the outer billiard about any lattice polygon has only periodic points (see for example [17]). \( \square \)

The set $D \setminus D_h$ is indicated in figure 5 in black. It is easy to see that all points in $D \setminus D_h$ have the same period—four. So $D \setminus D_h$ is fixed by $T^4$ and hence is an elliptic domain.

A cone $C(p) \subset T_p D_h$ will be defined by a pair of vectors $(u, v) \in T_p D_h^2$, $C(p) = \{ w \in T_p D_h | [u, w][w, v] > 0 \}$, where $[\cdot, \cdot]$ stands for the cross product induced by identifying $T_p D_h$ with the ambient $\mathbb{R}^2$. It is clear that scaling $u$ and $v$ in the above definition by positive constants does not change $C(p)$ so whenever we speak of the equality of vector defining cones we shall mean equality up to multiplication by a positive constant.

We will first define the invariant cone field on the points in $D_h$ for which $p_t$ is not a corner of $B$. The cone field can then be extended to all points in $D_h$ which do not land on discontinuities of $dT$ by pulling back the cones by $dT$. We set $C(p) = (u(p), v(p))$, where for $p$ with $x(p) = x, y(p) = y$

$$u(p) = (1, -y/x)$$

is a vector tangent to the homothetic hyperbola passing through $p$, and

$$v(p) = p - p_t = \left( x - \frac{a + \sqrt{a^2 - axy}}{y}, y - \frac{ay}{a + \sqrt{a^2 - axy}} \right),$$

(see figure 3). In the last formula we used

$$(x(p_t), y(p_t)) = \left( \frac{a + \sqrt{a^2 - axy}}{y}, \frac{ay}{a + \sqrt{a^2 - axy}} \right).$$
For convenience we will define the cone field at points of $\gamma$ (which are not in $D$) by taking $u(p)$ to be the tangent vector to $\gamma$ in the corresponding direction.

6. Hyperbolicity

Now we are ready to prove the main result.

**Theorem 6.1.** $C(p)$ is eventually strictly preserved under $T$ a.e. in $D_h$ for $a \in (0, (3-\sqrt{5})/8)$. Hence $T$ has non-vanishing Lyapunov exponents a.e. in $D_h$ for the corresponding set of parameter values.

We will prove this by showing that the cone field defined in the previous section is eventually strictly preserved.

**Definition 6.2.**

1. A cone field $C(p)$ is preserved at a point $p$ if $dT TC(p) \subset C(T(p))$.
2. A cone field is strictly preserved if $dTC(p) \subset \circ C(T(p))$.
3. A cone field is eventually strictly preserved if for almost every $p$ there exists an $n(p)$ such that the cone field is strictly preserved at $T^n(p)$.

The result above then follows from the result of Wojtkowski introduced in [18].

**Theorem 6.3.** If there exists a measurable bundle of sectors which is eventually strictly preserved by $\Phi : M^2 \to M^2$, where $M^2$ is a two-dimensional Riemannian manifold, satisfying the following.

1. $\Phi$ preserves a probabilistic measure $\mu$ which has a non-vanishing density with respect to the Riemann area element on $M^2$ and
2. The singularities of $\Phi$ satisfy
   $$\int_{M^2} \log^+ \|D\Phi\| \, d\mu(x) < +\infty,$$

   where $\log^+ t = \max(\log t, 0)$.

Then the Lyapunov exponent $\lambda_+$ of $\Phi$ is positive $\mu$ a.e.

$T$ satisfies the first condition above because it preserves the Lebesgue measure on $D_h$. It also satisfies the second condition because the differential of $T$ is bounded in norm on $D_h$. Indeed, the differential of an outer billiard map blows up only if the table has points of vanishing curvature, which $B$ does not. Thus the above result may be applied to $T$ if $C(p)$ is eventually strictly preserved. This will be proved in a series of lemmas because the argument naturally divides into cases. The argument is essentially different for points with $p_t$ and $T(p_t)$ belonging to the interior of the same hyperbolic segment of $B$ and for points for which they belong to interiors of different hyperbolic segments.

We begin by examining the first case.

**Lemma 6.4.** $C(p)$ is preserved at points such that $p_t$ and $T(p_t)$ belong to the interior of the same hyperbolic segment.

**Proof.** Since we are concerned with one hyperbolic arc we can, for the moment, forget about the rest of the table and consider the outer billiard about a single branch of hyperbola.

Simple algebra shows that $xy$ is an integral of motion for $T$ and hence the homothetic hyperbolas $xy = c$ are preserved by $T$. Since $T$ preserves the order of points on the invariant
hyperbolas the vector \( u(p) \) tangent to a hyperbola at \( p \) is mapped by \( dT \) to a vector tangent to the same hyperbola at \( T(p) \). That is, \( dT_p u(p) = u(T(p)) \). Also the tangent vector \( v(p) = p - p_t \) is mapped to minus itself \( dT_p v(p) = -v(p) \) since restricted to the line of tangency, the billiard map is simply a reflection in the tangency point. Hence the image cone \( dT_p C(p) = (u(T(p)), -v(p)) \). Noting that the vectors \( u(T(p)) \) and \( dT_p u(p) \) are the same up to rescaling by a positive constant it is enough to check that \( dT_p v(p) \in C(T(p)) \). We will show that the angle of \( C(p) \) is always obtuse and the angle of \( dT_p C(p) \) is always acute. Observe that \( v(p) \) always lies in the fourth quadrant while \( u(p) \) always lies in the second. So the angle between them is always obtuse. Similarly \( u(T(p)) \) is always in the second quadrant and so is \( -v(p) \), so the angle between them is always acute. Hence \( dT_p C(p) \subset C(T(p)) \) for points reflecting in a single hyperbolic segment.

Note that \( dT_p v(p) \) is strictly inside \( C(T(p)) \) although \( dT_p u(p) = u(T(p)) \) up to rescaling by a positive constant, so the inclusion is not strict. □

The following consequence of the proof is useful in itself.

**Lemma 6.5.** For every point \( p \in D_\alpha \) such that \( p_t \) is in the interior of a hyperbolic arc.

1. \( C(p) \) has an obtuse angle and
2. \( dT_p C(p) \) has an acute angle.

Before moving on to the second case we observe that cones \( C(p) \) are nested in a special way along the line through \( p, p_t \).

**Lemma 6.6.** Let \( p \) and \( p' \) be two points satisfying \( p_t = p'_t \) and \( |p - p_t| > |p' - p_t| \) then

1. \( C(p) \subset C(p') \)
2. \( dT_p C(p') \subset dT_p C(p) \)

where the tangent spaces are identified by parallel translation to determine the inclusion relationship between the cones (figure 4).

Proof of this statement involves only elementary geometry and is left to the reader. The above lemma provides an easy way of determining whether a cone field at \( p \) is preserved. Let \( p_0 \)
be the intersection of the ray from \( p_t \) through \( p \) with \( \gamma \); then to check that the cone field is preserved at \( T(p) \) it is enough to check the inclusion between cones at \( p_b \) and \( T(p)_b \). Indeed, if \( d_{T(p)}C(p_b) \subset C(T(p)_b) \) then

\[
d_{T(p)}C(p) \subset d_{T(p)}C(p_b) \subset C(T(p)_b) \subset C(T(p)),
\]

so

\[
d_{T(p)}C(p) \subset C(T(p)).
\]

Note that for \( p \in D_h \) if \( d_{T(p)}v(p_b) \) is in the interior of \( C(T(p)_b) \), then \( d_{T(p)}C(p) \not\subset C(T(p)) \), i.e. the inclusion becomes strict, because from the above lemma \( d_{T(p)}u(p) \) is in the interior of \( d_{T(p)}C(p_b) \) (figure 4).

We now proceed to examine the points for which \( p_t \) and \( T(p)_t \) belong to interiors of different hyperbolic segments of \( B \) or corners.

**Lemma 6.7.** \( C(p) \) is strictly preserved at points such that \( p_t \) and \( T(p)_t \) belong to interiors of different hyperbolic segments.

**Proof.** If \( p_t \) and \( T(p)_t \) belong to interiors of different hyperbolic segments these segments must be adjacent. In this case (figure 4) \( T(p)_b \) and \( T(p)_h \) belong to the same side of \( \gamma \) which means that \( d_{T(p)}u(p_b) = u(T(p)_b) \). Furthermore, \( d_{T(p)}v(p_b) = v(T(p)_b) \in C(T(p)_h) \) by lemma 6.5. This implies that \( d_{T(p)}C(p) \not\subset C(T(p)) \).

\[\square\]

**Lemma 6.8.** \( C(p) \) is strictly preserved at points such that \( p_t \) is a corner point.

**Proof.** Considering all points such that \( p_t \) is a corner point we obtain figure 5 which shows points reflecting in corners and their images.

In figure 5 all points reflecting in corners are shaded: those that reflect in one corner between reflections in the interiors of sides are shaded diagonally, those that reflect in two corners are shaded horizontally and those reflecting in three are doubly shaded. Any point that hits all four corners is periodic, as was proved in lemma 5.1, and these points are solidly shaded in black. The lines in the diagram are the lines of discontinuity of the derivative and
their preimages. We will use this diagram to reduce the verification of preservation of the cone field at shaded points to the consideration of orbits of a few representative points. For this it is important to note that the diagram in figure 5 is structurally invariant for $a \in (0, 1/4)$ (which includes the parameter range under consideration), the interval in which the lines tangent to $B$ and passing through the corners of $\gamma$ do not intersect in the interior of $D$.

We first argue that we only need to consider preservation inside a few of the shaded polygons. This follows because the cone field and the map are both preserved under the action of the rotation subgroup of $D_4$. Hence if the cone field is preserved inside some shaded polygon of figure 5 then it is preserved in every polygon of its orbit under the rotation subgroup. Therefore, it is enough to check preservation for one polygon per orbit.

Next, we show that for points in the shaded polygons only cones at the first and last points of the orbit segment inside the shaded set need to be compared. $T$ is a central symmetry for every point $p$ with $p_t$ a corner of $B$. So $dT_p$ at such a point is $-I$ and $dT_pC(p) = C(p)$. Hence it is enough to check inclusion of cones at successive points of the orbit that reflect in the interiors of sides. Thus we ignore the points of the orbit segment that reflect in corners, and let $p$ and $p'$ be the first and last points respectively such that $p_t$ and $p'_t$ belong to the interiors of sides. As before, $p_b$ and $p'_b$ will be intersections of the corresponding rays with $\gamma$.

We further show that for points in the interior of a given polygon of the above diagram there is only one way for the boundary points $p_b, T(p)_b, ..., p'_b$ to be distributed on the sides of $\gamma$. This allows us to determine the preservation of cones for points in a given polygon by examining a diagram like figure 4 for one of its interior points. More precisely, suppose the sides of $\gamma$ are numbered (how exactly is not important). Then to every point $p$ in one of the shaded polygons we can assign a sequence of numbers of length at most 4 such that if the $k$th symbol of the sequence is $j$ then $T^k(p)_b$ belongs to the side of $\gamma$ labelled by $j$.

**Claim 6.9.** The sequence described above is the same for every point in a given shaded polygon.

**Proof.** Suppose there are two points $p$ and $q$ for which $T^k(p)_b$ and $T^k(q)_b$ lie on different sides for some $k$. Since the polygons are convex $T^k(p)$ and $T^k(q)$ can be joined by a line segment contained in the interior of the polygon. By continuity there will be a point $r$ on this line segment such that $r_b$ will be a corner point. This is a contradiction since lines tangent to $B$ and passing through corners of $\gamma$ form the boundaries of the shaded polygons. □
In what follows we will say that a shaded polygon (or point) has order \( k \) if it belongs to an orbit segment that undergoes \( k \) reflections in corners between successive reflections in the interiors of sides. We will consider polygons of each order in turn.

(1) We first consider order one domain. Here there are two possibilities corresponding to two different kinds of order one domain in the above diagram (see figure 6). In this case, it is enough to look at the sides that the cones lie on to determine inclusion. In the first case (the left diagram in figure 6), the situation is identical to the case of a point touching adjacent hyperbolic segments, and so the cone field is strictly preserved for these points. In the second case the cones \( d_{T(p)b}C(p_b) \) and \( C(p'_b) \) lie on opposite sides so \( u(p'_b) = -d_{T(p)b}u(p_b) \). As before, because the angle between \( d_{T(p)b}u(p_b) \) and \( d_{T(p)b}v(p_b) \) is always acute and the angle between \( u(p'_b) \) and \( v(p'_b) \) is always obtuse (lemma 6.5), \( d_{T(p)b}v(p_b) \in C(p'_b) \) and we have the desired strict inclusion.

(2) For points touching two corners there is only one possible configuration because the restriction on \( a \) in the statement of the theorem guarantees that a point can touch only neighbouring corners of \( B \). Indeed, if there was a point that reflected in opposite corners then one easily checks that \( a \) must be greater than 1/4 which is ruled out by assumption. Hence the only possible configuration is as in figure 7.

The cone is strictly preserved if \( p'_b \) precedes \( p_b \) (the order is given by the clockwise orientation of the square). From the diagram it is clear that this is true when \( T(p) \) and \( p' \) are on the same side of the line \( \ell \), tangent to the midpoint of the hyperbolic segment containing \( p \) and \( p' \). Since \( T(p) \) reflects in a corner it has to lie between the tangents to the hyperbolic arcs that meet at that corner—call them \( \mu_1 \) and \( \mu_2 \), which are, in turn, lines in \( L \) that pass through corners of \( \gamma \). The same reasoning applies to \( p' \) for reverse time since \( T^{2}(p) \) also reflects in a corner. The above condition will be satisfied if the triangle bounded by \( \mu_1, \mu_2 \) and \( \gamma \), containing \( T(p) \) does not intersect \( \ell \). In this case \( T(p) \) is guaranteed to lie on the right side of \( \ell \). One easily checks that this happens exactly when \( 2\sqrt{a+4a < 1} \) or \( a < (3-\sqrt{5})/8 \) as in the statement of the theorem.

(3) Finally, for points touching three corners there is one possible configuration as in figure 8. From the diagram we see that the cone field is strictly preserved if \( p'_b \) precedes...
Observe that reflecting a point in the three consecutive corners of $B$ gives a central symmetry about the fourth corner since the corners of $B$ are vertices of a square. It follows that $p'$ and $T(p)$ are symmetric about the one corner untouched by the orbit segment from $p$ to $p'$ and so that $[p', T(p)]$ contains this corner. On the other hand, $[p', T(p)]$ must intersect the interior of $B$ because otherwise $p'_i$ is a corner and then $p$ has order four and so is not in $D_h$. So the line through $p$ and $T(p)$ must leave $p'$ on the same side as $B$. Noting that $p_i$ and $p'_i$ belong to the interior of the same hyperbolic arc it follows that $p'_i$ precedes $p_i$, if the boundary of $B$ is oriented clockwise, and so $p'_b$ precedes $p_b$.

We have proved that the cone field is strictly preserved at every point of $D_h$ with the exception of points that land on discontinuities of the derivative which form a set of measure zero. Non-vanishing of the Lyapunov exponents now follows.

7. Conclusion

The above result shows that chaotic behaviour is possible for outer billiards. The numerical studies [5] indicate that it coexists with KAM-type behaviour on the rest of the domain and that near-integrable behaviour may still persist in the neighbourhood of infinity. It also appears that this one-parameter family of examples is a member of a much larger class of chaotic outer billiards. This class, we expect, contains many other outer billiards obtained by the secant area construction from polygons. Some numerical studies along these lines for outer billiards obtained from a regular pentagon and a particular non-regular hexagon are also contained in the previous reference. Unfortunately, the current proof does not appear to extend easily to these potential candidates.

Acknowledgment

I am grateful to my advisor Sergei Tabachnikov for setting me on this problem and for the many valuable discussions.
References

[1] Berger M 1987 Geometry (Berlin: Springer)

[2] Boyland P 1996 Dual billiards, twist maps and impact oscillators Nonlinearity 9 1411–38

[3] Bunimovich L On the ergodic properties of nowhere dispersing billiards 1979 Commun. Math. Phys. 65 295–312

[4] Douady R 1982 These de Troisieme Cycle University of Paris 7

[5] Genin D 2005 Regular and Chaotic Dynamics of Outer Billiards PhD Thesis Penn State University

[6] Gutkin E and Katok A 1995 Caustics for inner and outer billiards Commun. Math. Phys. 173 101–133

[7] Gutkin E and Simanyi N 1989 Dual polygonal billiards and necklace dynamics Commun. Math. Phys. 143 431–50

[8] Gutkin E and Tabachnikov S 2005 Complexity of piecewise convex transformations in two dimensions with applications to polygonal billiards (preprint)

[9] Katok A 1987 The growth rate for the number of singular and periodic orbits for a polygonal billiard Commun. Math. Phys. 111 151–60

[10] Kolodziej R 1989 The antibilliard outside a polygon Bull. Pol. Acad. Sci. 37 163–8

[11] Lowenstein H, Kozlovsk K and Vivaldi F 2004 Recursive tiling and geometry of piecewise rotations by π/7 Nonlinearity 17 371–95

[12] Moser J 1973 Stable and Random Motions in Dynamical Systems (Princeton, NJ: Princeton University Press)

[13] Moser J 1978 Is the solar system stable? Math. Intell. 1 65–71

[14] Neumann B 1959 Sharing ham and eggs Iota Manchester University

[15] Shaidenko A and Vivaldi F 1987 Global stability of a class of discontinuous dual billiards Comm. Math. Phys. 110 625–40

[16] Tabachnikov S 1995 Billiards Panor. Synth. No 1 (Societe Mathematique De France) (ISBN: 2856290302)

[17] Tabachnikov S 1996 Asymptotic dynamics of the dual billiard transformation J. Stat. Phys. 83 27–37

[18] Wojtkowski M 1985 Invariant families of cones and Lyapunov exponents Ergod. Th. and Dyn. Syst. 5 145–61

[19] Wojtkowski M 1986 Principles for the design of billiards with non-vanishing Lyapunov exponents Commun. Math. Phys. 105 391–414