DESIGN OF HYPERBOLIC BILLIARDS

Maciej P. Wojtkowski

University of Arizona

September 12, 2006

Abstract. We formulate a general framework for the construction of hyperbolic billiards. Spherical symmetry is exploited for a simple treatment of billiards with spherical caps and soft billiards in higher dimensions. Other examples include the Papenbrock stadium.

1. Introduction

The purpose of this paper is to present a general framework for the construction of hyperbolic billiards, especially with some convex pieces in the boundary, and also "soft" billiards. Most of the examples of hyperbolic billiards that were constructed up to date can be understood in this framework, with the notable exception of systems studied in [W5],[W6].

Billiards are a class of dynamical systems with appealingly simple description. A point particle moves with constant velocity in a box of arbitrary dimension ("the billiard table") and reflects elastically from the boundary (the component of velocity perpendicular to the boundary is reversed and the parallel component is preserved). Mathematically it is a class of hamiltonian systems with collisions defined by symplectic maps on the boundary of the phase space, [W1]. Such systems are also called hybrid systems, being a concatenation of continuous time and discrete time dynamics.

The billiard dynamics defines a one parameter group of maps Φ^t of the phase space which preserve the Lebesgue measure, and are in general only measurable due to discontinuities. The boundaries of the box are made up of pieces: concave, convex and/or flat. Discontinuities occur in particular at the orbits tangent to concave pieces of the boundary of the box. The orbits hitting two adjacent pieces ("corners") have two natural continuations, which is another source of discontinuity. These singularities are not too severe so that the flow has well defined Lyapunov exponents and Pesin structural theory is applicable, [K-S]. A billiard system is called hyperbolic if it has nonzero Lyapunov exponents almost everywhere (or at least on a subset of positive Lebesgue measure). It is called completely hyperbolic if all of its Lyapunov exponents are nonzero almost everywhere, except for one zero exponent in the direction of the flow.

Billiards in smooth strictly convex domains have no singularities, but no such systems are known to be hyperbolic. In dimension 2 Lazutkin showed that near the boundary of such domains the system is near integrable. Applying the KAM theory he proved that for these "grazing orbits"

Key words and phrases. Hyperbolic billiards, Jacobi fields.

Much of the work on this paper was done while the author visited the Institute of Mathematics of the Polish Academy of Sciences in Warsaw and the Institute Henri Poincare in Paris. I am grateful for the warm hospitality and excellent working conditions that I enjoyed in Warsaw and Paris in the Spring of 2005. Supported in part by the NSF grant DMS-0406074. I am also grateful to Paul Wright and the anonymous referees for their valuable remarks.

there is always a family of invariant curves with positive total measure in the phase space, and with zero Lyapunov exponents.

In general billiards exhibit mixed behavior just like other hamiltonian systems, there are invariant tori intertwined with the "chaotic sea". In hyperbolic billiards stable behavior is excluded by the choice of the pieces in the boundary of the box, arbitrary concave pieces and special convex ones, and their particular placement, usually separation. Thus hyperbolicity is achieved by design, as in optical instruments.

Hyperbolicity is the universal mechanism for random behavior in deterministic dynamical systems. Under additional assumptions it leads to ergodicity, mixing, K-property, Bernoulli property, decay of correlations, central limit theorem, and other stochastic properties. Hyperbolic billiards provide a natural class of examples for which these properties were extensively studied. In this article we restrict ourselves to hyperbolicity itself.

The most prominent example of a hyperbolic billiard is the gas of hard spheres. This way of looking at the system was developed in the groundbreaking papers of Sinai, see [Ch-S] for an exhaustive list of references. The excellent collection of papers, [H], contains more up to date information. An important source on hyperbolic billiards is the book by Chernov and Markarian, [Ch-M]. The books by Kozlov and Treschev [K-T], and by Tabachnikov [T] provide broad surveys of billiards from different perspectives.

1. Jacobi fields and monotonicity

The key to understanding hyperbolicity in billiards lies in two essentially equivalent descriptions of infinitesimal families of trajectories. The basic notion is that of a Jacobi field along a billiard trajectory. Let $\gamma(t,u)$ be a family of billiard trajectories, where t is time and u is a parameter, $|u| < \epsilon$, for some $\epsilon > 0$. A Jacobi field $J(t)$ along $\gamma(t) = \gamma(t,0)$ is defined by $J(t) = \frac{\partial \gamma}{\partial u}_{|u=0}$.

Jacobi fields form a finite dimensional vector space that can be naturally identified with the tangent space of the phase space at any point on the trajectory. Jacobi fields contain the same information as the derivatives of the billiard flow $D\Phi^t$. Indeed, if we treat $(J(0), J'(0))$ and $(J(t), J'(t))$ as tangent vectors at $(\gamma(0), \gamma'(0))$ and $(\gamma(t), \gamma'(t))$ respectively then $D\Phi^t\left(J(0), J'(0)\right) = (J(t), J'(t))$. In particular the Lyapunov exponents are the exponential rates of growth of Jacobi fields.

Jacobi fields split naturally into parallel and perpendicular components to the trajectory, each of them a Jacobi field in its own right. The parallel Jacobi field carries the zero Lyapunov exponent. In the following we discuss only the perpendicular Jacobi fields until Sections 6, 7 and 8, where we are forced to consider general Jacobi fields. They form a codimension 1 subspace in the tangent to the unit tangent bundle, i.e., the phase space restricted by the condition that velocity has length one.

Since the billiard trajectories are geodesics of the Euclidean metric the Jacobi fields satisfy between collisions the differential equation

$$J'' = 0, \quad \text{and hence} \quad J(t) = J(0) + tJ'(0). \tag{1}$$

At a collision a Jacobi field undergoes a change by the map

$$J(t_c^+) = \mathcal{R}J(t_c^-) \qquad J'(t_c^+) = \mathcal{R}J'(t_c^-) + \mathcal{P}^*\mathcal{K}\mathcal{P}J(t_c^+), \tag{2}$$

where $J(t_c^-)$ and $J(t_c^+)$ are Jacobi fields immediately before and after collision, $\mathcal{K}$ is the shape operator of the piece of the boundary ($\mathcal{K} = \nabla n$, n is the inside unit normal to the boundary), and $\mathcal{P}$ is the projection along the velocity vector from the hyperplane perpendicular to the orbit to the hyperplane tangent to the boundary. Finally $\mathcal{R}$ is the orthogonal reflection in the hyperplane tangent to the boundary.

To measure the growth/decay of Jacobi fields we introduce a quadratic form in the tangent spaces, or equivalently on Jacobi fields, $\mathcal{Q}(J, J') = \langle J, J' \rangle$. Evaluation of $\mathcal{Q}$ on a Jacobi field is a function of time $\mathcal{Q}(t)$.

Definition 1. *A billiard trajectory $\gamma(t)$ is (strictly) monotone on a Jacobi field J, between its two points $q_0 = \gamma(0)$ and $q_1 = \gamma(t_1)$, or equivalently between time 0 and time t_1, if*

$$\mathcal{Q}(t_1)(>) \geq \mathcal{Q}(0).$$

A billiard trajectory $\gamma(t)$ is called (strictly) monotone between its two points $q_0 = \gamma(0)$ and $q_1 = \gamma(t_1)$ (or between time 0 and time t_1), if it is (strictly) monotone on any nonzero Jacobi field $J(t)$ between q_0 and q_1.

A nonzero Jacobi field is called parabolic between time 0 and time t_1 if $J'(0) = 0$ and $J'(t_1) = 0$.

A billiard trajectory $\gamma(t)$ is called parabolic between its two points $q_0 = \gamma(0)$ and $q_1 = \gamma(t_1)$, if it has a Jacobi field $J(t)$ which is parabolic between time 0 and time t_1 (i.e., $J'(0) = J'(t_1) = 0$). It is called completely parabolic if for every Jacobi field $J(t)$, if $J'(0) = 0$ then also $J'(t_1) = 0$.

Clearly any trajectory which is strictly monotone between q_0 and q_1 cannot be parabolic between q_0 and q_1.

Due to the reversibility of the billiard motion monotonicity is the property of a trajectory in the configuration space with chosen points q_0 and q_1. Indeed, if it holds for the trajectory traversed from q_0 to q_1 then it also holds for the reversed trajectory from q_1 to q_0. This is the subject of the following

Lemma 2. *If a trajectory is (strictly) monotone between its two points q_0 and q_1 then the reversed trajectory is also (strictly) monotone between q_1 and q_0.*

Proof. Let us consider a nonzero Jacobi field J along the orbit $\gamma(t), 0 \leq t \leq T$. The orbit $\tilde{\gamma}(t) = \gamma(T - t)$ is the reversed orbit and $\tilde{J}(t) = J(T - t)$ is a Jacobi field along the orbit $\tilde{\gamma}(t)$. We get further $\tilde{J}'(t) = -J'(T - t)$ and the change of $\mathcal{Q}$ on J along the orbit γ and on $\tilde{J}$ along the orbit $\tilde{\gamma}$ are the same. $\square$

It follows from (1) that if there are no collisions between two points on a trajectory then the trajectory is monotone between the points. Indeed $\mathcal{Q}(t) - \mathcal{Q}(0) = t|J'(0)|^2$. We have further

Lemma 3. *If a trajectory is monotone between two noncollision points q_0 and q_1 then it is also monotone between a point $\tilde{q}_0$ earlier than q_0 ($\tilde{q}_0 < q_0$) and a point $\tilde{q}_1$ later than q_1 ($q_1 < \tilde{q}_1$), provided that there are no collisions between $\tilde{q}_0$ and q_0, and q_1 and $\tilde{q}_1$, respectively. Moreover for two such points $\tilde{q}_0 < q_0$ and $\tilde{q}_1 > q_1$ either the orbit is strictly monotone between $\tilde{q}_0$ and $\tilde{q}_1$ or it is parabolic between the points.*

Proof. The first part of the Lemma is obvious. To prove the second part let us consider a nonzero Jacobi field J. If J' does not vanish at q_0 then the form $\mathcal{Q}$ increases strictly on J between $\tilde{q}_0$ and q_0. Similarly if J' does not vanish at q_1 then the form $\mathcal{Q}$ increases strictly between q_1 and $\tilde{q}_1$. $\square$

It turns out that a sufficiently long "free flight" ensures monotonicity for most trajectories.

Proposition 4. *If a trajectory is not parabolic between two noncollision points q_0 and q_1 then it is strictly monotone between a point $\tilde{q}_0$ sufficiently earlier than q_0 and a point $\tilde{q}_1$ sufficiently later than q_1, provided that there are no collisions between $\tilde{q}_0$ and q_0, and q_1 and $\tilde{q}_1$, respectively.*

If a trajectory is completely parabolic between two noncollision points q_0 and q_1 then it is monotone between a point $\tilde{q}_0$ sufficiently earlier than q_0 and a point $\tilde{q}_1$ sufficiently later than q_1, provided that there are no collisions between $\tilde{q}_0$ and q_0, and q_1 and $\tilde{q}_1$, respectively.

More precisely we extend the segments of the trajectory containing q_0 and q_1 into rays, which allows us to go arbitrarily far in the past and/or arbitrary far in the future without any new collisions.

Note also that in dimension 2 we can apply this result to any trajectory since then any parabolic trajectory is obviously completely parabolic. However for trajectories close to parabolic the necessary interval of free flight may be unbounded.

Proof. Let $q_0 = \gamma(0)$ and $q_1 = \gamma(t_1)$. We are seeking $T > 0$ so large that $\mathcal{Q}(t_1 + T) > \mathcal{Q}(-T)$ for any nonzero Jacobi field J. In other words we want the quadratic form $\mathcal{Q}(t_1 + T) - \mathcal{Q}(-T)$ on perpendicular Jacobi fields to be positive definite. We have

$$\begin{aligned} \mathcal{Q}(t_1+T) - \mathcal{Q}(-T) = \\ \mathcal{Q}(t_1+T) - \mathcal{Q}(t_1) + \mathcal{Q}(t_1) - \mathcal{Q}(0) + \mathcal{Q}(0) - \mathcal{Q}(-T) \\ = T\left(|J'(0)|^2 + |J'(t_1)|^2\right) + \mathcal{Q}(t_1) - \mathcal{Q}(0) \end{aligned} \tag{3}$$

If a trajectory is not parabolic between q_0 and q_1 then $|J'(0)|^2 + |J'(t_1)|^2$ is a positive definite quadratic form on perpendicular Jacobi fields. It follows that for sufficiently large T the quadratic form (3) is also positive definite. This proves the first part of the Proposition. To prove the second part let us recall that due to the symplectic nature of the billiard dynamics for trajectories completely parabolic between time 0 and time t_1 we have

$J'(t_1) = AJ'(0), J(t_1) = A^{*-1}J(0) + BJ'(0)$,

for some linear operators A and B such that A^*B is symmetric, see for example [W1]. Using (3) we get

$$\mathcal{Q}(t_1+T) - \mathcal{Q}(-T) = T\left(|J'(0)|^2 + |AJ'(0)|^2\right) + \langle A^*BJ'(0), J'(0)\rangle .$$

Clearly the quadratic form is positive semidefinite for sufficiently large $T > 0$, and hence the trajectory is monotone, but not strictly monotone. □

By (2) the monotonicity of a trajectory at a collision, i.e., $\mathcal{Q}(t_c^+) \geq \mathcal{Q}(t_c^-)$, is equivalent to the positive semidefiniteness of the shape operator $\mathcal{K} \geq 0$, it holds for concave pieces of the boundary. Billiards with only concave and flat pieces of the boundary are called semidispersing. If $\mathcal{K} > 0$ at a point of collision then by Lemma 3 we have strict monotonicity between a point before the collision and a point after the collision. In semidispersing billiards, where $\mathcal{K} \geq 0, \mathcal{K} \neq 0$, strict monotonicity may still occur after sufficiently many reflections.

Definition 5. *We say that a billiard system is eventually strictly monotone (ESM) on a subset X of positive Lebesgue measure in the phase space, if for almost every trajectory beginning in X there is a return time t_1 to X such that the trajectory is strictly monotone between time 0 and time t_1.*

The role of monotonicity is revealed in the following

Theorem 6 [W1]**.** *If a billiard system is ESM on a subset X of the phase space then for almost every orbit passing through X all Lyapunov exponents are different from zero.*

Theorem 6 is formulated here for billiard systems. However it can be generalized and applied to other systems, not even hamiltonian (see [W2] for precise formulations, references and the history of this idea).

3. Wave fronts and monotonicity

There is a geometric formulation of monotonicity (which historically preceded the one given above). Let us consider a local wavefront, i.e., a local hypersurface $W(0)$ perpendicular to a trajectory $\gamma(t)$ at $t = 0$. Let us consider further all billiard trajectories perpendicular to $W(0)$. The points on these trajectories at time t form a local hypersurface $W(t)$ perpendicular again to the trajectory (warning: in general at exceptional moments of time the wavefront $W(t)$ is singular). Infinitesimally wavefronts are described by the shape operator $U = \nabla n$, where n is the unit normal field. U is a symmetric operator on the hyperplane tangent to the wavefront (and perpendicular to the trajectory $\gamma(t)$. The evolution of infinitesimal wavefronts is given by the formulas

$$\begin{aligned} &U(t) = (tI + U(0)^{-1})^{-1} \qquad \text{without collisions} \\ &U(t_c^+) = \mathcal{R}U(t_c^-)\mathcal{R}^{-1} + \mathcal{P}^*\mathcal{K}\mathcal{P} \qquad \text{at a collision} \end{aligned} \tag{4}$$

It follows that between collisions a wavefront that is initially convex (i.e., diverging, or $U > 0$) will stay convex. Moreover any wavefront after a sufficiently long run without collisions will become convex (after which the normal curvatures of the wavefront will be decreasing). The second part of (4) shows that after a reflection in a strictly concave boundary a convex wavefront becomes strictly convex (and its normal curvatures increase). These properties are equivalent to (strict) monotonicity as formulated in Definition 1. Indeed in the language of Jacobi fields an infinitesimal wavefront represents a linear subspace in the space of perpendicular Jacobi fields (i.e., the tangent space). Moreover it is a Lagrangian subspace with respect to the standard symplectic form. We can follow individual Jacobi fields or whole subspaces of them. It explains the parallel of (1),(2) and (4). The form $\mathcal{Q}$ allows the introduction of positive and negative Jacobi fields and positive and negative Lagrangian subspaces. An infinitesimal convex wavefront represents a positive Lagrangian subspace. Monotonicity is equivalent to the property that for every positive Lagrangian subspace at time 0 its image under the derivative of the flow $D\Phi^t$ is also positive.

It may seem that there is loss of information in formulas (4) compared to (1) and (2). However the symplectic nature of the dynamics makes them actually equivalent, [W1].

4. Design of hyperbolic billiards

In view of (4) it seems that a convex piece in the boundary ($K < 0$) excludes monotonicity. There are two ways around this apparent obstacle to hyperbolicity.

First we could change the quadratic form $\mathcal{Q}$ at the convex boundary and consider monotonicity with respect to the modified form $\mathcal{Q}$. We follow here anther path. We treat convex pieces as "black boxes" and look only at incoming and outgoing trajectories. The first strategy is presented in [W1]. Although the second strategy seems more restrictive, the examples of hyperbolic billiards constructed to date fit the black box scenario with few exceptions, [W5], [W6].

To introduce this approach let us consider a billiard table with flat pieces of the boundary and exactly one convex piece. A trajectory in such a billiard experiences visits to the convex piece separated by arbitrary long sequences of collisions in flat pieces, which do not affect the geometry of a wavefront at all. Hence whatever is the geometry of a wavefront emerging from the curved piece it will become convex and very flat by the time it comes back to the curved piece of the boundary again. Hence it follows, at least heuristically, that we must study the complete passage through the convex piece of the boundary, regarding its effect on convex, and especially flat, wavefronts.

Important difference between convex and concave pieces is that a trajectory has usually several consecutive collisions in the same convex piece, moreover the number of such collisions is unbounded. A finite billiard trajectory is called *complete* if it contains reflections in one and the same piece of the boundary, and it is preceded and followed by reflections in other pieces.

We can now formulate two principles for the design of hyperbolic billiards.

1. *No parabolic trajectories*
Convex pieces may have no complete trajectories which are parabolic.

2. *Separation*
There must be sufficient separation, in space or time, between complete trajectories.

The relevance of these two principles can be seen in Proposition 4. For non-parabolic trajectories if there is enough separation we get strict monotonicity and Theorem 6 is applicable.

Definition 7. *A complete trajectory is (strictly) z-monotone (on a Jacobi field J) for some $z \geq 0$, if it is (strictly) monotone (on the Jacobi field J) between the point at the distance z before the first reflection and the point at the distance z after the last reflection.*

A convex piece of the boundary is (strictly) monotone if almost every complete trajectory is (strictly) z-monotone for some $z \geq 0$. Additionally the piece of the boundary is called finitely (strictly) monotone if the value of z is uniformly bounded for almost all complete trajectories in the piece.

In the language of wavefronts a complete trajectory is z-monotone if every diverging wavefront at a distance at least z from the first reflection becomes diverging at the distance z after the last reflection, or earlier.

Clearly a strictly concave piece is strictly monotone. Every complete trajectory has only one reflection and it is strictly ϵ-monotone for any small ϵ, the property we call 0-*monotone*.

It follows from Theorem 6 that we get a completely hyperbolic billiard if we put together curved strictly monotone pieces of the boundary and some flat pieces, in such a way that for every two consecutive complete trajectories, which are z_1-monotone and z_2-monotone respectively the distance from the last reflection in the first trajectory to the first reflection in the second one is bigger than $z_1 + z_2$. Indeed we can consider the subset X of the phase space containing appropriate midpoints of trajectories leaving one curved piece and hitting another one. We obtain immediately the property ESM on X.

This construction seems unlikely to succeed if there is no uniform bound on the distances z at which complete orbits are monotone, so that no separation of the pieces will be sufficient. However in the case of spherical caps, studied by Bunimovich and Rehacek, [B-R], we find a geometric scenario that works without the uniform bound on the value of z. We will discuss it in Section 6.

5. Hyperbolic billiards in dimension 2

In all of the examples of hyperbolic billiards constructed so far the convex pieces of the boundary have no parabolic complete trajectories.

Checking this property is nontrivial due to the unbounded number of reflections in complete trajectories close to tangency. It was accomplished so far only in integrable, or near integrable examples, with one exception described in the following.

Billiards in dimension 2 are understood best. First of all there is yet another way of describing infinitesimal families of trajectories. Every infinitesimal family of lines in the plane has a point of focusing (in linear approximation), possibly at infinity. This point of focusing contains the same information as the curvature of the infinitesimal wavefront (it is the center of curvature, rather than curvature itself) and it has the advantage that it does not change in free flight. The change in the focusing point after a reflection is described by the familiar mirror equation of the geometric optics

$$-\frac{1}{f_0} + \frac{1}{f_1} = \frac{2}{d}, \tag{5}$$

where f_0, f_1 are the signed distances of the points of focusing to the reflection point, $d = r\cos\theta$, r is the radius of curvature of the boundary piece ($r > 0$ for a strictly convex piece) and θ is the angle of incidence. The mirror equation is just the two dimensional version of (4). We can see that f_1 is a fractional linear function of f_0 as it should be, since the focusing point f is a projective coordinate in the projectivization of the two dimensional space of perpendicular Jacobi fields, [W3]. This fractional linear function gives us a mapping of the line extending the billiard segment, to the next one, the lines become topological circles with the addition of the points at infinity. These circles have natural orientations given by the direction of the billiard trajectory. Fractional linear maps given by the mirror equation preserve this orientation. Indeed the perpendicular Jacobi fields form a plane which has the canonical orientation defined by the symplectic form (the form does not vanish on the plane). This orientation, like the symplectic form, is invariant under the dynamics and it induces an orientation of the projectivization (the circle of focusing points with projective coordinate f).

By Proposition 4 in dimension 2 every convex piece is monotone. However in general it is not finitely monotone, i.e., the value of z may be unbounded. To examine this issue let us consider an incoming trajectory before the first collision and the outgoing trajectory after the last collision. The minimal value of z for which this trajectory is z-monotone can be obtained in terms of the linear fractional map that shows the dependence of the focusing point before the first reflection and the focusing point after last reflection. Let the focusing points before the first reflection and after the last reflection be denoted by f_0 and f_1 respectively. f_0 and f_1 are signed distances to the respective reflection points measured in the direction of the motion. There are two cases, parabolic and non-parabolic trajectory.

If the complete trajectory is parabolic then then for the parabolic Jacobi field $J(t)$ the focusing after the last reflection occurs at infinity and we have $f_1 = af_0 + b$, for some $a > 0$ and b. We obtain that this parabolic trajectory is z-monotone for $z = \max\{0, \frac{b}{a+1}\}$. Indeed it is the minimal $z \geq 0$ such that if $f_0 \leq -z$ then $f_1 \leq z$.

In the case of a nonparabolic complete trajectory let the Jacobi field that is initially focused at infinity ($J'(0) = 0$) be focused after the last reflection at finite $f_1 = c_1$, and the Jacobi field focused at infinity after the last reflection be focused at some finite $f_0 = c_0$ before the first reflection. We get then that

$$f_1 = \frac{c_1 f_0 + b}{f_0 - c_0}, \quad c_0c_1 + b < 0. \tag{6}$$

Let us note that the condition $c_0c_1 + b < 0$ is equivalent to the orientation preservation of the fractional linear map since $\frac{df_1}{df_0} = -\frac{c_0c_1+b}{(f_0-c_0)^2} > 0$.

Lemma 8. *The trajectory is z-monotone for*

$$2z = \max\{0, c_1 - c_0 + \sqrt{(c_0 + c_1)^2 - 4(c_0c_1 + b)}\}.$$

Proof. By direct calculation we obtain two values $2f_\pm = c_0 - c_1 \pm \sqrt{(c_0 + c_1)^2 - 4(c_0c_1 + b)}$ such that if $f_0 = f_\pm$ then $f_1 = -f_\pm$. We have also that $f_- < c_0 < f_+$ and $-f_+ < c_1 < -f_-$.

If $f_- \leq 0$ then the trajectory is z-monotone for $z = -f_-$. If $f_- > 0$ then the trajectory is 0-monotone (in the sense that it is ϵ-monotone for arbitrarily small $\epsilon > 0$).

The problem with the application of these formulas is that in general we cannot get the explicit dependence of f_1 on f_0 for complete trajectories with a large number of reflections. The exception is provided by integrable billiard tables, the disk and the ellipse. Billiard in a disk is integrable due to its rotational symmetry. Let J be a Jacobi field obtained by the rotation of a trajectory. This family of trajectories ("the rotational family") is focused exactly in the middle between two consecutive reflections (that is where J vanishes). It follows further from the mirror equation

(5) that a parallel family of orbits is focused at a distance $\frac{d}{2}$ after the reflection, and any family focusing somewhere between the parallel family and the rotational family will focus at a distance somewhere between $\frac{d}{2}$ and d, not only after the first reflection, but also after arbitrary long sequence of reflections. Hence any complete trajectory in an arc of a circle is z-monotone, where $2z$ is the length of a segment of the trajectory, and by Lemma 3 it is strictly z'-monotone, for any $z' > z$. Two arcs of a circle separated by parallel segments form the stadium of Bunimovich, [B1].

Lazutkin showed that billiards in smooth strictly convex domains are near integrable close to the boundary of the table, [L]. Donnay applied Lazutkin's coordinates to establish that for an arbitrary strictly convex arc the situation near the boundary is similar to that in a circle, i.e., in our language complete near tangent trajectories are z-monotone, where z is of the order of the length of a single segment. This crucial calculation shows that if a strictly convex arc is strictly monotone then any sufficiently small perturbation of this arc is strictly monotone.

In view of Proposition 4 the only obstacle to using a convex arc in the boundary of a hyperbolic billiard is the presence of parabolic orbits. Such orbits are not a problem in themselves since they are still z-monotone, but nearby orbits may have $c_0 << -1$ and $c_1 >> 1$ which by Lemma 8 will result in a large z value. The case when $c_0 >> 1$ and $c_1 << -1$ is safe, resulting in $z = 0$. However no examples of such "safe" parabolic orbits are known to us.

Bunimovich [B3] introduced the concept of absolutely focusing arcs which in our language means that parabolic trajectories are excluded and additionally for every complete trajectory $c_0 < 0$ and $c_1 > 0$. We do not need the last assumptions for our argument, but it is conceivable that our generalization is superficial, i.e., if there are no parabolic orbits then by necessity $c_0 < 0$ and $c_1 > 0$ for all complete trajectories.

The following Proposition reformulates in our language the result of Donnay [D],

Proposition 9. *Any convex arc without complete parabolic orbits is finitely strictly monotone (and hence can be used in designing completely hyperbolic billiards).*

Proof. In view of the Donnay's analysis of near tangent orbits we have to consider only compact families of complete trajectories with the number of reflections not exceeding certain fixed number. Under the assumption of the absence of parabolic complete trajectories, the values of $|c_0|$ and $|c_1|$ are uniformly bounded on these compact families. It follows from Lemma 8 that the z-value is uniformly bounded for these trajectories. □

It was also observed by Donnay [D] and Markarian [M] that for any strictly convex arc, its sufficiently short piece does not have parabolic orbits. Indeed, if the arc is very short, then any complete trajectory is either near tangent or it has only one reflection. No near tangent trajectory is parabolic and the mirror equation (5) does not allow a trajectory with one reflection to be parabolic.

In general checking that parabolic orbits are absent cannot be done by a direct calculation.

An arc satisfying $\frac{d^2r}{ds^2} < 0$, where r is the radius of curvature as a function of the arc length s, is strictly monotone, [W3]. Such an arc is called *convex scattering*. More precisely we have that any complete trajectory in a convex scattering arc is strictly z-monotone with $z = \max d$, maximum of the values of d from the mirror equations (5) for the first and the last segment of the trajectory. This property leads to examples of hyperbolic billiards with one convex piece of the boundary, like the domain bounded by the cardioid. Let us note that the convex scattering property stands out in not being associated with integrability or near integrability.

Integrability of the elliptic billiard allows one to establish finite strict monotonicity of the semi-ellipse with endpoints on the longer axis, [W3]. Donnay, [D], showed that the other semi-ellipse is also finitely strictly monotone provided that $a \leq \sqrt{2}b$, where $a \geq b$ are the semiaxes.

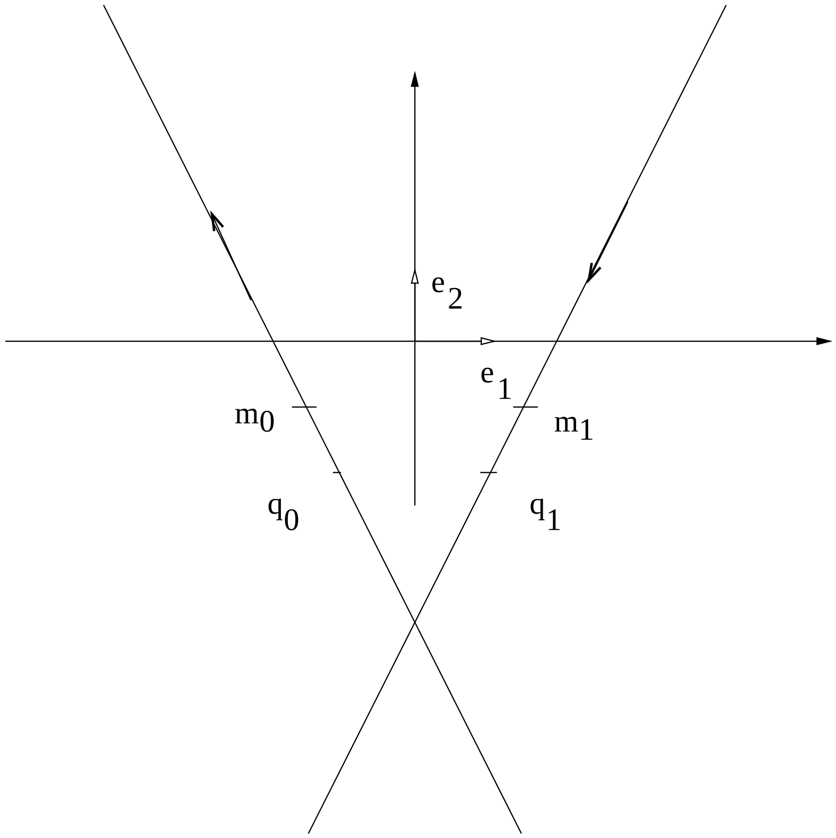

FIGURE 1

6. Systems with local spherical symmetry

Let us consider two segments of the same orbit lying in one plane $\mathbb{R}^2 \subset \mathbb{R}^n$, with mirror symmetry in the plane $\mathbb{R}^2$ reversing the direction of time, Fig. 1.

We further assume that there is the center of symmetry which we use as the origin of the coordinate system and that small rotations around it (of $\mathbb{R}^n$) take the two segments into two segments of another orbit, with the preservation of time. In such a case we say that our system has *local spherical symmetry*, on the orbit in question.

Examples of systems with local spherical symmetry are furnished by billiards with spherical caps, and by soft billiards with spherical scatterers, [B2],[W4],[B-R],[D-L], [B-T].

Let $\gamma(t)$ be the time parameterization of the segments, where $q_0 = \gamma(0)$ belongs to the first segment and $q_1 = \gamma(t_1)$ belongs to the second segment. It follows from the local spherical symmetry that one parameter groups of rotations produce families of orbits. Let $Z \in o(n)$ be an infinitesimal rotation (i.e. Z is an anti-symmetric matrix). We get the family of orbits $\gamma(t,u) = e^{uZ}\gamma(t)$ and the respective Jacobi field $J_Z(t) = Z\gamma(t), J'_Z(t) = Z\gamma'(t)$. This Jacobi field is known to us only on the two segments, but it is enough to check monotonicity between q_0 and q_1. We will call such Jacobi fields *spherical*.

We choose an orthonormal basis e_1, e_2 in the plane of the orbit so that its axis of symmetry has the direction of e_2, Fig. 1. We will be checking monotonicity of our orbit between $\gamma(0) = q_0 = ae_1 + be_2$ and $\gamma(t_1) = q_1 = -ae_1 + be_2$. We have $\gamma'(0) = \sin\alpha e_1 + \cos\alpha e_2$ with $-\frac{\pi}{2} \le \alpha \le \frac{\pi}{2}$, and

$\gamma'(t_1) = \sin\alpha e_1 - \cos\alpha e_2$. It follows that for the spherical Jacobi field J_Z

$$J_Z(0) = aZe_1 + bZe_2, \;\; J_Z'(0) = \sin\alpha Ze_1 + \cos\alpha Ze_2$$
$$J_Z(t_1) = -aZe_1 + bZe_2, \;\; J_Z'(t_1) = \sin\alpha Ze_1 - \cos\alpha Ze_2. \tag{7}$$

We will say that the orbit segments are in general position if their extensions do not contain the origin. Equivalently the segments are in general position if $a\cos\alpha - b\sin\alpha \neq 0$. For orbit segments in general position there are many spherical Jacobi fields. More precisely we have the following

Lemma 10. *If the orbit segments are in general position then spherical Jacobi fields form a linear subspace of dimension $2n-3$. Nonzero spherical Jacobi field are not necessarily perpendicular but none of them is parallel.*

Proof. Let us consider the linear map $Z \to J_Z$. It follows from the condition $a\cos\alpha - b\sin\alpha \neq 0$ that the kernel of the map coincides with antisymmetric matrices such that $Ze_1 = Ze_2 = 0$. Hence the kernel of the map has the dimension $\frac{1}{2}(n-2)(n-3)$ while the space of all matrices Z has the dimension $\frac{1}{2}n(n-1)$. This gives us the dimension of the space of spherical Jacobi fields.

To prove the second part of the Lemma let us consider a spherical Jacobi field such that $J_Z(0)$ and $J_Z'(0)$ are parallel to $\gamma'(0)$, and hence linearly dependent. For any antisymmetric matrix Z either Ze_1, Ze_2 are linearly independent or they are both perpendicular to both e_1 and e_2. It follows that Ze_1 and Ze_2 must be perpendicular to e_1 and e_2, and further that $J_Z(0)$ and $J_Z'(0)$ are both parallel and perpendicular to $\gamma'(0)$. Only the zero Jacobi field can satisfy it. □

We will consider only orbit segments in general position. It is enough for the study of hyperbolicity because the orbits for which this condition fails form a subset of the phase space of dimension n and can be safely ignored. It follows from (7) and the condition $a\cos\alpha - b\sin\alpha \neq 0$ that by an appropriate choice of a skewsymmetric matrix Z we can get arbitrary vectors perpendicular to the plane spanned by e_1 and e_2 as the values of $J(0)$ and $J'(0)$. Let us call such spherical Jacobi fields *transverse*. Transverse Jacobi fields form a linear subspace of the space of spherical Jacobi fields of dimension $2n-4$.

At this stage we need to invoke the symplectic nature of the dynamics. Jacobi fields form the tangent space to the phase space at any point on the orbit. Hence we get an identification of all of these tangent spaces. This identification amounts to the action of the derivative of the flow. The tangent spaces are equipped with the canonical symplectic form and hence the space of Jacobi fields is a linear symplectic space with the canonical symplectic form $\omega(J_1, J_2) = \langle J_1', J_2\rangle - \langle J_1, J_2'\rangle$ ('Wronskian'), where the scalar products are evaluated at any point on the orbit segments (in particular we get the same value independent of the point). It follows from this formula that for any Jacobi field J skeworthogonal to the space of transverse Jacobi fields the values of J and J' at any point are in the plane spanned by e_1 and e_2. We will call such Jacobi fields *planar*. The space of perpendicular Jacobi fields contains the 2 dimensional subspace of planar Jacobi fields.

Further we have the unique splitting of any Jacobi field $J = J_p + J_t$ into a planar, J_p, and a transverse, J_t, Jacobi fields. Moreover it follows from the definition of the form $\mathcal{Q}$ that $\mathcal{Q}(J) = \mathcal{Q}(J_p + J_t) = \mathcal{Q}(J_p) + \mathcal{Q}(J_t)$. Hence, if we establish monotonicity separately for planar and for transverse Jacobi fields then we get monotonicity for all Jacobi fields.

Note that we have used above the symplectic formalism to obtain the splitting into transverse and planar Jacobi fields. In the two classes of examples with spherical symmetry we can get it geometrically from an additional symmetry.

Monotonicity between the symmetric points q_0 and q_1 for spherical Jacobi fields can be analyzed by direct calculation.

$$\mathcal{Q}(t_1) - \mathcal{Q}(0) = \langle J_Z(t_1), J_Z'(t_1)\rangle - \langle J_Z(0), J_Z'(0)\rangle = -2a\sin\alpha\,\langle Ze_1, Ze_1\rangle - 2b\cos\alpha\,\langle Ze_2, Ze_2\rangle\,. \tag{8}$$

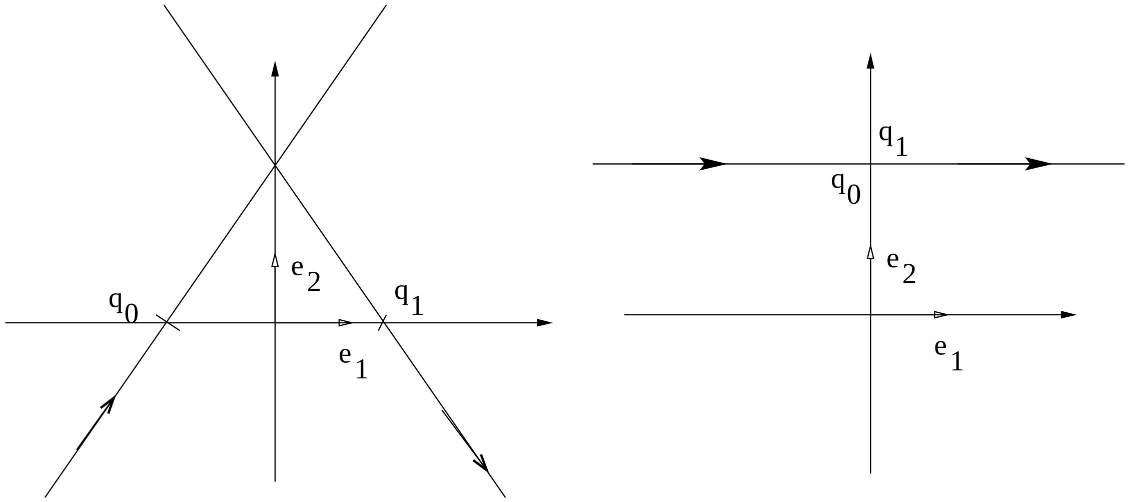

FIGURE 2. Configurations $\mathcal{A}$ (left) and $\mathcal{A}_s$ (right)

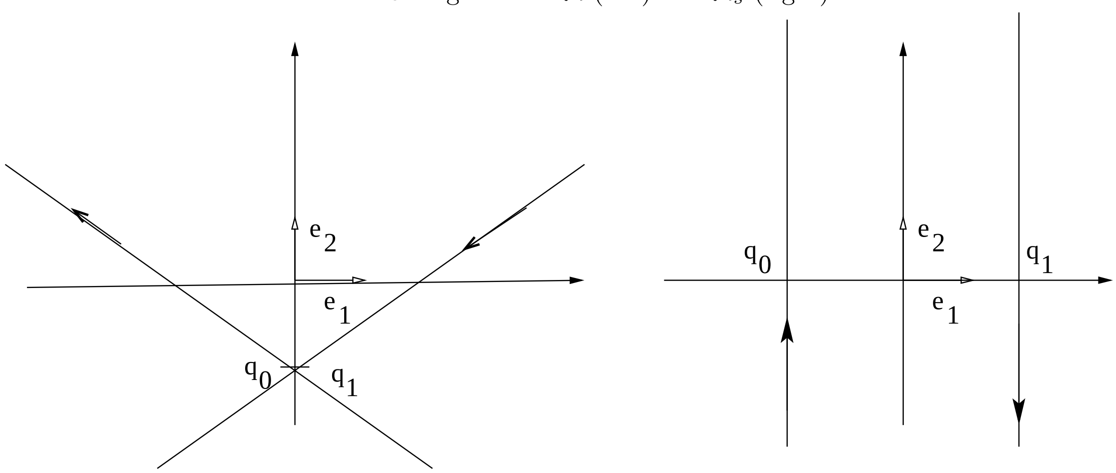

FIGURE 3. Configurations $\mathcal{B}$ (left) and $\mathcal{B}_s$ (right)

Hence we get monotonicity if and only if $b\cos\alpha \leq 0$ and $a\sin\alpha \leq 0$.

Let us assume that there is monotonicity on all spherical Jacobi fields between the points $q_0 = \gamma(0)$ and $q_1 = \gamma(t_1)$ and that monotonicity fails on some of these fields between $\gamma(\epsilon)$ and $\gamma(t_1 - \epsilon)$ for arbitrarily small $\epsilon > 0$. In the same way as in the formula (3) we obtain from (8)

$$\mathcal{Q}(t_1 - \epsilon) - \mathcal{Q}(\epsilon) = (-2a\sin\alpha - 2\epsilon\sin^2\alpha)\,\langle Ze_1, Ze_1\rangle + (-2b\cos\alpha - 2\epsilon\cos^2\alpha)\,\langle Ze_2, Ze_2\rangle\,.$$

Our assumptions lead to the conditions $b = 0, a\sin\alpha < 0$ or $a = 0, b\cos\alpha < 0$.

In this way we arrive at two generic configurations, the configuration $\mathcal{A}$ where $b = 0, a\sin\alpha < 0$, and the configuration $\mathcal{B}$ where $a = 0, b\cos\alpha < 0$. We also have two singular configurations, the configuration $\mathcal{A}_s$ with $\alpha = \pm\frac{\pi}{2}, a = 0$. and the configuration $\mathcal{B}_s$ with $\alpha = 0, b = 0$, Fig. 2 and Fig. 3.

In all cases we get monotonicity between q_0 and q_1 on all spherical Jacobi fields, but the points q_0 and q_1 are positioned differently in different configurations. In configuration $\mathcal{A}$ and $\mathcal{B}_s$ the points q_0 and q_1 lie on the e_1 axis and in configuration $\mathcal{B}$ and $\mathcal{A}_s$ they coincide geometrically with one point on the axis of symmetry, i.e., the e_2 axis. In the singular configuration $\mathcal{B}_s$ both segments are vertical and in the singular configuration $\mathcal{A}_s$ both segments lie on the same horizontal line. In any of the configurations the points are optimal in the sense that monotonicity fails for points past q_0 and before q_1.

The difference between a configuration $\mathcal{A}$ and a configuration $\mathcal{B}$ is in the location of the point of intersection of the line extensions of the segments. This point lies on the axis of symmetry, *above* the origin for a configuration $\mathcal{A}$, and *below* the origin for a configuration $\mathcal{B}$. (The terms 'below' and 'above' seem arbitrary. To remove this ambiguity we note that the two orbit segments are ordered by time. The first segment in a generic configuration allows us to orient canonically the line of symmetry, which was hidden earlier in the condition that $-\frac{\pi}{2} \leq \alpha \leq \frac{\pi}{2}$.)

We have thus established that monotonicity on spherical Jacobi fields depends only on the geometry of the incoming and outgoing segment and it is not affected by the dynamics.

The generic configurations have no parabolic spherical Jacobi fields but the singular configurations do. However even for the singular configurations we get monotonicity between appropriate points, i.e., the conclusions of Proposition 4 hold, even though we do not have a completely parabolic orbit. It follows from the splitting of an arbitrary perpendicular Jacobi field into a transverse field and a planar one. All transverse fields are spherical and hence are covered by the above analysis. The planar fields form a two dimensional invariant subspace and hence the proof of the completely parabolic part of Proposition 4 applies to them. Having monotonicity separately for transverse and planar fields is equivalent to monotonicity.

Let us analyze monotonicity on planar Jacobi fields in more detail. We will compare the Jacobi fields at the points $\gamma(0) = m_0$ and $\gamma(t_1) = m_1$ which are closest to the origin in the line extensions of the respective orbit segments, Fig. 1. (If the initial segments are too short to contain the points their status is somewhat abstract, they may or may not be actual orbit points. However it will not effect our analysis.)

Lemma 11. *The fractional linear representation of the dynamics* (6) *on the planar perpendicular Jacobi fields between m_0 and m_1 has the form*

$$f_1 = \frac{-cf_0}{f_0 - c}$$

where c is the unique value for which there is a planar perpendicular Jacobi field J with $J'(0) = 0$ and $J(t_1 - c) = 0$ (or a field with $J(c) = 0$ and $J'(t_1) = 0$).

Monotonicity holds for the planar Jacobi fields between $\gamma(-z)$ and $\gamma(t_1 + z)$ for $z = |c| - c$. In the limit case of $c \to \infty$ we get $f_1 = f_0$ and then there is monotonicity for $z = 0$.

Proof. The space of planar perpendicular Jacobi fields is 2 dimensional. The spherical planar Jacobi field J_Z generated by the infinitesimal rotation Z with $Ze_1 = e_2, Ze_2 = -e_1$ is not perpendicular but it has the nonzero perpendicular component which we will denote by J_r. By the choice of the points m_0 and m_1 we have that $J_r(0) = 0$ and $J_r(t_1) = 0$. Moreover introducing compatible orthonormal frames $v_0 = \gamma'(0), v_0^\perp$ and $v_1 = \gamma'(0), v_1^\perp$ at m_0 and m_1 respectively, we can calculate that $\langle J_r'(0), v_0^\perp \rangle = \langle J_Z'(0), v_0^\perp \rangle = 1$ and $\langle J_r'(t_1), v_1^\perp \rangle = \langle J_Z'(t_1), v_1^\perp \rangle = 1$. Now we use $\left(\langle J(0), v_0^\perp \rangle, \langle J'(0), v_0^\perp \rangle\right)$ and $\left(\langle J(t_1), v_1^\perp \rangle, \langle J'(t_1), v_1^\perp \rangle\right)$ as coordinates in the 2 dimensional space of planar perpendicular Jacobi fields, at m_0 and m_1 respectively. In these coordinates J_r is the second basic vector both at m_0 and at m_1. Hence the dynamics between m_0 and m_1 is described in these coordinates by the matrix $\begin{pmatrix} 1 & 0 \\ * & 1 \end{pmatrix}$. Since the focusing distance $f_i = -\frac{\langle J, v_i^\perp \rangle}{\langle J', v_i^\perp \rangle}, i = 0, 1,$ we obtain the Lemma from Lemma 8 by direct calculation. □

Note that the effort in the proof above is to show that $c_0 = -c_1 = c$ in (6). We get it from spherical symmetry alone. In the two classes of examples it follows easily from the additional reversible symmetry.

We have thus established that monotonicity of a complete trajectory in a system with spherical symmetry depends only on the geometry of the incoming and outgoing segments and the value of c from Lemma 11, which is the only information we need to extract from the dynamics.

In the case of spherical caps of radius R any complete trajectory lies in a plane passing through the center. Moreover the planar Jacobi fields are just the Jacobi fields of the trajectory in the billiard in the disk of radius R. It was observed in Section 5 that in such a case c is always positive and hence $z = 0$.

The analysis of monotonicity in the case of spherical caps is thus complete and can be summarized in the following Proposition which was essentially stated in [W4].

Proposition 12. *Any complete trajectory, in general position, in a spherical cap is monotone between* $\min(q_0, m_0)$ *and* $\max(q_1, m_1)$. *(The* min *and* max *are understood in the sense of the temporal ordering of the trajectory). Moreover only the trajectories in singular configurations are parabolic.*

Proof. It remains to analyze parabolic trajectories, i.e., we are looking for a perpendicular Jacobi field such that $J'(0) = 0, J'(t_1) = 0$. We know that there are no such nonzero planar Jacobi fields. It remains to check the transverse (and hence spherical) Jacobi fields. It follows from (7) that if $J_Z'(0) = \sin\alpha Z e_1 + \cos\alpha Z e_2 = 0$ and $J_Z'(t_1) = \sin\alpha Z e_1 - \cos\alpha Z e_2 = 0$ then either $\alpha = 0$ and the trajectory is in the configuration $\mathcal{B}_s$, or $\alpha = \pm\frac{\pi}{2}$ and the trajectory is in configuration $\mathcal{A}_s$.

We can now apply this analysis to specific examples of billiards with spherical caps, and to soft billiards with spherical scatterers.

The first construction of a three dimensional hyperbolic billiard with spherical caps was obtained in [B-R]. We are in a position to recover easily this construction, to see what the obstacles are and how to overcome them. We want to attach spherical caps to a box. We need separation of the caps so that it takes a long time from when an orbit leaves a cup until it reaches another one (or the same one after reflecting in flat pieces). By similarity considerations instead of separating the cups we may fix the rectangular box, and decrease the radius of the sphere.

The first observation is that configurations $\mathcal{B}_s$ are disastrous, because if they are present then in the same plane we will also have configurations $\mathcal{B}$ with the point q_0 arbitrarily far away. By elementary geometry we get

Proposition 13. *If the angle at which a piece S of a sphere is seen from the center is less than $\frac{\pi}{2}$ then all complete trajectories in S are in configuration $\mathcal{A}$, in particular there are no trajectories in configuration $\mathcal{B}_s$*

We will call such pieces *small spherical caps.*

One may get the impression that configurations $\mathcal{A}_s$ may pose similar difficulty because the points q_0 and q_1 may go to infinity. Indeed if we consider a small spherical cap and a plane of our complete orbit that cuts the edge of the cap then our complete orbit may have the points q_0 and q_1 far away. What saves the construction is that the points must stay on the line through the center, and hence they have a bounded displacement in one direction. So now the prescription for the design of the billiard with spherical caps is to place the small caps only at the bottom and the top of the box. Such a billiard is equivalent to the billiard between two parallel hyperplanes (the top and the bottom) with small spherical caps attached. It is clear that if the hyperplanes are sufficiently far apart then the configurations $\mathcal{A}$ do not pose any difficulty. The exact separation is such that the horizontal hyperplanes through the centers of the spheres of the top spherical caps should be above those for the bottom spherical caps. Clearly more complicated designs can also be produced. One finds several of them in [B-R].

7. Soft billiards

The analysis in Section 6 can be readily applied to soft billiards. These are systems with a point particle moving in a rectangular box, or a torus, with spherical scatterers. However the point particle does not collide elastically with the scatterer, but enters into it and is subjected to a field of force with a spherically symmetric potential. In the 2 dimensional case, after the work of Knauf, [K1],[K2] Donnay and Liverani, [D-L], gave general conditions on the potential that guarantee, in our present language, that all complete trajectories are z-monotone with uniformly bounded z. A complete trajectory through a scatterer is the piece of a trajectory from entering a scatterer to leaving it. This allowed them to construct a variety of completely hyperbolic soft billiards,. The case of higher dimensions remained open for 15 years. Recently Balint and Toth, [B-T], obtained additional conditions on the potential that guarantee complete hyperbolicity in arbitrary dimension. Our condition that no complete trajectory is parabolic is fully equivalent to those of [B-T]. Moreover our approach results in fairly explicit conditions on the required separation, while such conditions are absent both from [D-L] and [B-T].

The orbit of the point particle inside a scatterer is not in general a straight segment. However we restrict our attention to the incoming and outgoing segments of our trajectory which we denote by $\gamma(t)$, with $\gamma(0)$ being a point before the entrance into the scatterer and $\gamma(t_1)$ a point after the exit. For a family of trajectories $\gamma(t,u)$ we consider the Jacobi field $J(t) = \frac{\partial \gamma}{\partial u}_{|u=0}$. For Jacobi fields which are not spherical we cannot claim that if $J(0)$ is perpendicular to the trajectory then $J(t_1)$ is also perpendicular. However since $J'(t)$ is perpendicular to the trajectory outside of the scatterer (because the point particle has unit velocity there), then the values of $\mathcal{Q}(0)$ and $\mathcal{Q}(t_1)$ depend only on the perpendicular component of $J(t)$. We can then consider these perpendicular components in place of perpendicular fields and the analysis of Section 6 is perfectly valid. (What happens here is that an invariant codimension one subspace in the tangent bundle of the phase space is not a priori available and we have to work with a quotient space rather than a subspace. Perpendicular components of Jacobi fields form the quotient space, see [W1],[W2].)

Definition 14. *The halo of a scatterer in a soft billiard is a closed concentric ball of minimal radius, containing the scatterer and such that almost any complete trajectory through the scatterer is strictly monotone between two points outside of the ball.*

Our goal is to establish the existence of the halo for a given scatterer, and to determine its radius. By Theorem 6 if each scatterer in a soft billiard has a halo and the halos are mutually disjoint then the soft billiard is completely hyperbolic.

In dimension 3 and above if the spherical potential $V = V(r)$ is continuous (i.e. in particular it vanishes at the boundary) and attractive $V'(r) > 0$ then there is no halo. Indeed let us consider a straight line tangent to the scatterer. Perturbing it we obtain a complete trajectory “grazing” the scatterer. By necessity it is in configuration $\mathcal{A}$ with the points q_0 and q_1 at large distance from the center. Since this distance goes to infinity as the trajectory approaches the tangent line and the points must belong to the halo we conclude that there is no halo for our scatterer. Hence scatterers with continuous attractive potentials are not allowed in the design of a hyperbolic soft billiard in dimension ≥ 3. This was already observed by Balint and Toth, [B-T].

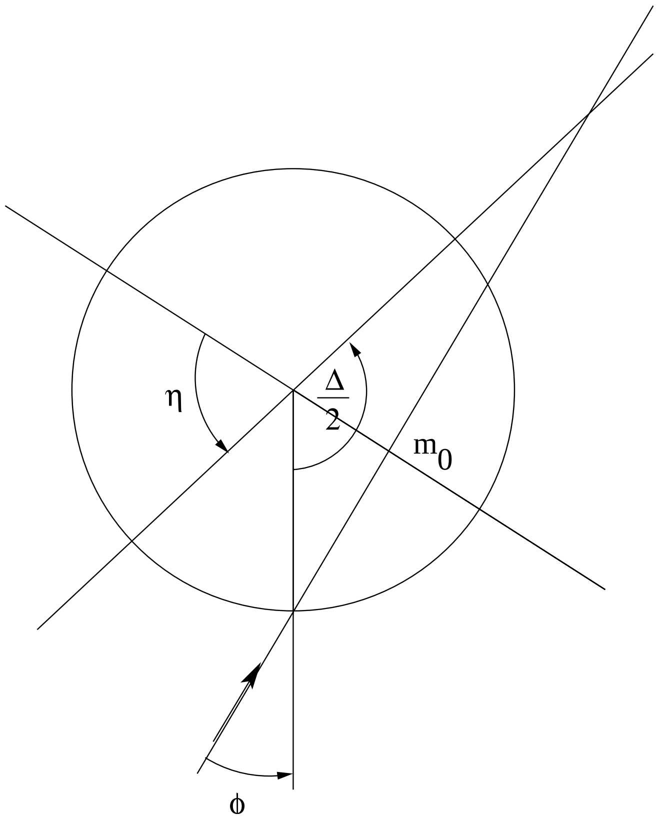

FIGURE 4

The passage through a scatterer is completely described by the rotation angle $\Delta = \Delta(\varphi), 0 \leq \varphi \leq \frac{\pi}{2}$, [B-T]. For a given angle of incidence φ the angle Δ is the angular difference between the entrance and the exit points on the complete trajectory that enters the scatterer with the incidence angle φ. With a fixed orientation of the circle the value of Δ differs by the sign when we switch the incoming and outgoing lines. For simplicity we consider only the case depicted in Fig. 4, with the counterclockwise orientation of the boundary of the scatterer.

It is convenient to introduce the angle $-\frac{\pi}{2} \leq \eta \leq \frac{\pi}{2}$ between the perpendicular axis of symmetry of the line of the incoming segment and the axis of symmetry of the configuration, Fig. 4.

In configuration $\mathcal{A}$ we have $0 < \eta < \frac{\pi}{2}$ and in configuration $\mathcal{B}$ we have $-\frac{\pi}{2} < \eta < 0$. Moreover $\Delta = 2\eta - 2\varphi + \pi$.

To find the radius of the halo of a scatterer we need to find the distance of the points q_0, q_1 to the center of symmetry for any complete trajectory passing through the scatterer. By simple geometric considerations we obtain that this distance is equal to

$$R\frac{\sin\varphi}{\sin\eta} \quad \text{in configuration } \mathcal{A} \quad \text{and} \quad R\frac{\sin\varphi}{\cos\eta}, \quad \text{in configuration } \mathcal{B} \tag{9}$$

where R denotes the radius of the scatterer, and φ is the angle of incidence for our complete trajectory. Hence the point q_0 is outside of the scatterer when $\eta < \varphi$ in configuration $\mathcal{A}$, and when $\eta < -\frac{\pi}{2} + \varphi$ in configuration $\mathcal{B}$. These formulas allow the direct calculation of the halo of a scatterer in examples. It will be large if for some configurations η is small and positive or close to $-\frac{\pi}{2}$. It is guaranteed to be finite if there are no singular configurations ($\eta = 0$ or $\eta = \pm\frac{\pi}{2}$).

There is also contribution into the halo from the planar Jacobi fields. More specifically we need to calculate the constant c in Lemma 11. It is sufficient to obtain one additional planar Jacobi field, not focused at m_0 (as in the rotational field). For that purpose let us consider the family of trajectories $\gamma(t,\varphi)$ entering a scatterer at one point $\gamma(0,\varphi)=(R,0), \gamma'(0,\varphi)=-cos\varphi e_1+\sin\varphi e_2$. At the exit time $t_1=t_1(\varphi)$ we get

$$\gamma(t_1(\varphi),\varphi)=(R\cos\Delta, R\sin\Delta)\,, \quad \gamma'(t_1(\varphi),\varphi)=\cos(\Delta+\varphi)e_1+\sin(\Delta+\varphi)e_2.$$

Hence we get the Jacobi field J

$$\begin{aligned} &J(0)=0, && J(t_1)=R\Delta'\left(-\sin\Delta e_1+\cos\Delta e_2\right),\\ &J'(0)=\sin\varphi e_1+\cos\varphi e_2, && J'(t_1)=(\Delta'+1)\left(-\sin(\Delta+\varphi)e_1+\cos(\Delta+\varphi)e_2\right). \end{aligned}$$

By direct calculation we obtain now $c=-\frac{R\cos\varphi}{\Delta'+2}$. Hence by Lemma 11 if $\Delta'+2<0$ we get $z=0$ and there is no contribution from planar Jacobi fields to the scatterer's halo. If however $\Delta'+2>0$ then $z=-2c$ which translates to the halo of radius

$$h=R\sqrt{p^2\cos^2\varphi+\sin^2\varphi}, \quad \text{where} \quad p=\left(\frac{\Delta'}{2}+1\right)^{-1}. \tag{10}$$

The application of these formulas, in obtaining explicit separation of scatterers for hyperbolicity, hinges on the representation of the rotation function Δ in terms of the potential, which is somewhat cumbersome. We have nothing new to add on this subject compared to the papers [D-L] and [B-T] where the reader can find a detailed discussion. We will consider here only the simple case of a constant potential $V=V_0<E=\frac{1}{2}$. In the two dimensional case it was studied by Baldwin, [Ba], and Knauf, [K2], who arrived at sharp conditions for complete hyperbolicity. It was shown in [B-T] that there is always sufficient separation of scatterers that will guarantee complete hyperbolicity also in the multidimensional case, without providing specific bounds. It turns out that the two dimensional conditions are also sufficient in higher dimension.

Soft billiards with constant potential are systems where the crossing of scatterers is governed by a version of the law of refraction. One needs to distinguish the case of the positive potential $0<2V_0<1$ and the negative potential $V_0<0$. We have

$$\frac{\Delta}{2}=\arccos\left(\frac{\sin\varphi}{\nu}\right),\ \nu=\sqrt{1-2V_0},$$

where in the case of a positive potential ($\nu<1$) the formula is valid for $0\le\varphi\le\varphi_0$ and φ_0 is such that $\sin\varphi_0=\nu$, and in the case of a negative potential ($\nu>1$) the formula is valid for all $0\le\varphi\le\frac{\pi}{2}$. It follows immediately that

$$\frac{\Delta'(\varphi)}{2}=-\frac{1}{\sqrt{1+\frac{\nu^2-1}{\cos^2\varphi}}}$$

Hence the derivative is a decreasing function in the case of a positive potential ($\nu<1$) and an increasing function in the case of a negative potential. It follows that in the case of positive potential $\frac{\Delta'(\varphi)}{2}\le-\frac{1}{\nu}<-1$, and there is no contribution into the halo from the planar Jacobi fields. In the case of a negative potential ($\nu>1$) $\frac{\Delta'(\varphi)}{2}\ge-\frac{1}{\nu}>-1$ and we get, using (10), that the halo radius $h=\frac{\nu R}{\nu-1}$ (the minimal value of h is assumed at $\varphi=0$ because p is a decreasing function of φ).

Further in the positive case we have only configurations $\mathcal{B}$ with $\eta < \varphi$ so that there is no contribution from them into the halo.

In the negative case we have only configurations $\mathcal{A}$ and using (9) we arrive readily at the same halo radius $h = \frac{\nu R}{\nu - 1}$ (it is again assumed at $\varphi = 0$). The explanation for this coincidence is that the halo is determined by trajectories in the limit of the incidence angle $\varphi \to 0$. In that limit the distinction between planar and transverse Jacobi fields is lost.

To summarize, no separation of scatterers is necessary for hyperbolicity in the case of a positive potential and in the case of a negative potential the scatterers should have non intersecting halos with the radius $h = \frac{\nu R}{\nu - 1}$, $\nu = \sqrt{1 - 2V_0}$. Baldwin, [Ba], showed that with the violation of these conditions one can construct systems with elliptic periodic orbits. The same can be claimed in higher dimensional systems. Hence our conditions are sharp.

8. Twisted cartesian products

We will describe here a construction of higher dimensional hyperbolic systems which generalizes the Papenbrock stadium, [P], and can be understood in the language of monotonicity as developed in this paper.

Let us consider two billiard systems, system 1 and system 2, and their cartesian product. Given monotone trajectories γ_1 and γ_2 in systems 1 and 2 respectively, we address monotonicity of the trajectory $(\gamma_1(t), \gamma_2(t))$ in the cartesian product.

We are faced with the basic difficulty that the moments of time between which there is monotonicity may be different for γ_1 and γ_2. This difficulty disappears if one of the systems has all trajectories monotone between any two points. The examples of such systems are semidispersing billiards and closely related geodesic flows on manifolds of nonpositive sectional curvature. The simplest example is the motion of a point particle in a segment. We will call such systems *universally monotone*.

Another new element in our construction is monotonicity in the full phase space, so far we have discussed montonicity of systems on one energy level. This restriction was somewhat hidden in the fact that all our Jacobi fields J satisfied $\langle J', \gamma' \rangle = 0$. When we allow all energy levels we have more Jacobi fields and a trajectory could fail to be monotone on some of the additional fields. Monotonicity on all Jacobi fields will be called *ambient*.

In the cartesian product the kinetic energy is split arbitrarily between system 1 and system 2. In other words γ_1 may be traversed fast while γ_2 is traversed slowly. In a pure cartesian product it is not an issue because both kinetic energies are first integrals of motion. However we are going to modify the cartesian product to obtain a hyperbolic system and such modifications are bound to destroy the first integrals, only the total kinetic energy remains constant. We need to consider each of the systems in all of the phase space and check for the ambient monotonicity. More precisely we need to allow more Jacobi fields by considering families of trajectories $\gamma(t, u)$ (cf., Section 2) in which $||\frac{\partial}{\partial t}\gamma(t, u)||$ depends on u. It turns out that for billiard systems (and geodesic flows) ambient monotonicity follows automatically from monotonicity.

Indeed in such systems the same trajectory can be traversed at different speeds. Hence in particular for any trajectory $\gamma(t)$ there is a constant $a > 0$ such that $\gamma(as)$ is its arc length parameterization. In other words a trajectory on an arbitrary energy level is a reparametrization of a trajectory with unit velocity. We have

$$\gamma(t, u) = \gamma(a(u)s, u), \qquad Y(t) = \frac{\partial}{\partial u}\gamma(t, u)_{|u=0}, \quad J(s) = \frac{\partial}{\partial u}\gamma(a(u)s, u)_{|u=0},$$

$$J(s) = Y(t) + a'sw, \qquad \frac{d}{dt}Y = \frac{1}{a}\frac{d}{ds}J - \frac{a'}{a}w, \quad \text{where} \quad a = a(0),\ a' = a'(0),\ w = \frac{\partial}{\partial t}\gamma(t, 0).$$

Adopting the convention that $'$ is $\frac{d}{dt}, \frac{d}{ds}$ or $\frac{d}{du}$, as appropriate, and using $\langle J', w\rangle = 0$, we get

$$\mathcal{Q}(Y) = \langle Y, Y'\rangle = \frac{1}{a}\langle J(s), J'\rangle - \frac{a'}{a}\langle J(s), w\rangle + \frac{(a')^2}{a^3}s = \frac{1}{a}\mathcal{Q}(J) - \frac{a'}{a}\langle J(s), w\rangle + \frac{(a')^2}{a^3}s.$$

Once we remember that $\langle J(s), w\rangle$ does not change along a trajectory (because of the invariant split of Jacobi fields into perpendicular and parallel Jacobi fields) we conclude that

$$\mathcal{Q}(Y)(t_1) - \mathcal{Q}(Y)(t_0) = \frac{1}{a}\left(\mathcal{Q}(J)(s_1) - \mathcal{Q}(J)(s_0)\right) + \frac{(a')^2}{a^3}(s_1 - s_0) \quad \text{where} \quad t_i = as_i, i = 0, 1. \tag{11}$$

We summarize the consequences of (11) into the following

Proposition 15. *If a trajectory traversed with speed one is monotone between two points then the same trajectory traversed at an arbitrary speed satisfies ambient monotonicity between the points.*

If a trajectory is strictly monotone in the restricted phase space between two points then in the full phase space the only Jacobi fields Y on which $\mathcal{Q}$ is not increased are parallel to the velocity and satisfy $Y' = 0$.

Proof. The first part follows immediately from (11). To prove the second part let us observe that by (11) if $\gamma(t)$ is monotone between $\gamma(t_0)$ and $\gamma(t_1)$ and there is no increase of $\mathcal{Q}$ on a Jacobi field Y then $a' = 0$, which means that the Jacobi field may be obtained from a family of trajectories in one energy level. Now since the trajectory is assumed to be strictly monotone, then the Jacobi field Y must be parallel, i.e., $Y = const\gamma', Y' = 0$. □

We are ready to proceed with the construction. We consider a euclidean space with coordinates $(x_0, x_1, \dots, x_k) = (x_0, x)$. Our system 1 is a billiard (or a geodesic flow) in a domain D in the upper halfspace $\{x_0 > 0\}$ with some boundary D_0 at $\{x_0 = 0\}$. We will remove the boundary D_0 and allow trajectories to enter freely into the lower halfspace $\{x_0 < 0\}$. We assume that in the system 1 any trajectory is strictly monotone between two consecutive visits to D_0. Our system 2 is a universally monotone system in the configuration space E which is isometric to D_0. It is clear that the product system has all trajectories monotone between consecutive visits of the first component to D_0, regardless of where the second component is in E. We will now examine the Jacobi fields on which our trajectory is parabolic.

Let the euclidean coordinates in E be denoted by $y = (y_1, \dots, y_k)$ and let the coordinate identity map $y = x$ furnish the isometry from D_0 to E. Let $\gamma(t) = (\gamma_1(t), \gamma_2(t)) \in D \times E$ be a trajectory of the cartesian product of our systems an let $\gamma(t_i) \in D_0 \times E, i = 0, 1$, be two consecutive visits to $D_0 \times E$. We consider a family of trajectories $\gamma(t, u) = (\gamma_1(t, u), \gamma_2(t, u)) = (x_0(t, u), x(t, u), y(t, u))$ including $\gamma(t) = \gamma(t, 0)$ and generating the Jacobi field $Y = \left(\frac{\partial}{\partial u}\gamma_1(t, 0), \frac{\partial}{\partial u}\gamma_2(t, 0)\right)$.

Proposition 15 leads immediately to the following

Lemma 16. *If there is no increase in Q on a Jacobi field Y between t_0 and t_1 then $Y' = 0$ and $\frac{\partial}{\partial u}\gamma_1(t, 0) = const\frac{\partial}{\partial t}\gamma_1(t, 0)$.*

To finish the construction of our system we consider another system in the domain $\tilde{D}$ in the lower halfspace $\{x_0 < 0\}$ such that the part of the boundary of $\tilde{D}$ at $\{x_0 = 0\}$ is the same D_0, and with strict monotonicity between consecutive visits to D_0. The simplest example would be the reflection of our system 1 into the lower halfspace $\{x_0 < 0\}$. We now glue the domains $D \times E$ and $\tilde{D} \times E$ along the common boundary $D_0 \times E$ by the isometric map $G(x_0, x, y) = (x_0, -y, x)$. When a trajectory in $D \times E$ reaches $\{x_0 = 0\}$ then it is continued into $\tilde{D} \times E$ after the change of position and velocity by the map G. We will call such a system *a twisted cartesian product*.

Theorem 17. *A trajectory in the twisted cartesian product is strictly monotone between every second visit to $\{x_0 = 0\}$. The twisted cartesian product is completely hyperbolic.*

Proof. We can see that the glueing map G preserves the form $\mathcal{Q}$. Hence in the twisted cartesian product trajectories are monotone between visits to $\{x_0 = 0\}$ We need to examine the Jacobi fields on which there is no increase in the value of the form $\mathcal{Q}$ between a first visit to $\{x_0 = 0\}$ and the third.

Let $Y(t) = (Y_0(t), Y_1(t), Y_2(t))$ be such a field with the second visit to $\{x_0 = 0\}$ at $t = 0$. By Lemma 16 if there is no increase of $\mathcal{Q}$ on Y between the first and the second visit then $Y'(-0) = 0$ and $(Y_0(-0), Y_1(-0)) = c_1\gamma_1'(-0)$, where the values at $-0(+0)$ denote the limits at 0 over negative (positive) t.

Further we get that if there is no increase of $\mathcal{Q}$ on the Jacobi field Y between the second and third visit then $(Y_0(+0), Y_1(+0)) = c_2\gamma_1'(+0)$. Taking into account the gluing map G applied at $t = 0$ we have $(Y_0(+0), Y_1(+0), Y_2(+0)) = (Y_0(-0), -Y_2(-0), Y_1(-0))$ and we can conclude that $c_1 = c_2 = c$ and $Y(t) = c\gamma'(t)$ both for $t < 0$ and $t > 0$, i.e., Y is a parallel Jacobi field in the product system. That means that our trajectory is actually strictly monotone between the first and the third visit,

It follows that our twisted cartesian product on one energy level is eventually strictly monotone and hence completely hyperbolic. Indeed if we consider the Poincare section of our flow $\{x_0 = 0, \frac{dx_0}{dt} > 0\}$ we see that consecutive visits to the section are separated by exactly one more visit to $\{x_0 = 0\}$. □

Let us consider specific examples of systems with the required properties and their twisted cartesian products.

Example 1. Let the system 1 be the Sinai billiard with one convex scatterer in a square D with D_0 being one of the sides of the square and the system 2 be the motion of a point particle in the segment $E = D_0$. The resulting twisted cartesian product is a billiard system in a rectangular box in three dimensions with two cylindrical scatterers having perpendicular directions. Such systems were introduced by Simanyi and Szasz, [S-S], in a more general case of cylinders with arbitrary directions.

Strictly speaking our analysis of strict monotonicity is incomplete for such a system since in the Sinai billiard the trajectories which do not collide with the scatterer between consecutive visits to D_0 are parabolic. However it follows easily from our analysis that every trajectory is strictly monotone if it encounters both scatterers. To establish complete hyperbolicity of such a system it remains to show that the trajectories that encounter at most one scatterer form a set of zero Lebesgue measure. This was established in the paper [S-S], in a more general case.

Example 2. Let the system 1 be the billiard in a convex domain D without corners with the curved part of the boundary satisfying the property of convex scattering ($\frac{d^2r}{ds^2} < 0$), and D_0 being the flat part of the boundary (Fig. 5a). We have that any trajectory in D is strictly monotone between consecutive visits to D_0, [W3]. The system 2 is again the motion of a point particle in the segment $E = D_0$. By Theorem 16 the twisted cartesian product is completely hyperbolic.

If instead the system 1 is "half of a Bunimovich stadium" as in Fig. 5b, then the twisted cartesian product is the Papenbrock stadium. The first proof that the Papenbrock stadium is completely hyperbolic was obtained by Bunimovich and del Magno, [B-M].

Example 3. We can take as system 1 a rectangular box in 3 dimensions with a spherical cup on one side and the square D_0 on the other, that is essentially the Bunimovich-Rehachek system discussed in Section 6,[B-R]. The system 2 is the uniform motion of a point particle in the square $E = D_0$. Theorem 16 is again not immediately applicable because there are orbits in D that visit D_0 twice without entering into the spherical cap. However the proof of Theorem 16 gives us strict monotonicity on trajectories that enter both caps, in $\{x_0 > 0\}$ and in $\{x_0 < 0\}$. The proof

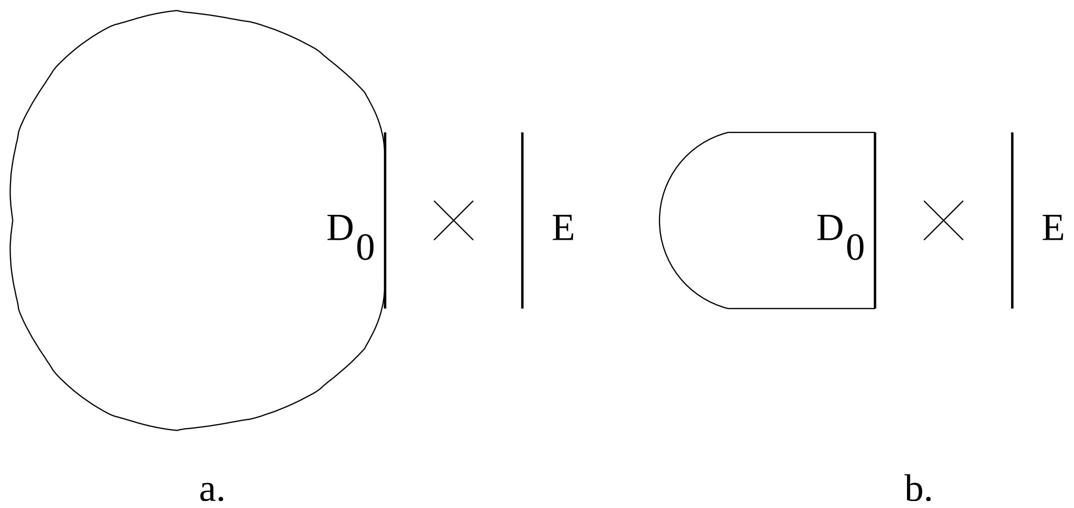

FIGURE 5

that almost all trajectories have this property is straightforward but cumbersome and we omit the details.

Let us finally observe that while the billiard domain in Example 1 can be modified with the preservation of strict monotonicity, as shown in [S-S], the billiard systems of Examples 2 and 3 are rigid in the sense that a typical perturbation of the billiard domain destroys the arguments used in Theorem 16. These arguments are based on partial integrability of cartesian products.

Thus again we are confronted with the fragility of complete monotonicity in billiards in higher dimensions (≥ 3) with convex pieces of the boundary. It is an open problem to produce a more robust construction, or to explain why it cannot be done.

REFERENCES

[Ba] P.R. Baldwin, *Soft Billiard Systems*, Pyhysica D. **29** (1988), 321 - 342.

[B1] L.A. Bunimovich, *On the ergodic properties of nowhere dispersing billiards*, Comm.Math.Phys. **65** (1979), 295 – 312.

[B2] L.A. Bunimovich, *Many-dimensional nowhere dispersing billiards with chaotic behavior*, Physica D **33** (1988), 58-64.

[B3] L.A. Bunimovich, *On absolutely focusing mirrors*, Ergodic Theory and related topics, III, Gustrow 1990, U. Krengel (Ed), Lecture Notes in Math. 1514 (1992), 62 – 82.

[B-D] L.A. Bunimovich, G. Del Magno, *Semi-Focusing Billiards: Hyperbolicity*, Comm.Math.Phys. **262** (2006), 17 – 32.

[B-R] L.A. Bunimovich, J. Rehacek, *How high-dimensional stadia look like*, Comm.Math.Phys. **197** (1998), 277 – 301.

[B-T] P. Balint , I.P. Toth, *Hyperbolicity in multi-dimensional Hamiltonian systems with applications to soft billiards*, Disc.Cont.Dyn.Syst. **15** (2006), 37 – 59.

[Ch-S] N.I. Chernov, Ya.G. Sinai, *Ergodic properties of some systems of* 2*-dimensional discs and* 3*-dimensional spheres*, Russ.Math.Surv. **42** (1987), 181 – 207.

[Ch-M] N.I. Chernov, R. Markarian, *Billiards*, AMS, 2005.

[D] V. Donnay, *Using integrability to produce chaos: billiards with positive entropy*, Comm.Math.Phys. **141** (1991), 225 - 257.

[H] *Hard ball systems and the Lorentz gas*, ed. D. Szasz, Springer-Verlag, 2000.

[K1] A. Knauf, *Ergodic and Topological Properties of Coulombic Periodic Potentials*, Comm.Math.Phys. **110** (1987), 89 - 112.

[K2] A. Knauf, *On Soft Billiard System*, Pyhysica D. **36** (1989), 259 - 262.

[K-S] A. Katok, J.-M. Strelcyn with the collaboration of F. Ledrappier and F. Przytycki, *Invariant manifolds, entropy and billiards; smooth maps with singularities*, Lecture Notes in Math. 1222, Springer-Verlag, 1986.

[K-T] V.V. Kozlov, D.V. Treschev, *Billiards. A genetic introduction to the dynamics of systems with impacts*, AMS, 1990.

[L] V.F. Lazutkin, *On the existence of caustics for the billiard ball problem in a convex domain*, Math. USSR Izv. **7** (1973), 185 – 215.

[M] R. Markarian, *Billiards with Pesin region of measure one*, Comm. Math. Phys. **118** (1988), 87 – 97.

[P] T. Papenbrock, *Numerical study of a three dimensional generalized stadium billiard*, Phys. Rev. E **61** (2000), 4626-4628.

[S-S] N. Simanyi, D. Szasz, *Nonintegrability of cylindric billiards and transitive Lie group actions*, Erg. Th. Dyn. Sys **20** (2000), 593 – 610.

[T] S. Tabachnikov, *Billiards*, Soc. Math. France, 1995.

[W1] M.P. Wojtkowski, *Systems of classical interacting particles with nonvanishing Lyapunov exponents*, Lyapunov Exponents, Proceedings, Oberwolfach 1990, L. Arnold, H. Crauel, J.-P. Eckmann (Eds), Lecture Notes in Math. 1486 (1991), 243 – 262.

[W2] M.P. Wojtkowski, *Monotonicity, J-algebra of Potapov and Lyapunov exponents*, Smooth Ergodic Theory and Its Applications, Proc. Symp. Pure Math. **69** (2001), AMS, 499 – 521.

[W3] M.P. Wojtkowski, *Principles for the design of billiards with nonvanishing Lyapunov exponents*, Comm. Math. Phys. **105** (1986), 391 – 414.

[W4] M.P. Wojtkowski, *Linearly stable orbits in* 3*-dimensional billiards*, Comm. Math. Phys. **129** (1990), 319 – 327.

[W5] M.P. Wojtkowski, *Hamiltonian systems with linear potential and elastic constraints,*, Fundamenta Matematicae **157** (1998), 305 – 341.

[W6] M.P. Wojtkowski, *Complete Hyperbolicity in Hamiltonian Systems with Linear Potential and Elastic Collisions,*, Reports on Mathematical Physics **44** (1999), 301 – 312.

Maciej P. Wojtkowski, Department of Mathematics, University of Arizona, Tucson, Arizona 85 721 USA

E-mail address: maciejw@math.arizona.edu