Principles for the Design of Billiards with Nonvanishing Lyapunov Exponents

Maciej Wojtkowski

University of Arizona, Tucson, AZ 85721, USA

Abstract. We introduce a large class of billiards with convex pieces of the boundary which have nonvanishing Lyapunov exponents.

0. Introduction

Results of Sinai on dispersing billiards [Sin] and the theorem of Lazutkin [Laz] on the existence of caustics for billiards in smooth convex domains gave credit to the idea that reflections in concave boundaries produce exponential instability as opposed to reflections in convex boundaries which are likely to produce elliptic (stable) behavior. The picture got more confusing after Bunimovich [Bun] gave examples of billiards in convex domains (the celebrated stadium) which possess strong mixing properties brought about by exponential instability. In these examples the convex pieces of the boundary of the billiard table are allowed to be only arcs of circles, and they have to be sufficiently far apart (connected by flat or concave pieces). So these examples have the unsatisfactory feature of imposing a rigid condition on the shape of the convex pieces of the boundary which would suggest that small perturbations of these examples will show elliptic behaviour.

In the present paper we show that reflections in convex pieces of the boundary satisfying certain condition (convex scattering) produce exponential instability. A curve is convex scattering if $\dfrac{d^2 r}{ds^2} \leq 0$, where $r(s)$ is the radius of curvature as the function of arc length parameter on the curve. The convex scattering condition is robust in that it allows for small perturbations of the curve (at least when the inequality is strict). Examples of convex scattering curves include (epi) (hypo) cycloids.

We are able to formulate the following "principles for the design of billiards with nonvanishing Lyapunov exponents":

P_1: All convex pieces of the boundary of the billiard table have to be convex scattering.

P_2: Any convex piece of the boundary has to be sufficiently far away from any other piece of the boundary (convex, flat or concave).

P_3: If two smooth pieces of the boundary meet at a vertex then the internal angle at the vertex has to be bigger than π when both pieces are convex, not less than π when one piece is convex and another is concave and bigger than $\frac{\pi}{2}$ when one piece is convex and another is flat.

If a billiard table satisfies these conditions then the Lyapunov exponents are nonzero. This is not a theorem since "sufficiently far away" has only a vague meaning and we do not attempt to formulate a general theorem along the lines of these principles. We feel such an attempt would only obscure the underlying ideas. Instead we consider several classes of examples. In particular the billiard inside a cardioid has nonvanishing Lyapunov exponents. Moreover this domain can be perturbed arbitrarily as long as its boundary stays convex scattering, so we get a large class of examples of this particularly simple shape. It should be noted that the billiard in a cardioid was studied numerically in [Rob].

Nonvanishing of Lyapunov exponents for a billiard leads automatically to strong mixing properties: countable number of ergodic components, positive entropy, Bernoulli property etc. Indeed Pesin theory of nonuniformly hyperbolic systems [Pes] can be generalized to systems with singularities as was shown in [Kat–Str]. What is more, it was shown there that plane billiards in particular can be treated from this point of view. It should be pointed out that ergodicity does not follow automatically from nonvanishing of Lyapunov exponents. In Appendix C we give an appropriate example in the class of billiards. Nevertheless we expect that typically (under some mild conditions) the billiards designed according to the above principles will be ergodic. It seems that the method of Sinai (see [Bun–Sin]) should apply, but the proof of ergodicity for a general class of billiards would require a detailed study of singularities which makes it a rather tedious task (for the billiard in a cardioid type domain the singularities are moderate and the Sinai proof should go through easily).

The plan of the paper is the following. In Sect. 1 we formulate the citerion (Theorem 1) for nonvanishing of Lyapunov exponents which was earlier introduced in [Wojt 1]. An important feature of this criterion is that it concerns only the projective action of the differential. It is interesting that the integral of the positive Lyapunov exponent (metric entropy) can be actually estimated from below in projective terms which is done in Theorem 2. This estimate is applied in Appendix B to the case of the billiard in the stadium close to a disk and it confirms the numerical results from [Ben–Str] and [Ben]. For the sake of completeness we include the proofs of Theorems 1 and 2 in Appendix A. It should be pointed out that these theorems are related to the Hilbert projective metric [Birk] which is made explicit in [Wojt 2].

In Sect. 2 we discuss billiards. We introduce the standard section of the flow generated by the billiard ball problem. To be able to apply Theorem 1 we describe the projective action of the differential of the section map in geometric terms (Lemmas 1 and 2 and Proposition 1). This description is very much in the spirit of the geometrical optics. We use it to give a simple proof of the theorem on nonexistence of caustics [Math], and to derive conditions of linear stability of a periodic billiard orbit (Proposition 3).

In Sect. 3 we define a bundle of sectors to be preserved by the differential of the standard section map and then we look for billiards that would actually preserve

this bundle. It turns out that the necessary condition is the convex scattering property for all convex pieces of the boundary (Definition 2 and Theorem 3). We formulate the sufficient conditions in a technical manner (not readily applicable) and then reformulate them heuristically in our three principles.

In Sect. 4 we consider several (classes of) examples and show how the method developed in Sect. 3 works in proving nonvanishing of Lyapunov exponents.

In Appendix C we discuss examples of billiards with nonvanishing Lyapunov exponents which do not satisfy our principles. It shows that the game can be played in other ways. We are also able to construct a billiard with nonvanishing Lyapunov exponents and at least two ergodic components.

1. The Projective Criterion

Let $\Phi: M^2 \to M^2$ be a piecewise differentiable invertible mapping (not necessarily continuous) of a two dimensional manifold M^2. Let us fix some Riemann metric on M^2. We further assume that Φ preserves a probabilistic measure μ which has nonvanishing density with respect to the Riemann area element v, i.e., $\dfrac{d\mu}{dv} = f$, $f > 0$, $\int_{M^2} f\, dv = 1$.

We restrict the singularities of ϕ by the following condition:

$$\int_{M^2} \log^+ \|D_x \Phi^{\pm 1}\|\, d\mu(x) < +\infty, \quad \text{where} \quad \log^+ t = \max(\log t, 0). \tag{1}$$

In this setting Oseledec Multiplicative Ergodic Theorem [Osel] applies and we obtain the existence μ almost everywhere (a.e.) of the Lyapunov exponents λ_+, λ_-, which in our case are equal to

$$\lambda_+(x) = \lim_{n \to +\infty} \frac{1}{n} \ln \|D_x \Phi^n\| \quad \text{and} \quad \lambda_-(x) = -\lambda_+(x).$$

We are going to formulate a condition on Φ under which $\lambda_+ > 0$ μ a.e.

Let $C(x)$, $x \in M^2$ be a measurable bundle of closed sectors in the tangent bundle of M^2 i.e., $C(x)$ is a closed sector in $T_x M^2$ [in appropriate linear coordinates (u_1, u_2) in $T_x M^2$ $C(x) = \{(u_1, u_2) | u_1 u_2 \geq 0\}$] defined μ a.e. and $C(x)$ depends on x measurably.

Definition 1. We say that the bundle $C(x)$, $x \in M^2$ is
   a) preserved by ϕ if $D_x \Phi(C(x)) \subset C(\Phi(x))$ μ a.e.,
   b) strictly preserved by Φ if it is preserved by Φ and if both boundary lines of $D_x \Phi(C(x))$ are contained strictly inside $C(\Phi(x))$ (μ a.e.),
   c) eventually strictly preserved by ϕ if it is preserved by ϕ and for μ almost all $x \in M^2$ there is $n(x)$ such that both boundary lines of $D_x \Phi^{n(x)}(C(x))$ are contained strictly inside $C(\Phi^{n(x)}(x))$.

Theorem 1. *If there exists a measurable bundle of sectors which is eventually strictly preserved by Φ satisfying the above properties, then the Lyapunov exponent λ_+ of Φ is positive μ a.e.*

   This theorem was proved in [Wojt 1]. Note that to check the preservation of sectors in the above sense we do not need to know $D_x\Phi$ itself but only how it acts on lines in T_xM^2 passing through the origin, i.e., on the 1-dimensional real projective space.

   It is interesting that the Lyapunov exponent λ_+ can be actually estimated below in projective terms. Indeed let us assume that for (μ almost) every $x \in M^2$ we have chosen a projective coordinate t, $-\infty \le t \le +\infty$, in the 1-dimensional projective space of lines in T_xM^2 [i.e. for appropriate euclidean coordinates (u_1, u_2) in T_xM^2 t, $-\infty < t < +\infty$, corresponds to the line $u_2 = tu_1$ and $t = \pm\infty$ corresponds to the line $u_1 = 0$]. Lines in the sector $C(x)$ form a closed interval $I(x)$ in the projective space, $I(x) = \{t \,|\, l(x) \le t \le r(x)\}$, where $l(x) < r(x)$. If the bundle of sectors $C(x)$, $x \in M^2$ is preserved by Φ, then the lines in the sector $D_x\Phi(C(x))$ form a subinterval $I_1(x)$ of $I(\Phi(x))$. If $I_1(x) = \{t \,|\, l_1(x) \le t \le r_1(x)\}$, $l_1(x) < r_1(x)$, then we have μ a.e. $l(\Phi(x)) \le l_1(x) < r_1(x) \le r(\Phi(x))$. Let $N = \{x \in M^2 \,|\, l(\Phi(x)) < l_1(x)$ and $r_1(x) < r(\Phi(x))\}$, and for $x \in N$ let $\zeta(x)$ be the cross ratio $[l_1(x), r_1(x), l(\Phi(x)), r(\Phi(x))]$ of the four points i.e.,

$$\zeta(x) = \frac{r(\Phi(x)) - l_1(x)}{r(\Phi(x)) - r_1(x)} \frac{r_1(x) - l(\Phi(x))}{l_1(x) - l(\Phi(x))}.$$

Clearly we have $\zeta(x) > 1$ for $x \in N$. $\zeta(x)$ does not depend on the choice of the projective coordinate t but only on the sectors $C(\Phi(x))$ and $D_x\Phi(C(x))$.

Theorem 2. *If a measurable bundle of sectors $C(x)$, $x \in M^2$ is preserved under Φ, then*

$$\int_{M^2} \lambda_+ \, d\mu \ge \int_N \ln \frac{\sqrt{\zeta} + 1}{\sqrt{\zeta} - 1} \, d\mu,$$

where N and ζ are defined above.

   In his theory of nonuniformly hyperbolic systems Pesin [Pes] established that nonvanishing of Lyapunov exponents implies automatically strong mixing properties. Under some technical conditions this theory was extended in [Kat–Str] to systems with singularities. In particular under these conditions the metric entropy $h_\mu(\Phi)$ is in the 2-dim setting equal to

$$h_\mu(\Phi) = \int_{M^2} \lambda_+ \, d\mu.$$

More generally whenever $h_\mu(\Phi)$ is estimated below by $\int_{M^2} \lambda_+ \, d\mu$ (for appropriate conditions see [Led–Str]) Theorem 2 yields an estimate of $h_\mu(\Phi)$.

   In the following we will apply Theorem 1 (and 2) to billiards. In general the theory of nonuniformly hyperbolic systems is applicable, as shown in [Kat–Str], to plane billiards. Hence if we establish nonvanishing of Lyapunov exponents for a billiard system, it follows that the system has at most countably many ergodic components, positive entropy, Bernoullian property on every ergodic component, etc.

   For completeness we include the proof of Theorems 1 and 2 in Appendix A. It is a variation of the argument given in [Wojt 1].

2. Billiards

We will be very brief with preliminaries since we can refer the reader to the detailed introductory discussion of billiards in Chap. 6 of [C–F–S].

Let Q be a connected domain in the plane with a piecewise smooth boundary. We distinguish three classes of smooth pieces of the boundary ∂Q: flat – straight segment, convex – the center of curvature lies on the inside of Q, and concave – the center of curvature lies on the outside of Q. The curvature of ∂Q receives a sign after we choose the unit normal vector to point inside Q, so that convex pieces of ∂Q have positive curvature and concave pieces – negative curvature.

By a billiard in Q we mean the dynamical system arising from the uniform motion of a point mass inside Q with elastic reflections ("the angle of reflection equals the angle of incidence") at the boundary. So we have the flow Ψ^t in the space $T_1 Q$ of unit tangent vectors of Q (with obvious identifications) preserving the Liouville measure (see Chap. 6 of [C–F–S]). Let $\pi : T_1 Q \to Q$ denote the natural projection.

We will deal with the following standard section $\Phi : S \to S$ of Ψ^t, rather than the flow Ψ^t itself. Let $S \subset T_1 Q$ be the set of unit vectors attached at the boundary ∂Q and pointing inside Q. For $p \in S$ draw a straight line with the direction of the unit vector, from the point $\pi(p)$ in ∂Q up to the next point of intersection with ∂Q. $\Phi(p) \in S$ is the unit vector attached at this point and symmetric to the direction of the straight line with respect to the boundary ∂Q (Fig. 1).

We introduce coordinates (s, α) in S where s is the arc length parameter along ∂Q and α is the angle, $0 \leq \alpha \leq \pi$, which the unit vector makes with ∂Q. Φ preserves the probabilistic measure $\mu = c \sin \alpha \, ds \, d\alpha$, where c is the normalizing constant. Typically Φ has singularities (is not defined or is discontinuous or nondifferentiable) but they occur only on a subset of S of measure zero. In particular the condition (1) is satisfied [Kat–Str].

Since eventually we want to apply Theorem 1 to Φ, we need information about $D\Phi$. It is possible to express $D\Phi$ in the coordinates but we are interested only in the projective action of $D\Phi$ which allows a straightforward geometric description. For this description we need some elementary differential geometry.

Let us consider a smooth family of rays $l(u)$, $|u| < \varepsilon$, in the plane i.e., a family of oriented lines. Its parametric representation is $l(u, t) = \gamma(u) + t v(u)$, where $\gamma(u)$, $|u| < \varepsilon$, is a parametrized curve such that $\gamma(u) \in l(u)$ and $v(u)$ is the unit direction vector of $l(u)$. Let further $t = f(u)$ be a function such that $l(u, f(u))$, $|u| < \varepsilon$, is the
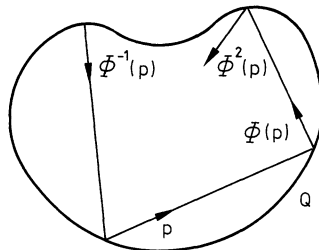

Fig. 1

envelope of the family. Equivalently $\dfrac{d}{du} l(u, f(u))$ is parallel to $v(u)$ i.e.,

$$f(u) = -\frac{(\gamma', v')}{(v', v')} \quad \text{if} \quad v' \neq 0, \tag{2}$$

where $'$ denotes differentiation with respect to u.

Geometrically $l(u, f(u))$ is the point on $l(u)$ where our family is focused in linear approximation around $l(u)$. If $v'(u) = 0$, then the family is parallel in linear approximation.

Let us assume that a family of rays is reflected from a smooth curve. Thus together with our family $l(u)$ we have another family $l_1(u)$ of reflected rays. In the parametric representation of the families we can choose γ to be the curve at which the reflection takes place. We want to find the relation between the envelopes of $l(u)$ and $l_1(u)$. Without loss of generality we can assume that $u = s$ is the arc length parameter on γ. Let further $\mathbf{t}(s)$ and $\mathbf{n}(s)$ be the unit tangent and normal vectors of γ at $\gamma(s)$, where $\mathbf{n}(s)$ points to the side of γ on which the reflection takes place. After we have chosen the normal vector $\mathbf{n}(s)$ the curvature $k(s)$ of γ has a sign.

Let $l(s, t) = \gamma(s) + t v(s)$ and $l_1(s, t) = \gamma(s) + t v_1(s)$ and $\alpha(s)$, $0 \leq \alpha(s) \leq \pi$, be the angle which $v_1(s)$ makes with $\mathbf{t}(s)$, so that $v_1 = \cos\alpha \mathbf{t} + \sin\alpha \mathbf{n}$. We have then $v = \cos\alpha \mathbf{t} - \sin\alpha \mathbf{n}$.

Lemma 1. *If* $t = f(s)$ *is the local envelope of a family of rays* l *and* $t = f_1(s)$ *the local envelope of the reflected family* l_1, *then*

$$-\frac{1}{f} + \frac{1}{f_1} = \frac{2k}{\sin\alpha}. \tag{3}$$

Proof. We have

$$v' = (-\alpha' + k) \sin\alpha \mathbf{t} + (-\alpha' + k) \cos\alpha \mathbf{n},$$

$$v_1' = -(\alpha' + k) \sin\alpha \mathbf{t} + (\alpha' + k) \cos\alpha \mathbf{n}.$$

From (2), since $\gamma' = \mathbf{t}$, we obtain

$$\frac{1}{f} = \frac{\alpha' - k}{\sin\alpha}, \qquad \frac{1}{f_1} = \frac{\alpha' + k}{\sin\alpha}. \qquad \square$$

The formula (3) is well known in geometrical optics as "the mirror equation."

Let us reformulate Lemma 1 in geometric terms. To this end let us consider an infinitesimal family of rays l, being reflected from γ at $\gamma(0)$ and producing the reflected infinitesimal family l_1. Let $F \in l(0)$ and $F_1 \in l_1(0)$ be the points where respectively the families l and l_1 are focused in linear approximation. f and f_1 are signed distances from $\gamma(0)$ to F and F_1 respectively, the sign chosen according to the orientation of $l(0)$ and $l_1(0)$. Let us further assume that the curvature k of γ at $\gamma(0)$ is different from 0 so that we can consider the following three closed disks tangent to γ at $\gamma(0)$: D_1 – the disk of curvature $\left(\text{with the radius } \dfrac{1}{|k|}\right)$, D_2 and D_4 – disks with radii $\dfrac{1}{2|k|}$ and $\dfrac{1}{4|k|}$ respectively lying on the same side of γ as D_1. Let d be the length of the segment of $l(0)$ [or $l_1(0)$] inside D_2.
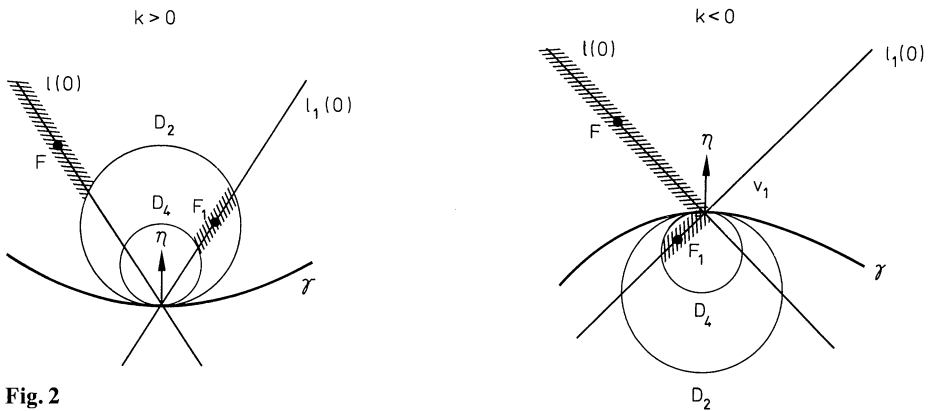
Fig. 2

From Lemma 1 we obtain immediately

Lemma 2.

$$-\frac{1}{f} + \frac{1}{f_1} = (\operatorname{sgn} k)\frac{2}{d}. \quad \square \tag{4}$$

Proposition 1. $F \notin D_2$ if and only if $F_1 \in \operatorname{int} D_2$. Moreover when $k > 0$, then if $f \leqq -d$, then $F_1 \in D_2 \backslash D_4$, and when $k < 0$, then if $f \leqq 0$, then $F_1 \in D_4$ (Fig. 2).

Proof. Straightforward from Lemma 2. $\square$

Remark. Using Lemmas 1, 2 and Proposition 1 we can draw some conclusions about the position of a caustic for a billiard problem in a convex domain Q. By a caustic for a billiard problem in Q we mean a smooth closed curve ξ in Q such that if one segment of the billiard trajectory is tangent to ξ then any other segment of this trajectory is also tangent to ξ. For a billiard trajectory tangent to a caustic ξ let $F_i, i = 0, 1, 2, \ldots$, be the consecutive points of tangency to ξ and $R_i \in \partial Q, i = 1, 2, \ldots$, the points at which the trajectory is reflected so that R_i lies between F_{i-1} and F_i on the trajectory. Let $D_2(i)$ and $D_4(i)$ denote the closed disks tangent to ∂Q at R_i as defined above. If we consider segments of nearby trajectories tangent to the caustic ξ, we obtain an infinitesimal family of rays focused in linear approximation at F_0. The reflected family of rays is focused at F_1. Hence by Proposition 1 we get that if $F_{i-1} \notin D_2(i)$, then $F_i \in \operatorname{int} D_2(i) \backslash D_4(i)$ and symmetrically if $F_i \notin D_2(i)$, then $F_{i-1} \in \operatorname{int} D_2(i) \backslash D_4(i)$.

  This line of thought allows for a very simple proof of the following theorem which is a weaker form of the theorem first formulated and proved by Mather [Math]; see also [Douad].

Theorem. *If Q is a convex domain and at some point of ∂Q the curvature is zero, then there exists no caustic (in the above sense) for the billiard problem in Q.*

Proof. Suppose there exists a caustic ξ. Let us consider segments of nearby trajectories tangent to ξ and reflected in a neighborhood of the flat point of ∂Q. They form an infinitesimal family of rays focused in linear approximation at ξ. The reflected family will be by Lemma 1 focused outside of Q so that it cannot be focused at ξ contrary to the definition of a caustic. $\square$
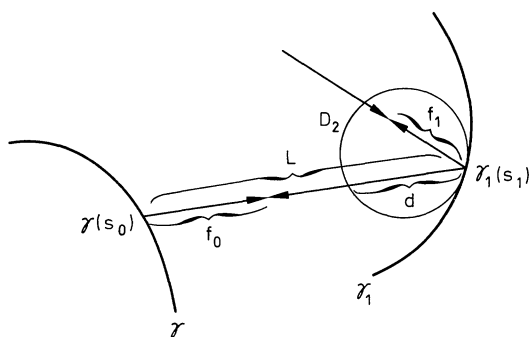
Fig. 3

We are now ready to describe geometrically the projective action of $D\Phi$ for a billiard problem in some domain Q. Let γ and γ_1 be two smooth pieces of ∂Q and $l(0)$ be a line intersecting γ and γ_1 at the points $\gamma(s_0)$ and $\gamma_1(s_1)$ respectively, oriented from $\gamma(s_0)$ to $\gamma_1(s_1)$. Let v_0 be the unit direction vector of $l(0)$. We will consider the mapping $\Phi: S \to S$ in the neighborhood of $p_0 = (\gamma(s_0), v_0) \in S$. Let $\Phi(p_0) = p_1 = (\gamma_1(s_1), v_1)$. A tangent vector to S at p_0 is an infinitesimal parametrized curve passing through p_0; in (s, α) coordinates in S: $\alpha = \alpha(s)$, $s = s(u)$. Such a curve gives rise to an infinitesimal family of rays l: $l(s - s_0, t) = \gamma(s) + tv(s)$, where $v(s)$ are unit vectors making the angle $\alpha(s)$ with the tangent to $\gamma(s)$, in particular $v(s_0) = v_0$. The signed distance f_0 from $\gamma(s_0)$ to the point on $l(0)$, where this family is focused in linear approximation equals [see (2) and the proof of Lemma 1]

$$f_0 = \frac{\sin\alpha}{-\dfrac{d\alpha}{ds} + k_0},$$

where k_0 is the signed curvature of $\gamma(s_0)$. So f_0 does not depend on the choice of parameter u on our infinitesimal curve i.e., f_0 depends only on the 1-dim subspace in the tangent space to S at p_0, $T_{p_0}S$. Moreover f_0 is actually a projective coordinate in the 1-dim projective space.

Let now f_0 and f_1 be respectively the projective coordinates of a 1-dim subspace of $T_{p_0}S$ and its image under $D_{p_0}\Phi$ in $T_{p_1}S$. Equivalently we consider an infinitesimal family of rays including $l(0) - f_0$ is the signed distance from $\gamma(s_0)$ to the point where the family is focused; and the family reflected from γ_1 at $\gamma_1(s_1) - f_1$ is the signed distance from $\gamma_1(s_1)$ to the point where the reflected family is focused (in linear approximation) (Fig. 3).

Hence we see that Lemmas 1, 2 and Proposition 1 contain the description of the action of $D\Phi$ on the 1-dim projective space. In particular the corresponding linear fractional map is given by (3) or (4). More precisely let L be the distance between $\gamma(s_0)$ and $\gamma_1(s_1)$, d be the length of the segment of $l(0)$ inside the disk D_2 constructed for γ_1 at $\gamma_1(s_1)$ and k_1 be the signed curvature of γ_1 at $\gamma_1(s_1)$.

Proposition 2.

$$f_1 = \frac{df_0 - dL}{2(\operatorname{sgn} k_1)f_0 - 2(\operatorname{sgn} k_1)L + d}. \tag{5}$$

Proof. Follows immediately from Lemma 2 after substitution $f = f_0 - L$. $\quad\square$

In Sect. 3 we will use this description of the projective action of $D\Phi$ to check the assumptions of Theorem 1 for a class of billiards. Before that let us use the formula (5) to derive a criterion of linear stability of a periodic billiard trajectory.

Let $l(0)$ be orthogonal to the boundary of Q at two consecutive points of intersection with ∂Q. The segment of $l(0)$ between these points supports a periodic trajectory for the billiard problem. For the section map Φ we get a periodic point of period 2.

Proposition 3. *This periodic trajectory is linearly stable if and only if*

$$Lk_0 k_1 - k_0 - k_1 < 0 \tag{6}$$

and

$$(Lk_0 - 1)(Lk_1 - 1) > 0, \tag{7}$$

where k_0, k_1 are signed curvatures of ∂Q at the two points and L is the distance between them (the sign of the curvature is obtained by the choice of the interior unit normal vector for ∂Q).

[Note that the condition for linear stability quoted in [Arn, p. 456] is thus only necessary but not sufficient. Indeed when $k_0 > 0$ and $k_1 > 0$, then the conditions (6) and (7) read

$$L - r_0 - r_1 < 0 \quad \text{and} \quad (L - r_0)(L - r_1) > 0,$$

where r_0, r_1 are the respective radii of curvature.]

Proof. Linear stability of the orbit is equivalent to $D\Phi^2$ being elliptic at the corresponding point. This can be decided from the knowledge of only the projectivization of $D\Phi^2$ which is a linear fractional map readily obtainable from Proposition 2. In our case the lengths of segments of $l(0)$ inside the D_2 disks at the points of reflection of our periodic trajectory are equal to $\dfrac{1}{|k_0|}$ and $\dfrac{1}{|k_1|}$. By simple transformations we obtain from (5) that $D\Phi^2$ is elliptic if and only if

$$|\operatorname{tr} A| < 2, \quad \text{where} \quad A = \begin{vmatrix} 1 & -L \\ 2k_1 & 1 - 2Lk_1 \end{vmatrix} \begin{vmatrix} 1 & -L \\ 2k_0 & 1 - 2Lk_0 \end{vmatrix}.$$

This yields $-2 < 2 - 4Lk_0 - 4Lk_1 + 4L^2 k_0 k_1 < 2$ which is equivalent to (6) and (7). $\square$

It follows from Proposition 3 that if we move sufficiently far apart the two pieces of ∂Q intersecting the line $l(0)$ perpendicularly, then the resulting periodic trajectory becomes linearly unstable: when the curvatures k_0, k_1 have the same sign, (6) is violated for large L, and when the curvatures k_0, k_1 have the opposite signs, then (7) is violated for large L.

3. Convex Scattering Property

Let Q be a connected domain in the plane or the flat torus with ∂Q being the union of smooth pieces: convex (positive curvature), concave (negative curvature) or flat

(zero curvature). Let S be the set of unit tangent vectors attached at ∂Q and pointing inwards and $S_1 \subset S$ be part of S over convex and concave pieces. Let $\Phi : S \to S$ be the section of the billiard flow as defined in Sect. 2 and let $\Phi_1 : S_1 \to S_1$, be the first return map of Φ. We assume that almost every trajectory of the billiard passes through S_1 (actually we do not know of any nontrivial example in which it would not be the case). We restrict our attention to S_1 and Φ_1 because the reflections in flat pieces of ∂Q are almost irrelevant for the exponential instability (or its absence) – indeed one can think that the trajectory continues freely into the reflected domain.

We will now define a bundle of sectors C in the tangent planes of S_1 and then look for conditions on Q under which C is strictly preserved (eventually strictly preserved) by ϕ_1 so that Theorem 1 applies.

According to the discussion in Sect. 2 a sector in the tangent plane of S_1 can be defined by a condition on the projective coordinate $f = $ the signed distance from the respective point in ∂Q to the "focus" of the infinitesimal family of rays corresponding to a vector tangent to S. Let for any convex piece γ of the boundary ∂Q $D_2(s)$ be the closed disk tangent to γ at $\gamma(s)$ with the radius equal to $\frac{1}{2}$ of the radius of curvature as defined in Sect. 2. For points in S_1 over the point $\gamma(s)$ in a convex piece γ the sector C is defined by the condition that the "focus" of the respective infinitesimal family of rays lies in $D_2(S)$; equivalently in the 1-dim projective space the sector C is defined as $\{f \,|\, 0 \leqq f \leqq d\}$, where d is the length of the segment of the billiard trajectory inside $D_2(s)$. For the points in S_1 over a concave piece of ∂Q the sector C is defined by the condition that the "focus" of the respective family of rays lies on the exterior side of the boundary ∂Q; equivalently in 1-dim projective space the sector C is defined as $\{f \,|\, -\infty \leqq f \leqq 0\}$.

Let us assume that Q is such that C is (strictly) preserved by Φ_1. We want to reveal what this assumption implies about Q. For this purpose let us consider a piece of the billiard trajectory starting at $\gamma(s_0)$ in a nonflat piece γ of ∂Q and ending at $\gamma_1(s_1)$ in another nonflat piece γ_1 (we do not exclude that $\gamma = \gamma_1$) and having in between an arbitrary number of reflections at the flat pieces. By Proposition 1 we obtain

Case 1. γ and γ_1 are concave.

The preservation of C by Φ_1 [along the billiard trajectory from $\gamma(s_0)$ to $\gamma_1(s_1)$] imposes no conditions whatsoever on γ and γ_1. What is more the preservation of C is automatically strict. Note that if ∂Q has no convex pieces then the bundle of sectors C is strictly preserved, and by Theorem 1 the mapping Φ_1, and hence also Φ have nonvanishing Lyapunov exponents. This is the celebrated class of Sinai dispersing billiards, [Sin].

Case 2. γ is convex and γ_1 is concave.

The preservation of C by ϕ_1 implies that $d \leqq L$, where L is the length of the trajectory from $\gamma(s_0)$ to $\gamma_1(s_1)$ and d is the length of the appropriate segment in the D_2 disk at $\gamma(s_0)$ (as in the definition of C) (Fig. 4). Strict preservation is equivalent to $d < L$.

Case 3. γ is concave and γ_1 is convex. Analogous to Case 2.

Case 4. γ and γ_1 are convex.
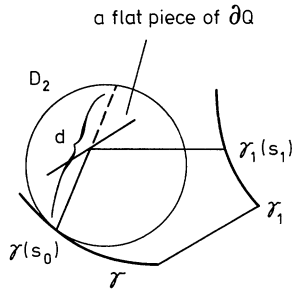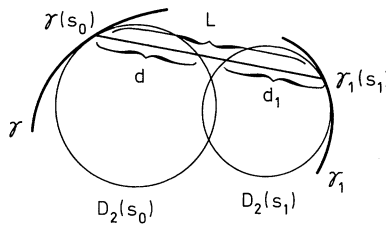
a flat piece of ∂Q

Fig. 4

Fig. 5

The preservation of C implies that $d + d_1 \leqq L$, where L is the length of the trajectory from $\gamma(s_0)$ to $\gamma_1(s_1)$ and d, d_1 are the lengths of the segments inside the respective D_2 disks at $\gamma(s_0)$ and $\gamma(s_1)$ (d, d_1 appear in the definition of the respective sectors C) (Fig. 5). The strict preservation of C is equivalent to $d + d_1 < L$.

We see that the preservation of the bundle of sectors C can be always arranged by moving the nonflat pieces of ∂Q sufficiently far apart, except for the Case 4, $\gamma = \gamma_1$ – the preservation of C in this case imposes a severe restriction on γ. Our next goal is to describe this restriction in more explicit terms.

Let γ be a smooth curve with nonvanishing curvature. At every point $\gamma(s)$ of γ we consider the closed disk $D_2(s)$ tangent to γ at $\gamma(s)$ with the radius equal to $\frac{1}{2}$ of the radius of curvature and lying on the same side of γ as the disk of curvature.

Definition 2. A smooth curve γ is (strictly) convex scattering if for every segment $[\gamma(s_0), \gamma(s_1)]$, $\gamma(s_0) \neq \gamma(s_1)$, such that the arc of γ between $\gamma(s_0)$ and $\gamma(s_1)$ lies entirely on one side of the straight line passing through $\gamma(s_0)$ and $\gamma(s_1)$, we have that

$$D_2(s_0) \cap D_2(s_1) \cap [\gamma(s_0), \gamma(s_1)]$$

contains at most one point (is empty).

Clearly the (strict) preservation of C in Case 4, $\gamma = \gamma_1$ is equivalent to γ being (strictly) convex scattering. That is why the following theorem is crucial in our approach to billiards.

Theorem 3. *A smooth curve γ is (strictly) convex scattering if and only if the radius of curvature $r(s)$ is a (strictly) concave function of the arc length parameter s, i.e.,*

$$\frac{d^2 r}{ds^2} \leqq 0 \quad \left(\frac{d^2 r}{ds^2} < 0 \text{ for an open dense set of } s \right).$$

Proof. Let $\gamma(s_0)$ and $\gamma(s_1)$ be two arbitrary points on γ such that the arc of γ between $\gamma(s_0)$ and $\gamma(s_1)$ lies entirely on one side of the straight line passing through $\gamma(s_0)$ and $\gamma(s_1)$. We choose cartesian coordinates (x, y) in such a way that $\gamma(s_0)$ is the beginning of coordinates, $\gamma(s_1)$ lies on the positive x-axis and the arc of γ between $\gamma(s_0)$ and $\gamma(s_1)$ lies below the x-axis. Let $\gamma(s) = (x(s), y(s))$ be the arc length parametrization of γ and $\phi(s)$ be the angle which $\dfrac{d\gamma}{ds}$ makes with the x-axis i.e., $\dfrac{d\gamma}{ds}$ $= (\cos\phi, \sin\phi)$. By $r(s)$ we denote the radius of curvature of γ at $\gamma(s)$ i.e., $r = \dfrac{ds}{d\phi}$. Let the length of the segment $[\gamma(s_0), \gamma(s_1)]$ be equal to L and the lengths of the segments $[\gamma(s_0), \gamma(s_1)] \cap D_2(s_i)$ be equal to d_i, $i = 0, 1$ i.e., $d_0 = -r(s_0)\sin\phi(s_0)$ and $d_1 = r(s_1)\sin\phi(s_1)$. We have

$$L = \int_{s_0}^{s_1} \frac{dx}{ds}\,ds = \int_{s_0}^{s_1} \cos\phi(s)\,ds = \int_{s_0}^{s_1} \frac{d\sin\phi}{ds}\frac{ds}{d\phi}\,ds = \int_{s_0}^{s_1} \frac{d\sin\phi}{ds}r(s)\,ds$$

$$= r(s_1)\sin\phi(s_1) - r(s_0)\sin\phi(s_0) - \int_{s_0}^{s_1} \sin\phi\frac{dr}{ds}\,ds.$$

Hence

$$L - d_0 - d_1 = -\int_{s_0}^{s_1} \sin\phi\frac{dr}{ds}\,ds = -\int_{s_0}^{s_1} \frac{dy}{ds}\frac{dr}{ds}\,ds = -y(s_1)\frac{dr}{ds}(s_1) + y(s_0)\frac{dr}{ds}(s_0)$$

$$+ \int_{s_0}^{s_1} y(s)\frac{d^2r}{ds^2}\,ds = \int_{s_0}^{s_1} y(s)\frac{d^2r}{ds^2}\,ds,$$

since $y(s_0) = y(s_1) = 0$. But $y(s) \leq 0$ for $s_0 \leq s \leq s_1$, so that if $\dfrac{d^2r}{ds^2} \leq 0$ for $s_0 \leq s \leq s_1$, then $d_0 + d_1 \leq L$. This proves that if $\dfrac{d^2r}{ds^2} \leq 0$ on γ then γ is convex scattering. If on the contrary $\dfrac{d^2r}{ds^2} > 0$ at some point $\gamma(s)$, then if we choose s_0, s_1 sufficiently close to s, we obtain $d_0 + d_1 > L$. The strict convex scattering version of the theorem is obtained in the same way. $\square$

Examples of convex scattering curves:

1. An arc of a circle; more generally an arc of a logarithmic spiral: $r(s) = as + b$ for some a, b (nonstrict convex scattering property).

2. An arc of a cardioid; more generally of any epi or hypocycloid including the limit case of a cycloid: $r^2(s) = -a^2s^2 + b^2$, $|a| < 1$ – epicycloid, $|a| = 1$ – cycloid, $|a| > 1$ – hypocycloid (strict convex scattering property).

3. The two arcs of an ellipse $\dfrac{x^2}{a^2} + \dfrac{y^2}{b^2} = 1$, $a > b > 0$, on which $|x| \leq a/\sqrt{2}$. Indeed if the ellipse is parameterized by $\gamma(t) = (a\cos t, b\sin t)$, then $\dfrac{dr}{ds} = \dfrac{3}{2}\dfrac{a^2 - b^2}{ab}\sin 2t$ (strict convex scattering property).

Note that the strict convex scattering property is an open condition i.e., it is preserved under small (in C^4 topology) perturbation of the curve.

The necessary and sufficient conditions for the preservation of the bundle of sectors C which we have formulated so far do not seem to be readily applicable. Nevertheless we will not attempt to formulate simpler sufficient conditions. Instead in the next section we will consider some specific shapes of Q and show that our conditions can be actually verified.

We can summarize the above results into the following somewhat vague principles for the design of billiards with nonvanishing Lyapunov exponents:

P_1: All the convex pieces of ∂Q have to be convex scattering.

P_2: Any convex piece of ∂Q has to be sufficiently far away from any other piece of the boundary.

P_3: If two smooth pieces of ∂Q meet at a vertex, then the interval angle at the vertex has to be bigger than π when both pieces are convex, not less than π when one piece is convex and the other concave and bigger than $\frac{\pi}{2}$ when one piece is convex and the other flat.

The meaning of "sufficiently far away" in P_2 is described in terms of the D_2 disks in Cases 2, 3, and 4 above. The condition in P_3 arises naturally from the consideration of trajectories reflecting at ∂Q in the neighborhood of the vertex. We skip the details of the geometrical analysis leading to this condition.

4. Examples

1. Q bounded by the cardioid (Fig. 6).

∂Q is not smooth but has only one convex smooth piece which is strictly convex scattering, so that the sector bundle C is strictly preserved by Φ. By Theorem 1 Φ has nonvanishing Lyapunov exponents.

Since the strict convex scattering property is robust, we get a whole class of examples by perturbing Q. There is no convex domain Q with only one smooth piece of ∂Q which is strictly convex scattering, but if we take Q bounded by the cardioid and consider its convex hull $\tilde{Q}$ we get a convex example of a billiard with nonvanishing Lyapunov exponents. Indeed $\partial\tilde{Q}$ contains two pieces, one convex-strictly convex scattering, another flat. The D_2 disks for the convex piece lie entirely in Q so that the sector bundle C is preserved also on trajectories reflecting at the flat piece.

2. Q bounded by an epicycloid e.g., nephroid (Fig. 7).

The smooth pieces of the boundary of an epicycloid are strictly convex scattering. To establish strict preservation of the sector bundle C we have yet to check that a segment connecting two different smooth pieces is not contained in
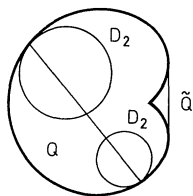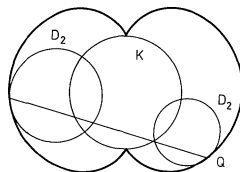

Fig. 6

Fig. 7

the union of the respective D_2 disks. This follows immediately from the following geometric property of (epi) (hypo) cycloids.

Proposition 4. (a) *The D_2 disks of an epi- or hypocycloid, obtained by rolling a circle upon the fixed circle K, intersect K orthogonally.*

(b) *Let γ be the maximal smooth arc of the epi- or hypocycloid and let $\tilde{\gamma}$ be the inversion of γ with respect to K; then $\tilde{\gamma}$ is the envelope of the family of D_2 disks of γ.*

Proposition 4 holds also for a cycloid if we replace the circle K by a straight line and the inversion by the symmetry with respect to the line.

Proof. Since a circle orthogonal to K is invariant under inversion with respect to K, part (b) follows immediately from part (a). To establish (a) let us assume that K is the unit circle with the center at the beginning of coordinates and the x-axis bisects the smooth arc γ between two successive cusps. We have

$$\gamma(t) = (x(t), y(t)),$$

$$x(t) = (1 \pm w)\cos t \pm w \cos\left(1 \pm \frac{1}{w}\right)t,$$

$$y(t) = (1 \pm w)\sin t \pm w \sin\left(1 \pm \frac{1}{w}\right)t,$$

where w is the radius of the circle rolled upon K, t, $-\pi w < t < \pi w$, has clear geometric meaning, and we take $+$ for an epicycloid and $-$ for a hypocycloid. In case of a hypocycloid we further assume that $2w < 1$ which does not reduce the generality because the rolling circles with radii w and $1 - w$ generate the same hypocycloid (Daniel Bernoulli 1725). By straightforward computation we find the radius of curvature $r(t)$ and the unit normal vector $\eta(t)$ of γ at $\gamma(t)$:

$$r(t) = \frac{4w(2 \pm w)}{1 \pm 2w}\cos\frac{1}{2w}t, \quad \eta(t) = \mp\left(\cos\left(1 \pm \frac{1}{2w}\right)t, \sin\left(1 \pm \frac{1}{2w}\right)t\right).$$

$\gamma + \frac{1}{2}r\eta$ is the center of the D_2 disk so we have to check that $(\gamma + \frac{1}{2}r\eta)^2 - \frac{1}{4}r^2 = 1$, which is the condition for orthogonality. It is equivalent to $\gamma^2 + (\gamma, r\eta) = 1$, which can be readily verified. ☐

3. $Q_a = \{(x, y) \in R^2 \mid |x| \leq \frac{1}{2}, |y| \leq \frac{1}{2}, |x|^{2/3} + |y|^{2/3} \geq a^{2/3}\}$ i.e., Q_a is a square with a hole in the shape of an astroid (four cusped hypocycloid).

We are going to establish that if $a \leq \sqrt{2}/4$ then the billiard in Q_a has nonvanishing Lyapunov exponents. Reflecting Q_a in its sides we obtain the domain $\tilde{Q}_a$ (Fig. 8). We identify the sides of $\tilde{Q}_a$ in the usual way and obtain the torus with four holes. The billiard in Q_a is equivalent to the billiard in this toral domain, where reflections take place only at the smooth pieces of the hypocycloids 1, 2, 3, 4. These pieces being strictly convex scattering, we have only to check that they are sufficiently far apart. Using Proposition 4 it is easy to establish that $a < \dfrac{\sqrt{2}}{4}$ is equivalent to the fact that the unions of D_2 disks of hypocycloids 1 and 3 in Fig. 8 do not intersect. This condition is not sufficient for the unions of D_2 disks of

Fig. 8

Fig. 9

hypocycloids 1 and 2 to be disjoint, but this is not necessary for our purpose. We have only to check that under this condition any segment connecting the hypocycloids 1 and 2 is not contained in the union of the respective D_2 disks. This is a cumbersome task which is significantly facilitated by the following general observation.

Let us consider two curves γ_1 and γ_2 and a family of parallel segments I_t, $|t| < \varepsilon$ connecting γ_1 and γ_2, parameterized by the distance from the segment I_0 (Fig. 9). Let us introduce arc length parameterizations $\gamma_1(s)$ and $\gamma_2(s)$ in such a way that the segment I_t has end points $\gamma_1(s_1(t))$ and $\gamma_2(s_2(t))$ and $\dfrac{ds_1}{dt} < 0$ and $\dfrac{ds_2}{dt} > 0$. Let $L(t)$ be the length of I_t and $d_1(t)$, $d_2(t)$ be the lengths of the intersections of I_t with D_2 disks of respectively γ_1 at $\gamma_1(s_1(t))$ and γ_2 at $\gamma_2(s_2(t))$; we assume γ_1 and γ_2 are curved in such a way that the disks are on the side of the segments.

Proposition 5.

$$\frac{d}{dt}(L(t) - d_1(t) - d_2(t)) = -\frac{dr_1}{ds}(s_1(t)) + \frac{dr_2}{ds}(s_2(t)),$$

where $r_i(s)$ is the radius of curvature of γ_i at $\gamma_i(s)$, $i = 1, 2$.

Proof. Without loss of generality we can assume that γ_1 and γ_2 are arcs of one smooth curve. Let us choose cartesian coordinates so that I_0 lies on the x-axis and I_t lies on the line $y = t$. If we denote by $r(s)$ the radius of curvature of γ as the function of the arc length and by $\phi(s)$ the angle which the tangent vector of γ makes

with the x-axis then as was shown in the proof of Theorem 3

$$L(t) - d_1(t) - d_2(t) = - \int\limits_{s_1(t)}^{s_2(t)} \sin\phi \frac{dr}{ds} ds,$$

so

$$\frac{d}{dt}(L(t) - d_1(t) - d_2(t)) = - \sin\phi(s_2(t)) \frac{dr}{ds}(s_2(t)) \frac{ds_2}{dt}$$

$$+ \sin\phi(s_1(t)) \frac{dr}{ds_1}(s_1(t)) \frac{ds_1}{dt}.$$

But $\dfrac{ds_1}{dt} = \dfrac{1}{\sin\phi}$ and $\dfrac{ds_2}{dt} = \dfrac{1}{\sin\phi}$. $\quad\square$

Proposition 5 allows to pick easily from a family of parallel segments the segment for which $L - d_1 - d_2$ has minimal value. Our task is to show that $L - d_1 - d_2 > 0$ for all segments connecting two pieces of the hypocycloids 1 and 2. Now it is much easier to establish that indeed the condition $a < \dfrac{\sqrt{2}}{4}$ is sufficient for this.

4. Stadium-like Q. Let Q be the domain bounded by two convex scattering pieces γ_1, γ_2 and two flat pieces γ_+, γ_-. In coordinates: $\gamma_\pm = \{-a \leq x \leq a, y = \pm w\}$, $a \geq 0$, $w > 0$. γ_1 connects $(-a, w)$ with $(-a, -w)$ and γ_2 is symmetric to γ_1 with respect to the y-axis. For simplicity we assume that γ_1 and γ_2 are symmetric with respect to the x-axis (Fig. 10).

It is clear that for given shapes of γ_1, γ_2 there is a_0 such that if $a > a_0$ then the billiard in the domain Q described above has nonvanishing Lyapunov exponents. Proposition 5 is very useful for finding minimal a_0. In particular the segment of x-axis in Q supports a periodic orbit which becomes linearly unstable if $L - 2r > 0$, where L is the length of the segment and r is the radius of curvature of ∂Q at the endpoints (Proposition 3). It looks plausible that if this orbit is linearly unstable then the pieces γ_1 and γ_2 are already sufficiently far apart, but it turns out not to be true in general – still it is true sometimes e.g., if γ_1, γ_2 are convex scattering arcs of
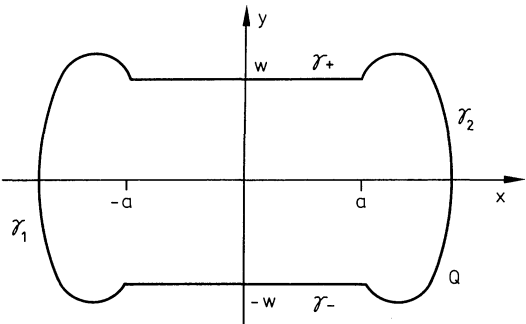

Fig. 10

an ellipse, or if γ_1, γ_2 are smooth arcs of a cycloid (which incidentally follows without any computations from the cycloid version of Proposition 4).

5. Bunimovich billiards, [Bun]. In this class of billiards the convex pieces of the boundary have to be only arcs of circles which (in the simplest version) must be contained inside the billiard table. So clearly for such billiards the bundle of sectors C as defined in Sect. 3 is preserved. These billiards were already treated in [Wojt 1] from the point of view of the projective criterion, using a different bundle of sectors. Bunimovich announced in [Bun] that these billiards are actually ergodic.

Appendix A. The Proofs of Theorems 1 and 2

Let us choose a basis $\{e_1(x), e_2(x)\}$ in $T_x M^2$ such that $e_1(x)$, $e_2(x)$ belong to the boundary lines of $C(x)$ and

$$C(X) = \{v \in T_x M^2 \mid v = u_1 e_1 + u_2 e_2, u_1 u_2 \geqq 0\} .$$

Moreover we require that the area defined by the Riemann metric of the parallelogram spanned by $e_1(x)$, $e_2(x)$ is equal to $\dfrac{1}{f(x)}$, where f is the density of the invariant measure μ with respect to the Riemann area element. Clearly we can make the choice so that $e_1(x)$ and $e_2(x)$ depend on x measurably.

The choice of bases leads to the identification of all $T_x M$ with the standard plane $\mathbb{R}^2$ and $D_x \Phi$ is described then by a matrix $A(x) = \begin{vmatrix} a(x) & b(x) \\ c(x) & d(x) \end{vmatrix}$. In terms of these matrices preservation of the bundle $C(x)$ means that entries of $A(x)$ are either all nonnegative or all nonpositive and invariance of the measure μ means that the determinant of $A(x)$ is equal to ± 1. Strict preservation of $C(x)$ by $D_x \Phi$ implies that $A(x)$ has no zero entries. The cross ratio $\zeta(x)$ is then equal to the cross ratio $\left[0, +\infty, \dfrac{c(x)}{a(x)}, \dfrac{d(x)}{b(x)} \right]$ if $\det A(x) = 1$ and the cross ratio $\left[0, +\infty, \dfrac{d(x)}{b(x)}, \dfrac{c(x)}{a(x)} \right]$ if $\det A(x) = -1$, i.e.,

$$\zeta(x) = \begin{cases} \dfrac{a(x) \, d(x)}{b(x) \, c(x)} & \text{if} \quad \det A(x) = 1 \\[3mm] \dfrac{b(x) \, c(x)}{a(x) \, d(x)} & \text{if} \quad \det A(x) = -1 . \end{cases}$$

So by the above choice of bases we have reduced our problem to a question about growth of products of 2×2 matrices with all nonnegative or all nonpositive entries. We shall now prove several lemmas in this direction.

Let $0 = \{(u_1, u_2) \in \mathbb{R}^2 \mid u_1 u_2 \geqq 0\}$ and $F : 0 \to \mathbb{R}$, $F(u) = u_1 u_2$ if $u = (u_1, u_2)$. Let further $\mathbf{P} = \{A \in \mathrm{GL}(2, \mathbb{R}) \mid A 0 \subset 0\}$ and $SP = \{A \in \mathbf{P} \mid |\det A| = 1\}$. For $A \in \mathbf{P}$ we set $\varrho(A) = \sup\limits_{u \in \mathrm{int} 0} \left(\dfrac{F(Au)}{F(u)} \right)^{1/2}$. $\| \cdot \|$ denotes the euclidean norm in $\mathbb{R}^2$ and the corresponding matrix norm.

Lemma A.1. *If $u \in 0$, then $\|u\| \geqq \sqrt{2} \sqrt{F(u)}$.* $\quad \square$

Lemma A.2. *If $A, B \in \mathbf{P}$, then $\varrho(AB) \geq \varrho(A)\varrho(B)$.*

Proof. Since $B \in \mathbf{P}$ $B \operatorname{int}0 \subset \operatorname{int}0$, so that if $u \in \operatorname{int}0$, then $F(Bu) \neq 0$. We have

$$\varrho(AB) = \inf_{u \in \operatorname{int}0} \left(\frac{F(ABu)}{F(Bu)} \right)^{1/2} \left(\frac{F(Bu)}{F(u)} \right)^{1/2} \geq \varrho(A)\varrho(B) . \quad \square$$

Lemma A.3. *If $A \in \mathbf{P}$, then $\|A\| \geq \varrho(A)$.*

Proof. Let $\mathbf{1} = (1, 1)$. Then

$$\|A\| \geq \frac{\|A\mathbf{1}\|}{\|\mathbf{1}\|} \geq \sqrt{F(A\mathbf{1})} = \left(\frac{F(A\mathbf{1})}{F(\mathbf{1})} \right)^{1/2} \geq \varrho(A) . \quad \square$$

Lemma A.4. *If $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \in \mathbf{P}$, then $\varrho(A) = \sqrt{ad} + \sqrt{bc}$. In particular if $A \in \mathbf{SP}$ and it has all nonzero entries, then $\varrho(A) = \dfrac{\sqrt{\zeta} + 1}{\sqrt{\zeta} - 1}$, where*

$$\zeta = \begin{cases} \dfrac{ad}{bc} & \text{if} \quad \det A = 1 \\[3mm] \dfrac{bc}{ad} & \text{if} \quad \det A = -1 . \end{cases}$$

Proof. Straightforward computation. $\quad \square$

Equipped with these lemmas we go back to our mapping Φ. We have $\lambda_+(x) = \lim\limits_{n \to +\infty} \dfrac{1}{n} \ln \|D_x \Phi^n\|$, where the norm of $D_x \Phi^n$ is defined by the Riemann metric $\left(\text{note that the condition (1) implies that } \int\limits_{M^2} \lambda_+ \, d\mu < +\infty \right)$. Let us consider the family of norms $\|\cdot\|_x$, $x \in M^2$, in $\mathbb{R}^2$ induced from the Riemann metric on M^2 by the identification of $T_x M^2$ with $\mathbb{R}^2$. The norms $\|\cdot\|_x$ are equivalent to the standard euclidean norm $\|\cdot\|$ for all x, so that we have $\alpha(x)\|\cdot\|_x \leq \|\cdot\| \leq \beta(x)\|\cdot\|_x$, where $\beta(x) \geq \alpha(x) > 0$, but $\alpha(x)$ and $\beta(x)$ are in general only measurable functions of x. Let for $d > 0$ $M_d = \{x \in M^2 \mid \alpha(x) \geq d^{-1} \text{ and } \beta(x) \leq d\}$. We have clearly $\lim\limits_{d \to +\infty} \mu(M_d) = 1$. For almost all $x \in M_d$ there are infinitely many n_k, $k = 1, 2, \ldots$ such that $\Phi^{n_k} x \in M_d$ (actually Φ_x^n return to M_d with positive frequency). If $x \in M_d$ and $\Phi_x^n \in M_d$, then

$$\|D_x \Phi^n\| = \sup_{u \neq 0} \frac{\|A(\Phi^{n-1}x) \ldots A(x)u\|_{\Phi_x^n}}{\|u\|_x} \geq \frac{1}{d^2} \sup_{u \neq 0} \frac{\|A(\Phi^{n-1}x) \ldots A(x)u\|}{\|u\|}$$

$$= \frac{1}{d^2} \|A(\Phi^{n-1}x) \ldots A(x)\| .$$

So we get for $x \in M_d$, using Lemma A.3,

$$\lambda_+(x) = \lim_{n \to +\infty} \frac{1}{n} \ln \|D_x \Phi^n\| = \lim_{k \to +\infty} \frac{1}{n_k} \ln \|D_x \Phi^{n_k}\|$$

$$\geq \limsup_{k \to +\infty} \frac{1}{n_k} \ln \frac{1}{d^2} \|A(\Phi^{n_k-1}x) \ldots A(x)\|$$

$$\geq \limsup_{k \to +\infty} \frac{1}{n_k} \ln \varrho(A(\Phi^{n_k-1}x) \ldots A(x))$$

$$\geq \limsup_{k \to +\infty} \frac{\ln \varrho(A(x)) + \ldots + \ln \varrho(A(\Phi^{n_k-1}x))}{n_k}.$$

Let $h(x) = \ln \varrho(A(x))$. A priori we do not know if $\int_{M^2} h \, d\mu < +\infty$, so for $m \geq 1$ we introduce the "cut off" function $h_m(x) = \min(h(x), m)$, and its ergodic mean $h_m^*(x)$ $= \lim_{n \to +\infty} \frac{1}{n}(h_m(x) + \ldots + h_m(\Phi^{n-1}x))$. From the above inequality we get for any fixed m that

$$\lambda_+(x) \geq \limsup_{k \to +\infty} \frac{1}{n_k}(h_m(x) + \ldots + h_m(\Phi^{n_k-1}x)) = h_m^*(x).$$

Since the right-hand side does not depend on the choice of d we obtain that for a.e. $x \in M^2$ $\lambda_+(x) \geq h_m^*(x)$. Integrating we get $\int_{M^2} \lambda_+ \, d\mu \geq \int_{M^2} h_m^* \, d\mu = \int_{M^2} h_m \, d\mu$ (the last equality is one of the conclusions of the Birkhoff ergodic theorem). Since this inequality is valid for all m we get finally that

$$\int_{M^2} \lambda_+ \, d\mu \geq \int_{M^2} h \, d\mu$$

$\left(\text{by the way we see that actually } \int_{M^2} h \, d\mu < +\infty\right)$. Clearly we have for (almost) all $x \in M^2$ $h(x) \geq 0$ and $N = \{x \in M^2 \,|\, h(x) > 0\}$. By Lemma A.4 we obtain immediately Theorem 2.

To prove Theorem 1 let us note that $\lambda_+(x) \geq h^*(x)$, where

$$h^*(x) = \lim_{n \to +\infty} \frac{1}{n}(h(x) + \ldots + h(\Phi^{n-1}x)).$$

Applying the Birkhoff ergodic theorem to the subset $\{x \in M^2 \,|\, h^*(x) = 0\}$ and the function h, we get that on this subset $h = 0$ a.e. So we get that $\lambda_+(x) > 0$ except possibly on the set where $\varrho(A(x)) = 1$ i.e., where $A(x)$ has some zero entries. The Lyapunov exponent for Φ^n is equal to $n\lambda_+$ and applying the above agreement to Φ^n we get that $n\lambda_+ > 0$, except possibly on the set where $\varrho(A(\Phi^{n-1}x) \ldots A(x)) = 1$. Since Φ eventually strictly preserves C, then this last set is smaller and smaller as $n \to +\infty$ so that finally we can conclude that $\lambda_+(x) > 0$ for a.e. $x \in M^2$.

Appendix B

Let us consider the billiard in the stadium Q i.e., a special case of the class of billiards considered in Example 4, Sect. 4 in which γ_1, γ_2 are half-circles with radius r and $\gamma_\pm = \{-a \le x \le a, y = \pm r\}$, so that the domain Q depends on two parameters a and r. This billiard was originally studied by Bunimovich [Bun], who in particular announced that the system is actually ergodic (for $a > 0$). If this is the case the Lyapunov exponent λ_+ for the corresponding section map Φ is constant a.e. (and thus it is equal to the metric entropy of Φ). In [Ben–Str] λ_+ was investigated numerically as a function of $\varepsilon = \dfrac{a}{r}$ at a fixed value ($= 4 + \pi$) of the area of Q. Their results indicate that for ε small λ_+ grows with ε faster than linearly (note that for $\varepsilon = 0$ we get the billiard in the disk which is integrable so that $\lambda_+ = 0$). The actual asymptotics was found numerically by Benettin [Ben] to be $\sqrt{\varepsilon}$ and he gave a good explanation for it. Using Theorem 2 we are able to support this fact rigorously. More precisely we will show that for sufficiently small ε $\lambda_+ \ge \mathrm{const}\sqrt{\varepsilon}$.

Let as in Sect. 3 S_1 be the part of S over the convex pieces γ_1 and γ_2 i.e., $\pi(S_1) = \gamma_1 \cup \gamma_2$, where $\pi: T_1 Q \to Q$ is the natural projection. The bundle of sectors C defined in Sect. 3 is preserved by Φ and on $S_c = \{x \in S_1 \,|\, \pi(x) \text{ and } \pi(\Phi x) \text{ belong to different pieces } \gamma_1, \gamma_2\}$ the preservation is strict. A convenient parameterization of S_c is obtained in the following way. For $x \in S_c$ consider the segment of the billiard trajectory with endpoints $\pi(x)$ and $\pi(\Phi x)$: let s be the y-coordinate of the point of intersection of the segment with the y-axis and let α be the angle which the segment makes with the y-axis. Clearly (s, α) are smooth coordinates in S_c and the invariant probability measure in S has in (s, α) coordinates the form $\mathrm{const} |\sin \alpha| \, ds \, d\alpha$. S_c is in (s, α) coordinates a subset of the rectangle $R_0 = [-r, r] \times [0, 2\pi]$, which fills the rectangle as $\varepsilon \to 0$ [the explicit description of S_c in (s, α) coordinates is unacceptably involved].

By Proposition 2 the projectivization of $D_x \Phi$ at $x \in S_c$ has in the appropriate coordinates the form

$$f_1 = \frac{d_1 f_0 - d_1 L}{2 f_0 - 2L + d_1},$$

and it maps the sector $C(x) = \{f_0 \,|\, 0 \le f_0 \le d_0\}$ strictly into the sector $C(\Phi x) = \{f_1 \,|\, 0 \le f_1 \le d_1\}$ [L is the length of the segment between $\pi(x)$ and $\pi(\Phi x)$]. By straightforward computation we obtain that ζ from Theorem 2 is equal to

$$\zeta = \frac{(L - d_0)(L - d_1)}{(L - d_0 - d_1) L}$$

and

$$\frac{\sqrt{\zeta} + 1}{\sqrt{\zeta} - 1} = \sqrt{1 + \omega} + \sqrt{\omega},$$

where

$$\omega = \frac{(L - d_0 - d_1)L}{d_0 d_1}.$$

By elementary geometry we get

$$L - d_0 - d_1 = 2a|\sin\alpha| = 2\varepsilon|\sin\alpha|r.$$

(Note that by Proposition 5 $L - d_0 - d_1$ should depend only on α.)
Furthermore

$$d_{0,1} = \sqrt{r^2 - (s\sin\alpha \pm a\cos\alpha)^2}.$$

By Theorem 2 we have

$$\lambda_+ = \int_S \lambda_+ \, d\mu \geq \int_{S_c} \ln(\sqrt{1+\omega} + \sqrt{\omega}) \, d\mu.$$

We would like to use the approximation $\ln(\sqrt{1+\omega} + \sqrt{\omega}) \approx \sqrt{\omega}$ which is valid for small ω. The trouble is that actually ω is always unbounded on S_c. Nevertheless if we fix a rectangle $R_\delta = \{(s,\alpha) | -r + \delta \leq s \leq r - \delta, |\sin\alpha| \geq \delta\}$ for small $\delta > 0$, then for sufficiently small ε, $R_\delta \subset S_c$ and ω converges uniformly to zero on R_δ as $\varepsilon \to 0$. Thus for sufficiently small ε we get

$$\int_{R_\delta} \ln(\sqrt{1+\omega} + \sqrt{\omega}) \, d\mu \approx \int_{R_\delta} \sqrt{\omega} \, d\mu \geq \text{const}\sqrt{\varepsilon}.$$

So finally $\lambda_+ \geq \text{const}\sqrt{\varepsilon}$ for sufficiently small ε.

It is interesting that when ε is far from zero one can get estimates of λ_+ better than that of Theorem 2, using some norm to measure the expansion of vectors in the sector $C(x)$ under the action of $D_x\Phi$. But as it turns out for $\varepsilon \to 0$ they would give only the linear estimate (constε) so that Theorem 2 seems to yield the best asymptotic estimate.

Appendix C

Let us consider the billiard in a stadium-like domain Q of the Example 4 in Sect. 4, where γ_1, γ_2 are two halves of an ellipse cut along the longer axis so that in particular the longer axis of the ellipse is equal to $2w$. Let the shorter axis be equal to $2u$. The domain Q depends thus on three parameters: a, u, and w. We will show that if $a \geq \dfrac{w^2 - u^2}{u}$ then the billiard in Q has nonvanishing Lyapunov exponents. This is exactly the condition under which the periodic orbit supported by the segment of the x-axis inside Q is no longer a linearly stable orbit (the pieces γ_1 and γ_2 are sufficiently far apart – Proposition 3). In this example γ_1 and γ_2 are not convex scattering so that the bundle of sectors defined in Sect. 3 is not preserved. It turns out that another bundle of sectors is preserved.

Let $\Phi : S \to S$ be the standard section of the billiard flow and S_1 be the part of S over the convex pieces γ_1 and γ_2. As in Sect. 3 we consider the first return map $\Phi_1 : S_1 \to S_1$.

Let us consider the bundle of sectors $C(x) = \{f \mid 0 \leqq f \leqq d(x)\}$, $x \in S_1$, as defined in Sect. 3. If $\pi(x) \in \gamma_{1,2}$ and $\pi(\Phi x) \in \gamma_{2,1}$, then it can be shown by straightforward calculation that $D_x \Phi(C(x)) \subset C(\Phi x)$ $\left(\text{provided } a \geqq \dfrac{w^2 - u^2}{u}\right)$. Indeed we need to show that $d(x) + d(\Phi x)$ is not bigger than the length of the segment between $\pi(x)$ and $\pi(\Phi x)$. This can be done quite simply using Proposition 5.

If for $x \in S_1$, $\pi(\Phi x)$ belongs to a flat piece, then again $D\Phi_1(C(x)) \subset C(\Phi_1 x)$. This can be established purely geometrically if we think about reflection at a flat piece as merely passing freely into the reflected domain.

Let us consider a subset $S_2 \subset S_1$ consisting of $x \in S_1$ such that the segment between $\pi(x)$ and $\pi(\Phi x)$ intersects the segment between the foci of γ_i if $\pi(x) \in \gamma_i$, $i = 1, 2$. For $x \in S_1 \backslash S_2$ the segment between $\pi(x)$ and $\pi(\Phi x)$ is tangent to some ellipse confocal with γ_i if $\pi(x) \in \gamma_i$, $i = 1, 2$. Let for $x \in S_1 \backslash S_2$ $e(x)$ be the distance between $\pi(x)$ and the point of tangency.

Now we define another bundle of sectors $D(x)$, $x \in S_1$ by

$$D(x) = \begin{cases} C(x) & \text{if } x \in S_2 \\ \{f \mid 0 \leqq f \leqq e(x)\} & \text{if } x \in S_1 \backslash S_2. \end{cases}$$

Φ_1 preserves $D(x)$, $x \in S_1$. Indeed if $x \in S_1 \backslash S_2$ and $\Phi x \in S_1 \backslash S_2$, then the preservation follows from the fact that confocal ellipses form a family of caustics for the billiard in the ellipse and from Proposition 1. When $x \in S_1 \backslash S_2$ and $\Phi x \in S_2$, it is not difficult to show that $e(x) \leqq d(x)$, so that $D(x) \subset C(x)$ and the preservation of D follows from the preservation of C. The case $x \in S_2$ and $\Phi x \in S_1 \backslash S_2$ is similar [now $e(\Phi x) \geqq d(\Phi x)$ so that $D(\Phi x) \supset C(\Phi x)$]. The remaining cases [$\pi(\Phi x)$ belongs to a flat piece] are easy to consider using the reflected domains and we omit the details.

Thus $D(x)$, $x \in S_1$ is eventually strictly preserved by Φ_1, so that by Theorem 1 the billiard in Q has nonvanishing Lyapunov exponents.

This example shows that the convex scattering property of convex pieces is not necessary for nonvanishing of Lyapunov exponents. But it should be pointed out that in this example we used heavily the fact that the billiard in an ellipse is integrable which allowed us to control trajectories which have very many reflections at one half of the ellipse before crossing the billiard table to the other half. So it may well turn out that typically the convex scattering property is in some sense a necessary condition for nonvanishing of Lyapunov exponents.

We can use the above example to construct a billiard with Lyapunov exponents nonvanishing almost everywhere and with at least two ergodic components. For that purpose let Q_1 be the domain obtained from Q by removing interiors of two rectangles as shown in Fig. 11, where F_1, F_2 are foci of γ_1 and F_3, F_4 are foci of γ_2. In the limit when the width of the rectangles shrinks to zero we obtain a billiard in Q in which additionally the trajectory is reflected at two segments connecting F_1, F_3 and F_2, F_4, respectively.

The bundle of sectors $D(x)$, $x \in S_1$ defined above will be preserved also by the new Φ_1. There are many cases one has to consider to check this. We omit the details, but let us note that in the new billiard we will have extra reflections only at the sides of the rectangles which are flat pieces of the boundary of Q_1, and generally
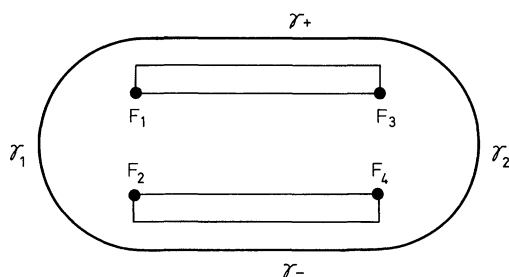
Fig. 11

we can expect the reflections at flat pieces only to improve the preservation of the sectors (see Sect. 3).

S_2 is an invariant subset for the new Φ_1, and clearly S_2 does not have full measure in S_1, so that the billiard in Q_1 has at least two ergodic components.

The above construction could be repeated if γ_1 and γ_2 would be pieces of an ellipse bigger than a half and could not be repeated if they would be smaller.

References

[Arn]   Arnold, V.I.: Mathematical methods of classical mechanics. Berlin, Heidelberg, New York: Springer 1978

[Ben]   Benettin, G.: Power law behavior of Lyapunov exponents in some conservative dynamical systems. Physica **13** D, 211–220 (1984)

[Ben–Str] Benettin, G., Strelcyn, J.-M.: Numerical experiments on the free motion of a point mass moving in a plane convex region: stochastic transition and entropy. Phys. Rev. A **17**, 773–785 (1978)

[Birk]   Birkhoff, G.: Lattice theory, AMS Colloquium Publications, Vol. 25, 3rd Ed. Chap. XVI (1968)

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

[Bun–Sin] Bunimovich, L.A., Sinai, Ya.G.: On a fundamental theorem in the theory of dispersing billiards. Math. USSR, Sb. **19**, 3, 407–423 (1973)

[C–F–S]   Cornfeld, I.P., Fomin, S.V., Sinai, Ya.G.: Ergodic theory. Berlin, Heidelberg, New York: Springer 1982, Chap. 6

[Douad]   Douady, R.: Applications du théorème des tores invariants. Thesis, University of Paris VII (1982)

[Kat–Str] Katok, A., Strelcyn, J.-M. In collaboration with Ledrappier, F., Przytycki, F.: Smooth maps with singularities: Invariant manifolds entropy and billiards. Lecture Notes in Mathematics (to appear)

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

[Led–Str] Ledrappier, F., Strelcyn, J.-M.: A proof of the estimation from below in Pesin's entropy formula. Ergodic Theory Dyn. Syst. **2**, 203–219 (1982)

[Math]   Mather, J.N.: Glancing billiards. Ergodic Theory Dyn. Syst. **2**, 397–403 (1982)

[Osel]   Oseledec, V.I.: The multiplicative ergodic theorem. The Lyapunov characteristic numbers of a dynamical system. Trans. Mosc. Math. Soc. **19**, 197–231 (1968)

[Pes]   Pesin, Ya.B.: Lyapunov characteristic exponents and smooth ergodic theory. Russ.
            Math. Surveys **32**, 4, 55–114 (1977)
[Rob]   Robnik, M.: Classical dynamics of a family of billiards with analytic boundaries. J.
            Phys. A: Math. Gen. **16**, 3971–3986 (1983)
[Sin]   Sinai, Ya.G.: Dynamical systems with elastic reflections. Ergodic properties of
            dispersing billiards. Russ. Math. Surveys **25**, 2, 137–189 (1970)
[Wojt 1]   Wojtkowski, M.: Invariant families of cones and Lyapunov exponents. Ergodic
            Theory Dyn. Syst. **5**, 145–161 (1985)
[Wojt 2]   Wojtkowski, M.: On uniform contraction generated by positive matrices. In:
            Random matrices and their applications. Cohen, J.E., Kesten, H., Newman, C.M.
            (eds.). Contemporary Mathematics Vol. 50. AMS 1986