ON ANOTHER EDGE OF DEFOCUSING: HYPERBOLICITY OF ASYMMETRIC LEMON BILLIARDS

LEONID BUNIMOVICH, HONG-KUN ZHANG, AND PENGFEI ZHANG

ABSTRACT. Defocusing mechanism provides a way to construct chaotic (hyperbolic) billiards with focusing components by separating all regular components of the boundary of a billiard table sufficiently *far away* from each focusing component. If all focusing components of the boundary of the billiard table are circular arcs, then the above separation requirement reduces to that all circles obtained by completion of focusing components are contained in the billiard table. In the present paper we demonstrate that a class of convex tables–*asymmetric lemons*, whose boundary consists of two circular arcs, generate hyperbolic billiards. This result is quite surprising because the focusing components of the asymmetric lemon table are *extremely close* to each other, and because these tables are perturbations of the first convex ergodic billiard constructed more than forty years ago.

1. INTRODUCTION

Billiards are dynamical systems generated by the motion of a point particle along the geodesics on a compact Riemannian manifold Q with boundary. Upon hitting the boundary of Q, the particle changes its velocity according to the law of elastic reflections. The studies of chaotic billiard systems were pioneered by Sinaĭ in his seminal paper [19] on dispersing billiards. A major feature of billiards which makes them arguably the most visual dynamical systems is that all their dynamical and statistical properties are completely determined by the shape of the billiard table Q and in fact by the structure of the boundary ∂Q.

Studies of convex billiards which started much earlier demonstrated that the convex billiards have regular dynamics and are even integrable. Such examples are billiards in circles or in squares, which everybody studied (without knowing that they study billiards) in a middle or in a high school. Jacobi proved integrability of billiards in ellipses by introducing elliptical coordinates in which the equations of motion are separated. Birkhoff conjectured that ellipses are the only integrable two dimensional smooth convex tables which generate completely integrable billiards. Later Lazutkin [12] proved that all two-dimensional convex billiards with sufficiently smooth boundary admit caustics and hence they can not be ergodic (see also [9]).

The first examples of hyperbolic and ergodic billiards with dispersing as well as with focusing components were constructed in [1]. A closer analysis of these examples allowed one to realize that there is another mechanism of chaos (hyperbolicity) than the mechanism of dispersing which generates hyperbolicity in dispersing billiards. This makes it possible to construct hyperbolic and ergodic billiards which do not have dispersing components on the boundary [2, 3]. Some billiards on convex tables also belong to this catergory. The *first* one was a table with boundary component consisting of a major arc and a chord connecting its two end points. Observe that this billiard is essentially equivalent to the one enclosed by two circular arcs symmetric with respect to the cutting chord. This billiard belongs to the class of (chaotic) flower-like billiards. The boundaries of these tables have the smoothness of order C^0. The stadium billiard (which became strangely much more

2000 *Mathematics Subject Classification.* 37D50.
Key words and phrases. Chaotic billiards, defocusing mechanism, hyperbolicity, asymmetric lemons, curvature, continued fractions, first return map.

popular than the others) appeared as one of many examples of convex ergodic billiards with smoother (C^1) boundary. Observe that in a flower-like billiard, all the circles generated by the corresponding petals (circular arcs) completely lie within the billiard table (flower).

In the present paper we consider (not necessarily small) perturbations of the first class of chaotic focusing billiards, i.e., with boundary made of a major arc and a chord. While keeping the major arc, we replace the chord (a neutral or zero curvature component of the boundary) by a circular arc with smaller curvature. This type of billiards were constructed in [5], with certain numerical results. We prove rigourously that the corresponding billiard tables generate hyperbolic billiards under the conditions that the chord is not too long, and the new circular arc has sufficiently small curvature. More precisely, we assume that the length of the chord does not exceed the radius of the circular component of the boundary, see Theorem 2. This condition is a purely technical one, and we conjecture that the hyperbolicity holds without this restriction.

It is worthwhile to recall that defocusing mechanism of chaos was one of a few examples where discoveries of new mechanisms/laws of nature were made in mathematics rather than in physics. No wonder that physicists did not believe that, even though rigorous proofs were present, until they check it numerically. After that, stadia and other focusing billiards were built in physics labs over the world. However, intuitive "physical" understanding of this mechanism always was that defocusing must occur after any reflection off the focusing part of the boundary. Indeed, there are just a few very special classes of chaotic (hyperbolic) billiards where this condition was violated (see e.g. [4]). However, these billiards were specially constructed to get hyperbolic billiards. To the contrary, hyperbolicity of asymmetric lemons came as a complete surprise to everybody. This class of chaotic billiards forces physicists as well as mathematicians to reconsider their understanding of this fundamental mechanism of chaos.

Our billiards can also be viewed as far-reaching generalizations of classical lemon-type billiards. The lemon billiards were introduced by Heller and Tomsovic [10] in 1993, by taking the intersection of two unit disks, while varying the distance between their centers, say b. This family of billiards have been extensively studied numerically in physics literature in relation with the problems of quantum chaos (see [16, 18]). The coexistence of the elliptic islands and chaotic region has also been observed numerically for all of the lemon tables as long as $b \neq 1$. Therefore, the lemon table with $b = 1$ is the *only* possible billiard system with complete chaos in this family. See also [13, 14] for the studies of classical and quantum chaos of lemon-type billiards with general quadric curves.

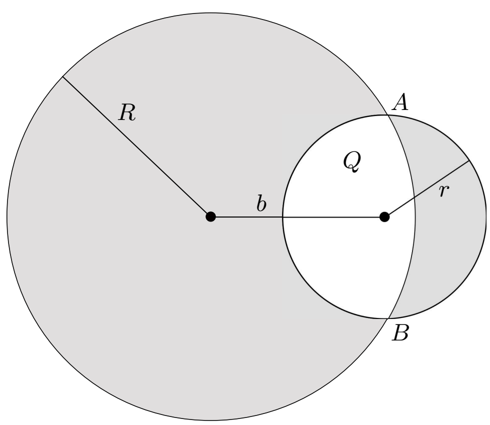

FIGURE 1. Basic construction of an asymmetric lemon table $Q(b, R)$.

The lemon tables were embedded into a 3-parameter family–the *asymmetric lemon billiards* in [5], among which the ergodicity is *no longer* an exceptional phenomenon. More precisely, let $Q(r, b, R)$ be the billiard table obtained as the intersection of a disk D_r of radius r with another disk D_R of radius $R > r$, where $b > 0$ measures the distance between the centers of these two disks (see Fig. 1). Without loss of generality, we will assume $r = 1$ and denote the lemon table by $Q(b, R) = Q(1, b, R)$. Restrictions on b and R will be specified later on to ensure the hyperbolicity of the billiard systems on these asymmetric lemon tables. On one hand, these billiard tables have extremely simple shape, as the boundary of the billiard table $Q(b, R)$ only consists of two circular arcs. Yet on the other hand, these systems already exhibit rich dynamical behaviors, as it has been numerically observed in [5] that there exists an infinite strip $\mathcal{D} \subset [1, \infty) \times [0, \infty)$, such that for any $(b, R) \in \mathcal{D}$, the billiard system on $Q(b, R)$ is ergodic.

In this paper we give a rigorous proof of the hyperbolicity on a class of asymmetric lemon billiards $Q(b, R)$. Our approach is based on the analysis of continued fractions generated by the billiard orbits, which were introduced by Sinaĭ [19], see also [1]. Continued fractions are intrinsic objects for billiard systems, and therefore they often provide sharper results than those one gets by the abstract cone method, which deals with hyperbolic systems of any nature and does not explore directly some special features of billiards. In fact, already in the fundamental paper [19] invariant cones were immediately derived from the structure of continued fractions generated by dispersing billiards. The study of asymmetric lemon billiards demonstrates that defocusing mechanism can generate chaos in much more general setting than it was thought before.

1.1. **Main results.** Let $b > 0$ and $R > 1$ be two positive numbers, $Q = Q(b, R)$ be the asymmetric lemon table obtained by intersecting the unit disc D_1 with D_R, where $b > 0$ measures the distance between the two centers of D_1 and D_R. Let $\Gamma = \partial Q$ be the boundary of Q, and Γ_1 be the circular boundary component of Q on the disk D_1, and Γ_R be the circular boundary component of Q on the the disk D_R. Let A and B be the points of intersection of Γ_1 and Γ_R, whom we will call the *corner* points of Q. It is easy to see the following two extreme cases: $Q(b, R) = D_1$ when $b \le R - 1$, and $Q(b, R) = \emptyset$ when $b \ge 1 + R$. So we will assume $b \in (R - 1, R + 1)$ for the rest of this paper.

We first review some properties of periodic points of the billiard system on $Q(b, R)$. It is easy to see that there is no fixed point, and exactly one period 2 orbit colliding with both arcs[1], say $\mathcal{O}_2$, which moves along the segment passing through both centers. The following result is well known, see [20] for example.

Lemma 1. *The orbit $\mathcal{O}_2$ is hyperbolic if $1 < b < R$, is parabolic if $b = 1$ or $b = R$, and is elliptic if $b < 1$ or $b > R$.*

It has been observed in [5] that under the condition $b < 1$ or $b > R$, $\mathcal{O}_2$ is actually nonlinearly stable (see also [17]). That is, the orbit $\mathcal{O}_2$ is surrounded by some islands. Therefore, the following is a necessary condition such that the billiard system on $Q(b, R)$ is hyperbolic.

(A0) The parameters (b, R) satisfy $\max\{R - 1, 1\} < b < R$.

In this paper we prove that the billiard system on $Q(b, R)$ is completely hyperbolic under the assumption **(A0)** and some general assumptions **(A1)**–**(A3)**. As these assumptions are rather technical, we will state them in Section 4.1.

Theorem 1. *Let $Q(b, R)$ be an asymmetric lemon table satisfying the assumptions **(A0)**–**(A3)**. Then the billiard system on $Q(b, R)$ is hyperbolic.*

The proof of Theorem 1 is given in Section 4.2.

To provide more intuitions for these conditions, we consider a special class of asymmetric lemon billiards. We first cut the unit disk D_1 by a chord with end point A and B, and let Γ_1 be the major arc of the unit circle with end points A, B. Denoted by Q_0 the larger part of the disk whose boundary contains Γ_1. By the classical defocusing mechanism, the billiard on Q_0 is hyperbolic and

[1]There are some other period 2 orbits which only collide with Γ_1. These orbits are parabolic.

ergodic. Now replace the chord with a circular arc on the circle D_R, for some large radius R. Note that the distance between the two centers is given by $b = (R^2 - |AB|^2/4)^{1/2} - (1 - |AB|^2/4)^{1/2}$. The resulting table $Q(R) := Q(b, R)$ can be viewed as a perturbation of Q_0. The following theorem shows that the billiard system on Q_0, while being nonuniformly hyperbolic, is robustly hyperbolic under suitable perturbations.

Theorem 2. *Let Γ_1 be the major arc of the unit circle whose end points A, B satisfy $|AB| < 1$. Then there exists $R_* > 1$ such that for each $R \geq R_*$, the billiard system on the table $Q(R)$ with two corners at A, B is hyperbolic.*

The proof of Theorem 2 is given in Section 4.3.

Remark 1. The hyperbolicity of the billiard system guarantees that a typical (infinitesimal) wave front in the phase space grows exponentially fast along the iterations of the billiard map. Therefore, one can say that these billiards in Theorem 2 still demonstrate the defocusing mechanism. However, the circle completing each of the boundary arcs of the table $Q(b, R)$ *contains* the entire table. Therefore, the defocusing mechanism can generate hyperbolicity even in the case when the separation condition is strongly violated.

The assumption $|AB| < 1$ in Theorem 2 is a purely technical one, and it is used only once in the proof of Theorem 2 to ensure $n^* \geq 6$ (see the definition of n^* in §4.3). Clearly this assumption $|AB| < 1$ is stronger than the assumption that Γ_1 is a major arc. We conjecture that as long as Γ_1 is a major arc, the billiard system on $Q(R)$ is completely hyperbolic for any large enough R.

Conjecture 1. Fix two points A and B on ∂D_1 such that Γ_1 is a major arc. Then the billiard system on $Q(b, R)$ is hyperbolic if the center of the disc D_R lies out side of the table.

To ease a task of reading we provide hereby a list of notations that we use in this paper.

The List of Notations

$Q(r, b, R) = D_r \cap D_R$	the asymmetric lemon table as the intersection of D_r with D_R. We usually set $r = 1$ and denote it by $Q(b, R) = Q(b, 1, R)$. We also denote $Q(R) = Q(b, R)$ if b is determined by R.
$\Gamma = \partial Q(b, R)$	the boundary of $Q(b, R)$, which consists of two arcs: Γ_1 and Γ_R.
$\mathcal{M} = \Gamma \times [-\pi/2, \pi/2]$	the phase space with coordinate $x = (s, \varphi)$, which consists of $\mathcal{M}_1$ and $\mathcal{M}_R$.
$\mathcal{F}$	the billiard map on the phase space $\mathcal{M}$ of $Q(b, R)$.
$\mathcal{S}_1$	the set of points in $\mathcal{M}$ at where $\mathcal{F}$ is not well defined or not smooth.
$\mathcal{B}^{\pm}(V)$	the curvature of the orthogonal transversal of the beam of lines generated by a tangent vector $V \in T_x\mathcal{M}$ before and after the reflection at x, respectively.
$d(x) = \rho \cdot \cos\varphi$	the half of the chord cut out by the trajectory of the billiard orbit in the disk D_ρ, where $\rho \in \{1, R\}$ is given by $x \in \mathcal{M}_\rho$.
$\mathcal{R}(x) = -\frac{2}{d(x)}$	the reflection parameter that measures the increment of the curvature after reflection.
$\tau(x)$	the distance between the current position of x with the next reflection with Γ.
$\chi^{\pm}(\mathcal{F}, x)$	the Lyapunov exponents of $\mathcal{F}$ at the point x.

$\eta(x)$	the number of successive reflections of x on the arc Γ_σ, where σ is given by $x \in \mathcal{M}_\sigma$.
$\hat{\mathcal{M}}_1 = \mathcal{M}_1 \backslash \mathcal{F}\mathcal{M}_1$	the set of points that first enter $\mathcal{M}_1$. Similarly we define $\hat{\mathcal{M}}_R$.
$M_n = \{x \in \hat{\mathcal{M}}_1 : \eta(x) = n\}$	the set of points in $\hat{\mathcal{M}}_1$ having n reflections on Γ_1 before hitting Γ_R.
$\hat{F}(x) = \mathcal{F}^{j_0+j_1+2}x$	the first return map of $\mathcal{F}$ on the subset $\hat{\mathcal{M}}_1$, where j_0 is the number of reflections of x on Γ_1, and j_1 is the number of reflections of $x_1 = \mathcal{F}^{j_0+1}x$ on Γ_R.
$\hat{\tau}_k = \tau_k - j_k\hat{d}_k - j_{k+1}\hat{d}_{k+1}$	a notation for short, where $\tau_k = \tau(x_k)$, $d_k = d(x_k)$, $\hat{d}_k = \frac{d_k}{j_k+1}$.
$\hat{\mathcal{R}}_k = -\frac{2}{\hat{d}_k}$	a notation for short.
$M = \bigcup_{n\geq 0} \mathcal{F}^{\lceil n/2 \rceil} M_n$	a subset of $\mathcal{M}_1$. Compare with the set $\hat{\mathcal{M}}_1 = \bigcup_{n\geq 0} M_n$.
$F(x) = \mathcal{F}^{i_0+i_1+i_2+2}x$	the first return map of $\mathcal{F}$ on M, where i_0 is the number of reflections of x on Γ_1, i_1 is the number of reflections of $x_1 = \mathcal{F}^{i_0+1}x$ on Γ_R, and i_2 is the number of reflections of $x_2 = \mathcal{F}^{i_1+1}x_1$ on Γ_1 before entering M.
$\bar{\tau}_k = \tau_k - i_k\hat{d}_k - d_{k+1}$	a notation for short. Only $\bar{\tau}_1$ is used in this paper.

2. Preliminaries for general convex billiards

Let $Q \subset \mathbb{R}^2$ be a compact convex domain with piecewise smooth boundary, $\mathcal{M}$ be the space of unit vectors based at the boundary $\Gamma := \partial Q$ pointing inside of Q. The set $\mathcal{M}$ is endowed with the topology induced from the tangent space TQ. A point $x \in \mathcal{M}$ represents the initial status of a particle, which moves along the ray generated by x and then makes an elastic reflection after hitting Γ. Denote by $x_1 \in \mathcal{M}$ the new status of the particle right after this reflection. The billiard map, denoted by $\mathcal{F}$, maps each point $x \in \mathcal{M}$ to the point $x_1 \in \mathcal{M}$. Note that each point $x \in \mathcal{M}$ has a natural coordinate $x = (s, \varphi)$, where $s \in [0, |\Gamma|)$ is the arc-length parameter of Γ (oriented counterclockwise), and $\varphi \in [-\pi/2, \pi/2]$ is the angle formed by the vector x with the inner normal direction of Γ at the base point of x. In particular, the phase space $\mathcal{M}$ can be identified with a cylinder $\Gamma \times [-\pi/2, \pi/2]$. The billiard map preserves a smooth probability measure μ on $\mathcal{M}$, where $d\mu = (2|\Gamma|)^{-1} \cdot \cos\varphi \, ds \, d\varphi$.

For our lemon table $Q = Q(b, R)$, the boundary $\Gamma = \partial Q$ consists of two parts Γ_1 and Γ_R. Collision vector starting from a corner point at A or B has s-coordinate $s = 0$ or $s = |\Gamma_1|$, respectively. Then we can view $\mathcal{M}$ as the union of two closed rectangles:

$$\mathcal{M}_1 := \{(s, \varphi) \in \mathcal{M} : 0 \leq s \leq |\Gamma_1|\} \quad \text{and} \quad \mathcal{M}_R := \{(s, \varphi) \in \mathcal{M} : |\Gamma_1| \leq s \leq |\Gamma|\}.$$

For any point $x = (s, \varphi) \in \mathcal{M}$, we define $d(x) = \cos\varphi$ if $x \in \mathcal{M}_1$, and $d(x) = R\cos\varphi$ if $x \in \mathcal{M}_R$. Geometrically, the quantity $2d(x)$ is the length of the chord in the complete disk (D_1 or D_R) decided by the trajectory of x.

Let $\mathcal{S}_0 = \{(s, \varphi) \in \mathcal{M} : s = 0 \text{ or } s = |\Gamma_1|\}$ be the set of post-reflection vectors $x \in \mathcal{M}$ that pass through one of the corners A or B. We define $\mathcal{S}_1 = \mathcal{S}_0 \cup \mathcal{F}^{-1}\mathcal{S}_0$ as the set of points on which $\mathcal{F}$ is not well-defined. Note that $\mathcal{F}^{-1}\mathcal{S}_0$ consists of 4 monotone curves $\varphi = \varphi_i(s)$ $(1 \leq i \leq 4)$ in $\mathcal{M}$. Moreover, we define $\mathcal{S}_{-1} := \mathcal{S}_0 \cup \mathcal{F}\mathcal{S}_0$. The set $\mathcal{S}_{\pm 1}$ is called the *singular set* of the billiard map $\mathcal{F}^{\pm 1}$.

A way to understand chaotic billiards lies in the study of infinitesimal families of trajectories. More precisely, let $x \in \mathcal{M} \setminus \mathcal{S}_1$, $V \in T_x\mathcal{M}$, $\gamma : (-\varepsilon_0, \varepsilon_0) \to \mathcal{M}$, $t \mapsto \gamma(t) = (s(t), \varphi(t))$ be a smooth curve for some $\varepsilon_0 > 0$, such that $\gamma(0) = x$ and $\gamma'(0) = V$. Clearly the choice of such a smooth curve is not unique. Each point in the phase space $\mathcal{M} \subset TQ$ is a unit vector on the billiard table

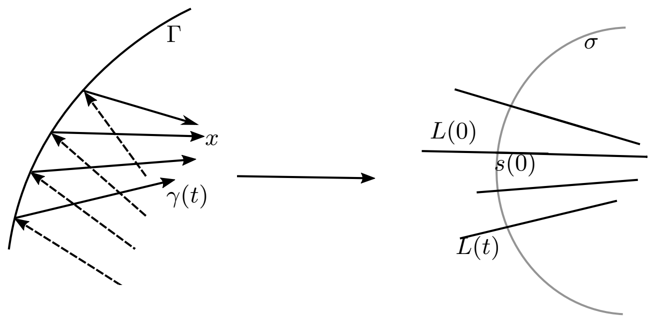

FIGURE 2. A bundle of lines generated by γ, and the cross-section σ.

Q. Let $L(t)$ be the line that passes through the vector $\gamma(t)$, see Fig. 2. Putting these lines together, we get a beam of post-reflection lines, say $\mathcal{W}^+$, generated by the path γ. Let σ be the orthogonal cross-section of this bundle passing through the point $s(0) \in \Gamma$. Then the post-reflection curvature of the tangent vector V, denoted by $\mathcal{B}^+(V)$, is defined as the curvature of σ at the point $s(0)$. Similarly we define the pre-reflection curvature $\mathcal{B}^-(V)$ (using the beam of dashed lines in Fig. 2).

Note that $\mathcal{B}^\pm(V)$ depend only on V, and are independent of the choices of curves tangent to V. These two quantities are related by the equation

$$\mathcal{B}^+(V) - \mathcal{B}^-(V) = \mathcal{R}(x), \tag{2.1}$$

where $\mathcal{R}(x) := -2/d(x)$ is the *reflection parameter* introduced in [19], see also [6, §3.8]. In fact, (2.1) is the well-known *Mirror Equation* in geometric optics. Note that $\mathcal{R}(x) > 0$ on dispersing components and $\mathcal{R}(x) < 0$ on focusing components of the boundary ∂Q. Since we mainly use $\mathcal{B}^-(V)$ in this paper, we drop the minus sign, simply denote it by $\mathcal{B}(V) = \mathcal{B}^-(V)$.

Let $\tau(x)$ be the distance from the current position of x to the next reflection with Γ. According to (2.1), one gets the evolution equation for the curvatures of the pre-reflection wavefronts of V and its image $V_1 = D\mathcal{F}(V)$ at $\mathcal{F}x$:

$$\mathcal{B}_1(V) := \mathcal{B}(V_1) = \cfrac{1}{\tau(x) + \cfrac{1}{\mathcal{R}(x) + \mathcal{B}(V)}}. \tag{2.2}$$

More generally, let $x \in \mathcal{M} \setminus \mathcal{S}_1$ be a point with $\mathcal{F}^k x \notin \mathcal{S}_1$ for all $1 \le k \le n$, $V \in T_x\mathcal{M}$ be a nonzero tangent vector, and $V_n = D\mathcal{F}^n V$ be its forward iterations. Then by iterating the formula (2.2), we get

$$\mathcal{B}(V_n) = \cfrac{1}{\tau(x_{n-1}) + \cfrac{1}{\mathcal{R}(x_{n-1}) + \cfrac{1}{\tau(x_{n-2}) + \cfrac{1}{\mathcal{R}(x_{n-2}) + \cfrac{1}{\ddots + \cfrac{1}{\tau(x) + \cfrac{1}{\mathcal{R}(x) + \mathcal{B}(V)}}}}}}} \tag{2.3}$$

with $x_k = \mathcal{F}^k x$. See also [6, §3.8] for Eq. (2.2) and (2.3).

For convenience, we introduce the standard notations for continued fractions [11]. In the following, we will denote $[a] := \frac{1}{a}$. The reader should not be confused by the integral part of a, which is never used in this paper[2].

Definition 1. Let a_n, $n \geq 0$ be a sequence of real numbers. The finite continued fraction $[a_1, a_2, \cdots, a_n]$ is defined inductively by:

$$[a_1] = \frac{1}{a_1},\ [a_1, a_2] = \frac{1}{a_1 + [a_2]},\ \cdots,\ [a_1, a_2, \cdots, a_n] = \frac{1}{a_1 + [a_2, \cdots, a_n]}.$$

Moreover, we denote $[a_0; a_1, a_2, \cdots, a_n] = a_0 + [a_1, a_2, \cdots, a_n]$.

Using this notation, we see that the evolution (2.3) of the curvatures of $V_n = D\mathcal{F}^n(V)$ can be re-written as

$$\mathcal{B}(V_n) = [\tau(x_{n-1}), \mathcal{R}(x_{n-1}), \tau(x_{n-2}), \mathcal{R}(x_{n-2}), \cdots, \tau(x), \mathcal{R}(x) + \mathcal{B}(V)]. \tag{2.4}$$

Note that Eq. (2.4) is a recursive formula and hence can be extended *formally* to an infinite continued fraction.

We will need the following basic properties of continued fractions to perform some reductions. Let $x = [a_1, \cdots, a_n]$ be a finite continued fraction. Then we can combine two finite continued fractions in the following ways:

$$[b_1, \cdots, b_n + x] = [b_1, \cdots, b_n, a_1, \cdots, a_n], \tag{2.5}$$

$$[b_1, \cdots, b_n, x] = [b_1, \cdots, b_n + a_1, a_2, \cdots, a_n], \tag{2.6}$$

$$[b_1, \cdots, b_n, 0, a_1, \cdots, a_n] = [b_1, \cdots, b_n + a_1, a_2, \cdots, a_n]. \tag{2.7}$$

Proposition 1. *Suppose a, b, c are real numbers such that $B := a + c + abc \neq 0$. Then the relation*

$$[\cdots, x, a, b, c, y, \cdots] = [\cdots, x + A, B, C + y, \cdots] \tag{2.8}$$

holds for any finite or infinite continued fractions, where $A = \frac{bc}{B}$ and $C = \frac{ab}{B}$.

Let Q be a bounded domain with piecewise smooth boundary, $\mathcal{F}$ be the billiard map on the phase space $\mathcal{M}$ over Q. Then the limit $\chi^+(\mathcal{F}, x) = \lim\limits_{n\to\infty} \dfrac{1}{n} \log \|D_x \mathcal{F}^n\|$, whenever it exists, is said to be a *Lyapunov exponent* of the billiard map $\mathcal{F}$ at the point x. Since $\mathcal{F}$ preserves the smooth measure μ, the other Lyapunov exponent at x is given by $\chi^-(\mathcal{F}, x) = -\chi^+(\mathcal{F}, x)$. Then the point x is said to be hyperbolic, if $\chi^+(\mathcal{F}, x) > 0$. Moreover, the billiard map $\mathcal{F}$ is said to be (completely) *hyperbolic*, if μ-almost every point $x \in \mathcal{M}$ is a hyperbolic point. By Oseledets *Multiplicative Ergodic Theorem*, we know that $\chi^+(\mathcal{F}, x)$ exists for μ-a.e. $x \in \mathcal{M}$, and there exists a measurable splitting $T_x\mathcal{M} = E^u_x \oplus E^s_x$ over the set of hyperbolic points, see [6].

It is well known that the hyperbolicity of a billiard map is related to the *convergence* of the continued fraction given in Eq. (2.4) as $n \to \infty$. In particular, the following proposition reveals the relations between them. See [1, 6, 19].

Proposition 2. *Let $x \in \mathcal{M}$ be a hyperbolic point of the billiard map $\mathcal{F}$. Then the curvature $\mathcal{B}^u(x) := \mathcal{B}(V^u_x)$ of a unit vector $V^u_x \in E^u_x$ is given by the following infinite continued fraction:*

$$\mathcal{B}^u(x) = [\tau(x_{-1}), \mathcal{R}(x_{-1}), \tau(x_{-2}), \mathcal{R}(x_{-2}), \cdots, \tau(x_{-n}), \mathcal{R}(x_{-n}), \cdots].$$

Finally we recall an invariant property for consecutive reflections on focusing boundary components by comparing the curvatures of the iterates of different tangent vectors. Given two distinct points a and b on the unit circle $\mathbb{S}^1$, denote by (a, b) the interval from a to b counterclockwise. Given three distinct points a, b, c on $\mathbb{S}^1$, denote by $a \prec b \prec c$ if $b \in (a, c)$. Endow $\mathbb{R} \cup \{\infty\} \simeq \mathbb{S}^1$ with the relative position notation $\prec$ on $\mathbb{S}^1$.

[2] We use the ceiling function $\lceil t \rceil = \min\{n \in \mathbb{Z} : n \geq t\}$ in §3.2, and the floor function $\lfloor t \rfloor = \max\{n \in \mathbb{Z} : n \leq t\}$ in §4.3.

Proposition 3 ([8]). *Let $X, Y, Z \in T_x\mathcal{M}$ be three tangent vectors at $x \in \mathcal{M}$ satisfying $\mathcal{B}(X) \prec \mathcal{B}(Y) \prec \mathcal{B}(Z)$. Then for each $n \in \mathbb{Z}$, the iterates $D\mathcal{F}^n X$, $D\mathcal{F}^n Y$ and $D\mathcal{F}^n Z$ satisfy*

$$\mathcal{B}(D\mathcal{F}^n X) \prec \mathcal{B}(D\mathcal{F}^n Y) \prec \mathcal{B}(D\mathcal{F}^n Z).$$

3. Continued fractions for asymmetric lemon billiards

In this section we construct two induced maps of the billiard system $(\mathcal{M}, \mathcal{F})$ on two different but closely related subsets of the phase space $\mathcal{M}$, and then study the evolutions of continued fractions of the curvatures $\mathcal{B}(V)$ under these induced maps. Let $Q(b, R)$ be an asymmetric lemon table obtained as the intersection of a disk of radius 1 with a disk of radius $R > 1$, $\Gamma = \partial Q$, $\mathcal{M} = \Gamma \times [-\pi/2, \pi/2]$ be the phase space of the billiard map on Q. Note that $\mathcal{M}$ consists of two parts: $\mathcal{M}_1 := \Gamma_1 \times [-\pi/2, \pi/2]$ and $\mathcal{M}_R = \Gamma_R \times [-\pi/2, \pi/2]$, the sets of points in $\mathcal{M}$ based on the arc Γ_1 and Γ_R, respectively. Assume that Γ_1 is a major arc.

For $\sigma \in \{1, R\}$, $x \in \mathcal{M}_\sigma \setminus \mathcal{S}_1$, let $\eta(x)$ be the number of successive reflections of x on the arc Γ_σ. That is,

$$\eta(x) = \sup\{n \geq 0 : \mathcal{F}^k x \in \mathcal{M}_\sigma \text{ for all } k = 0, \cdots, n\}. \tag{3.1}$$

For example, $\eta(x) = 0$ if $\mathcal{F}x \notin \mathcal{M}_\sigma$, and $\eta(x) = \infty$ if $\mathcal{F}^k x \in \mathcal{M}_\sigma$ for all $k \geq 0$. Let $N = \{x \in \mathcal{M}_1 : \eta(x) = \infty\}$. One can easily check that each point $x \in N$ is either periodic or belongs to the boundary $\{(s, \varphi) \in \mathcal{M}_1 : \varphi = \pm\pi/2\}$. In particular, N is a null set with $\mu(N) = 0$.

Let $\hat{\mathcal{M}}_1 := \{x \in \mathcal{M}_1 : \mathcal{F}^{-1}x \notin \mathcal{M}_1\}$ be the set of points first entering $\mathcal{M}_1$. Similarly we define $\hat{\mathcal{M}}_R$. The restriction of η on $\hat{\mathcal{M}}_1$ induces a measurable partition of $\hat{\mathcal{M}}_1$, whose cells are given by $M_n := \eta^{-1}\{n\} \cap \hat{\mathcal{M}}_1$ for all $n \geq 0$. Each cell M_n contains all first reflection vectors on the arc Γ_1 that will experience exactly n reflections on Γ_1 before hitting Γ_R. Then it is easy to check that

$$\mathcal{M}_1 = N \cup \bigcup_{n \geq 0} \bigcup_{0 \leq k \leq n} \mathcal{F}^k M_n. \tag{3.2}$$

3.1. The first induced map of $\mathcal{F}$ on $\mathcal{M}_1$.

Let $x_0 \in \hat{\mathcal{M}}_1$, and $j_0 = \eta(x_0)$ be the numbers of successive reflections of x_0 on Γ_1. Similarly, we denote $x_1 = \mathcal{F}^{j_0+1}x_0$, and $j_1 = \eta(x_1)$. Then the first return map $\hat{F}$ of $\mathcal{F}$ on $\hat{\mathcal{M}}_1$ is given by

$$\hat{F}x := \mathcal{F}^{j_0+j_1+2}x.$$

Note that the similar induced systems appeared in many references about billiards with convex boundary components, see [6, 7, 15]. In the systems considered in these references, the induced systems were shown to be (uniformly) hyperbolic. However, for our billiard systems on $Q(b, R)$, it is rather difficult to prove the hyperbolicity for this type of the induced map. Thus we introduce a new induced map in the next subsection. To make a comparison, we next investigate the properties of the induced map $(\hat{\mathcal{M}}_1, \hat{F})$.

To simplify the notations, we denote by $\tau_0 := \tau(\mathcal{F}^{j_0}x_0)$ the length of the free path of $\mathcal{F}^{j_0}x_0$, and by $\tau_1 := \tau(\mathcal{F}^{j_1}x_1)$ the length of the free path of $\mathcal{F}^{j_1}x_1$. Moreover, let $d_k = d(x_k)$, $\hat{d}_k = \frac{d_k}{j_k+1}$, $\hat{\mathcal{R}}_k = -2/\hat{d}_k$, $\hat{\tau}_k = \tau_k - j_k\hat{d}_k - j_{k+1}\hat{d}_{k+1}$, for $k = 0, 1$. Note that $\hat{d}_k = d_k$ and $\hat{\mathcal{R}}_k = \mathcal{R}_k$ if $j_k = 0$, and $\hat{\tau}_k = \tau_k$ if $j_k = j_{k+1} = 0$.

Using the relations in Proposition 1, we can reduce the long continuous fraction to a shorter one:

Lemma 2. *Let $x \in \hat{\mathcal{M}}_1$, $V \in T_x\mathcal{M}$ and $\hat{V}_1 = D\hat{F}(V)$. Then $\mathcal{B}(\hat{V}_1)$ is given by the continued fraction:*

$$\mathcal{B}(\hat{V}_1) = [\tau_1 - j_1\hat{d}_1, \hat{\mathcal{R}}_1, \hat{\tau}_0, \hat{\mathcal{R}}_0, -j_0\hat{d}_0, \mathcal{B}(V)]. \tag{3.3}$$

Remark 2. Note that in the case $j_0 = 0$ and $j_1 = 0$, $\hat{F}x = \mathcal{F}^2 x$, and the relation (3.3) reduces to the formula (2.4) with $n = 2$: $\mathcal{B}(\hat{V}_1) = [\tau_1, \mathcal{R}_1, \tau_0, \mathcal{R}_0 + \mathcal{B}(V)]$.

Proof. Suppose a point $x \in \mathcal{M}$ have m consecutive reflections on a circular arc Γ_σ, where $\sigma \in \{1, R\}$. That is, $\mathcal{F}^i x \in \mathcal{M}_\sigma$, for $i = 0, \cdots, m$. In this case we always have

$$d(\mathcal{F}^i x) = d(x),\ \mathcal{R}(\mathcal{F}^i x) = -2/d(x),\ 0 \le i \le m,\ \text{and}\ \tau(\mathcal{F}^i x) = 2d(x),\ 0 \le i < m.$$

Then for any $V \in T_x\mathcal{M}_\sigma$,

$$\begin{aligned}\mathcal{B}(D\mathcal{F}^{m+1}V) &= [\tau(\mathcal{F}^m x), \underbrace{\mathcal{R}(x), 2d(x), \mathcal{R}(x), \cdots, 2d(x)}_{m\ \text{times}}, \mathcal{R}(x) + \mathcal{B}(V)] \\ &= [\tau(\mathcal{F}^m x), \mathcal{R}(x)/2, -2m \cdot d(x), \mathcal{R}(x)/2 + \mathcal{B}(V)],\end{aligned} \tag{3.4}$$

see [6, §8.7]. Applying this reduction process for each of the two reflection series on Γ_1 and Γ_R respectively, we see that

$$\mathcal{B}(\hat{V}_1) = [\tau_1, \mathcal{R}_1/2, -2j_1 d_1, \mathcal{R}_1/2, \tau_0, \mathcal{R}_0/2, -2j_0 d_0, \mathcal{R}_0/2 + \mathcal{B}(V)]. \tag{3.5}$$

Now we rewrite the last segment in (3.5) as $[\cdots, \mathcal{R}_0/2 + \mathcal{B}(V)] = [\cdots, \mathcal{R}_0/2, 0, \mathcal{B}(V)]$ by Eq. (2.7). Then applying Eq. (2.8) to the segment $(a, b, c) = (\mathcal{R}_0/2, -2j_0 d_0, \mathcal{R}_0/2)$ in Eq. (3.5), we get

$$\begin{aligned} B_0 &:= a + c + abc = \mathcal{R}_0 - (\mathcal{R}_0/2)^2 \cdot 2j_0 d_0 = -\frac{2 + 2j_0}{d_0} = -\frac{2}{\hat{d}_0} = \hat{\mathcal{R}}_0, \\ A_0 &:= \frac{bc}{B_0} = -2j_0 d_0 \cdot \mathcal{R}_0/2 \cdot (-\frac{\hat{d}_0}{2}) = -j_0 \cdot \hat{d}_0, \\ C_0 &:= \frac{ab}{B_0} = -2j_0 d_0 \cdot \mathcal{R}_0/2 \cdot (-\frac{\hat{d}_0}{2}) = -j_0 \cdot \hat{d}_0. \end{aligned}$$

Putting them together with (3.5), we have

$$\begin{aligned}\mathcal{B}(\hat{V}_1) &= [\tau_1, \mathcal{R}_1/2, -2j_1 d_1, \mathcal{R}_1/2, \tau_0, \mathcal{R}_0/2, -2j_0 d_0, \mathcal{R}_0/2, 0, \mathcal{B}(V)] \\ &= [\tau_1, \mathcal{R}_1/2, -2j_1 d_1, \mathcal{R}_1/2, \tau_0 + A_0, B_0, 0 + C_0, \mathcal{B}(V)] \\ &= [\tau_1, \mathcal{R}_1/2, -2j_1 d_1, \mathcal{R}_1/2, \tau_0 - j_0 \cdot \hat{d}_0, \hat{\mathcal{R}}_0, -j_0 \cdot \hat{d}_0, \mathcal{B}(V)].\end{aligned} \tag{3.6}$$

Similarly we can apply Eq. (2.8) to the segment $(\mathcal{R}_1/2, -2j_1 d_1, \mathcal{R}_1/2)$, and get $B_1 = \hat{\mathcal{R}}_1$, $A_1 = C_1 = -j_1 \cdot \hat{d}_1$. Then we can continue the computation from (3.6) and get

$$\begin{aligned}\mathcal{B}(\hat{V}_1) &= [\tau_1, \mathcal{R}_1/2, -2j_1 d_1, \mathcal{R}_1/2, \tau_0 - j_0 \cdot \hat{d}_0, \hat{\mathcal{R}}_0, -j_0 \cdot \hat{d}_0, \mathcal{B}(V)] \\ &= [\tau_1 + A_1, B_1, \tau_0 + C_1 - j_0 \cdot \hat{d}_0, \hat{\mathcal{R}}_0, -j_0 \cdot \hat{d}_0, \mathcal{B}(V)] \\ &= [\tau_1 - j_1 \hat{d}_1, \hat{\mathcal{R}}_1, \hat{\tau}_0, \hat{\mathcal{R}}_0, -j_0 \hat{d}_0, \mathcal{B}(V)],\end{aligned}$$

where $\hat{d}_k = \frac{d_k}{j_k + 1}$, $\hat{\mathcal{R}}_k = -2/\hat{d}_k$ for $k = 0, 1$, and $\hat{\tau}_0 = \tau_0 - j_0 \hat{d}_0 - j_1 \hat{d}_1$. This completes the proof. □

It is clear that the formula in Eq. (3.3) is recursive. For example, let $\hat{V}_2 = D\hat{F}^2(V)$, $\tau_2 = \tau_0(\hat{F}x)$ and $\tau_3 = \tau_1(\hat{F}x)$ be the lengths of the free paths for $\hat{F}x$. Then we have

$$\mathcal{B}(\hat{V}_2) = [\tau_3 - j_3 \hat{d}_3, \hat{\mathcal{R}}_3, \hat{\tau}_2, \hat{\mathcal{R}}_2, \hat{\tau}_1, \hat{\mathcal{R}}_1, \hat{\tau}_0, \hat{\mathcal{R}}_0, -j_0 \hat{d}_0, \mathcal{B}(V)].$$

More generally, using the backward iterates $x_{-n} = \hat{F}^{-n} x$ of x, and the related notations (for example, $d_{-2n} = d_0(x_{-n})$, and $d_{1-2n} = d_1(x_{-n})$), we get a formal continued fraction

$$[\tau_0 - j_0 \hat{d}_0, \hat{\mathcal{R}}_0, \hat{\tau}_{-1}, \hat{\mathcal{R}}_{-1}, \hat{\tau}_{-2}, \cdots, \hat{\mathcal{R}}_{1-2n}, \hat{\tau}_{-2n}, \hat{\mathcal{R}}_{-2n}, \hat{\tau}_{-2n-1}, \cdots] \tag{3.7}$$

where $\hat{d}_k = \frac{d_k}{j_k + 1}$, $\hat{\mathcal{R}}_k = -2/\hat{d}_i$, $\hat{\tau}_k = \tau_k - j_k \hat{d}_k - j_{k+1} \hat{d}_{k+1}$, for each $k \le -1$.

Remark 3. In the dispersing billiard case, each entry of the continued fraction (2.4) is positive. Then Seidel–Stern Theorem (see [11]) implies that the limit of (2.4) (as $n \to \infty$) always exists, since the total time $\tau_0 + \cdots + \tau_n \to \infty$. For the reduced continued fraction (3.7), it is clear that $\hat{\mathcal{R}}_{-n} < 0$ for all $n \geq 0$. Moreover, $\hat{\mathcal{R}}_{-2n} \leq \mathcal{R}(x_{-n}) \leq -2$ for each $n \geq 1$, since the radius of the small disk is set to $r = 1$. Therefore, $\sum_n \hat{\mathcal{R}}_{-n}(x)$ always diverges. However, Seidel–Stern Theorem is not applicable to determine the convergence of (3.7), since the terms $\hat{\tau}_k$ have no definite sign. This is the reason that we need to introduce a new return map (M, F) instead of using $(\hat{M}_1, \hat{F})$ to investigate the hyperbolicity.

3.2. **New induced map and its analysis.** Denote by $\lceil t \rceil$ the smallest integer larger than or equal to the real number t. We consider a new subset, which consists of "middle" sliding reflections on Γ_1. More precisely, let

$$M := \bigcup_{n\geq 0} \mathcal{F}^{\lceil n/2\rceil} M_n = M_0 \cup \mathcal{F} M_1 \cup \mathcal{F} M_2 \cup \cdots \cup \mathcal{F}^{\lceil n/2\rceil} M_n \cup \cdots . \tag{3.8}$$

Let F be the first return map of $\mathcal{F}$ with respect to M. Clearly the induced map $F : M \to M$ is measurable and preserves the conditional measure μ_M of μ on M, which is given by $\mu_M(A) = \mu(A)/\mu(M)$, for any Borel measurable set $A \subset M$.

For each $x \in M$, we introduce the following notations:

(1) let $i_0 = \eta(x) \geq 0$ be the number of forward reflections of x on Γ_1, $\tau_0 := \tau(\mathcal{F}^{i_0} x)$ be the distance between the last reflection on Γ_1 and the first reflection on Γ_R. Let $d_0 := d(x)$ and $\mathcal{R}_0 := \mathcal{R}(x)$, which stay the same along this series of reflections on Γ_1;

(2) let $i_1 = \eta(x_1) \geq 0$ be the number of reflections of $x_1 = \mathcal{F}^{i_0+1}x$ on Γ_R, $\tau_1 := \tau(\mathcal{F}^{i_1} x_1)$ be the distance between the last reflection on Γ_R and the next reflection on Γ_1. Let $d_1 = d(x_1)$, and $\mathcal{R}_1 = \mathcal{R}(x_1)$, which stay the same along this series of reflections on Γ_R;

(3) let $i_2 = \lceil \eta(x_2)/2 \rceil \geq 0$ be the number[3] of reflections of $x_2 = \mathcal{F}^{i_1+1}x_1$ on Γ_1 till the return to M. Let $d_2 = d(x_2)$, $\mathcal{R}_2 = \mathcal{R}(x_2)$, which stay the same along this series of reflections on Γ_1.

Then the first return map F on M is given explicitly by $Fx = \mathcal{F}^{i_0+i_1+i_2+2}x$. Note that $x_2 = Fx$ and $d_2 = d(Fx)$ (in above notations).

The following result is the analog of Lemma 2 on the reduction of continued fractions for the new induced return map F:

Lemma 3. *Let $x \in M$, $V \in T_x\mathcal{M}$ and $V_1 = DF(V)$. Then $\mathcal{B}(V_1)$ is given by the continued fraction:*

$$\mathcal{B}(V_1) = [d_2, \frac{2i_2}{d_2}, \bar{\tau}_1, \hat{\mathcal{R}}_1, \hat{\tau}_0, \hat{\mathcal{R}}_0, -i_0\hat{d}_0, \mathcal{B}(V)], \tag{3.9}$$

where $\hat{d}_k = \frac{d_k}{j_k+1}$, $\hat{\mathcal{R}}_k = -2/\hat{d}_k$, $\hat{\tau}_0 = \tau_0 - i_0\hat{d}_0 - i_1\hat{d}_1$, and $\bar{\tau}_1 = \tau_1 - i_1\hat{d}_1 - d_2$.

Note that (3.9) may not be as pretty as (3.3). It involves three types of quantities: the original type (i_0, i_2 and d_2), the first variation ($\hat{R}_k$, $\hat{\tau}_0$ and $\hat{d}_0$), and the second variation $\bar{\tau}_1$. The quantity $\hat{d}_2$ appears only in the intermediate steps of the proof, and does not appear in the final formula (3.9).

Remark 4. Note that in the case $i_2 = 0$, (3.9) reduces to (3.3):

$$\begin{aligned}\mathcal{B}(V_1) &= [d_2, 0, \bar{\tau}_1, \hat{\mathcal{R}}_1, \hat{\tau}_0, \hat{\mathcal{R}}_0, -i_0\hat{d}_0, \mathcal{B}(V)] = [d_2 + \bar{\tau}_1, \hat{\mathcal{R}}_1, \hat{\tau}_0, \hat{\mathcal{R}}_0, -i_0\hat{d}_0, \mathcal{B}(V)] \\ &= [\tau_1 - i_1\hat{d}_1, \hat{\mathcal{R}}_1, \hat{\tau}_0, \hat{\mathcal{R}}_0, -i_0\hat{d}_0, \mathcal{B}(V)].\end{aligned}$$

[3] Our choice of $i_2 = \lceil \eta(x_2)/2 \rceil$ in Item (3), instead of using $\eta(x_2)$, is due to the fact that M is the union of the sets $\mathcal{F}^{\lceil n/2\rceil} M_n$, $n \geq 0$.

Proof. We first consider an intermediate step. That is, let $\hat{V}_1 = D\mathcal{F}^{i_0+i_1+2}(V)$ and $V_1 = D\mathcal{F}^{i_2}(\hat{V}_1)$. Applying the reduction process (3.4) for each of the series of reflection of lengths (i_0, i_1) (as in the proof of Lemma 2), we get that

$$\mathcal{B}(\hat{V}_1) = [\tau_1 - i_1\hat{d}_1, \hat{\mathcal{R}}_1, \hat{\tau}_0, \hat{\mathcal{R}}_0, -i_0\hat{d}_0, \mathcal{B}(V)].$$

This completes the proof of (3.9) when $i_2 = 0$ (see Remark 4). In the following we assume $i_2 \geq 1$. In this case we have $\mathcal{B}(V_1) = [\underbrace{2d_2, \mathcal{R}_2, \cdots, 2d_2, \mathcal{R}_2}_{i_2 \text{ times}} + \mathcal{B}(\hat{V}_1)]$. To apply the reduction (3.4), we need to consider

$$\frac{1}{\mathcal{B}(V_1) + \mathcal{R}_2} = [\underbrace{\mathcal{R}_2, 2d_2, \cdots, \mathcal{R}_2, 2d_2}_{i_2 \text{ times}}, \mathcal{R}_2 + \mathcal{B}(\hat{V}_1)]. \tag{3.10}$$

Then we apply the reduction (3.4) to Eq. (3.10) and get $\dfrac{1}{\mathcal{B}(V_1) + \mathcal{R}_2} = [\mathcal{R}_2/2, -i_2\hat{d}_2, \mathcal{R}_2/2 + \mathcal{B}(\hat{V}_1)]$, which is equivalent to

$$\mathcal{B}(V_1) = -\mathcal{R}_2/2 + [-i_2\hat{d}_2, \mathcal{R}_2/2 + \mathcal{B}(\hat{V}_1)] = [0, -\mathcal{R}_2/2, -i_2\hat{d}_2, \mathcal{R}_2/2, 0, \mathcal{B}(\hat{V}_1)]. \tag{3.11}$$

Applying the relation (2.8) to the segment $(a, b, c) = (-\mathcal{R}_2/2, -2i_2d_2, \mathcal{R}_2/2)$, we get

$$\begin{aligned} B &:= a + c + abc = 0 + (\mathcal{R}_2/2)^2 \cdot 2i_2d_2 = \frac{2i_2}{d_2}, \\ A &:= \frac{bc}{B} = -2i_2d_2 \cdot \mathcal{R}_2/2 \cdot \frac{d_2}{2i_2} = d_2, \\ C &:= \frac{ab}{B} = 2i_2d_2 \cdot \mathcal{R}_2/2 \cdot \frac{d_2}{2i_2} = -d_2. \end{aligned}$$

Putting them together with Eq. (3.11), we have

$$\mathcal{B}(V_1) = [d_2, \frac{2i_2}{d_2}, -d_2, \mathcal{B}(\hat{V}_1)] = [d_2, \frac{2i_2}{d_2}, \bar{\tau}_1, \hat{\mathcal{R}}_1, \hat{\tau}_0, \hat{\mathcal{R}}_0, -i_0\hat{d}_0, \mathcal{B}(V)],$$

where $\bar{\tau}_1 = \tau_1 - i_1\hat{d}_1 - d_2$ follows from (2.6). This completes the proof of (3.9). □

4. Hyperbolicity of asymmetric lemon billiards

In this section we first list several general sufficient conditions that ensure the hyperbolicity of the asymmetric lemon-type billiards (see the statement below and the proof of Theorem 1), then we verify these conditions for a set of asymmetric lemon tables. Let $Q(b, R)$ be an asymmetric lemon table satisfying (A0), that is, $\max\{R-1, 1\} < b < R$. Let M be the subset introduced in §3.2. We divide the set M into three disjoint regions X_k, $k = 0, 1, 2$, which are given by

(a) $X_0 = \{x \in M : i_1(x) \geq 1\}$;
(b) $X_1 = \{x \in M : i_1(x) = 0, \text{ and } i_2(x) \geq 1\}$;
(c) $X_2 = \{x \in M : i_1(x) = i_2(x) = 0\}$.

We first make a simple observation:

Lemma 4. *For each x with $i_1(x) = 0$, one has $\tau_0 + \tau_1 > d_0 + d_2$.*

Proof. Suppose $i_1(x) = 0$, and p_1 be the reflection point of x_1 on Γ_R. Then we take the union of $Q(b, R)$ with its mirror, say $Q^*(b, R)$, along the tangent line L of Γ_R at p_1, and extend the pre-collision path of x_1 beyond the point p_1, which will intersect $\partial Q^*(b, R)$ at the mirror point of the reflection point p_2 of $\mathcal{F}x_1$, say p_2^* (see Fig .3). Clearly the distance $|p_2^* - p_1| = |p_2 - p_1| = \tau_1$. By the assumption that $b > R - 1$, one can see that the tangent line L cuts out a major arc on ∂D_1 (clearly larger than Γ_1), and the point p_2^* lies outside of the unit disk D_1. Therefore, $\tau_0 + \tau_1 > 2d_0$. Similarly, we have $\tau_0 + \tau_1 > 2d_2$. Putting them together, we get $\tau_0 + \tau_1 > d_0 + d_2$. □

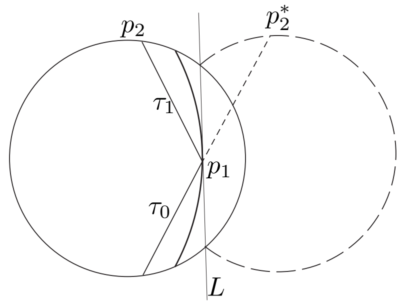

FIGURE 3. The case with $i_1(x) = 0$: there is only one reflection on Γ_R.

4.1. **Main Assumptions and their analysis.** In the following we list the assumptions on X_i's.

(**A1**) For $x \in X_0$: $i_0 \geq 1$, $i_2 \geq 1$ and

$$\tau_0 < (1 - \frac{1}{2(1+i_0)})d_0 + \frac{i_1}{1+i_1}d_1, \quad \tau_1 < (1 - \frac{1}{2i_2})d_2 + \frac{i_1}{1+i_1}d_1; \tag{4.1}$$

(**A2**) For $x \in X_1$: $\frac{d_0}{1+i_0} < d_1$ and $\tau_0 + \tau_1 < (1 - \frac{1}{2(1+i_0)})d_0 + d_1 + (1 - \frac{1}{2i_2})d_2$;

(**A3**) For $x \in X_2$: $\tau_0 \leq d_1/2$.

To prove Theorem 1, it suffices to verify hyperbolicity of the first return map $F : M \to M$, obtained by restricting $\mathcal{F}$ on M. For each $x \in M$, $V \in T_x\mathcal{M}$, we let $\mathcal{B}(V) = \mathcal{B}^-(V)$ be the pre-reflection curvature of V. Note that $\mathcal{B}(V)$ determines V uniquely up to a scalar. Let $V_x^d \in T_x\mathcal{M}$ be the unit vector corresponding to the incoming beam with curvature $\mathcal{B}(V_x^d) = 1/d(x)$, and $V_x^p \in T_x\mathcal{M}$ be the unit vector corresponding to the parallel incoming beam $\mathcal{B}(V_x^p) = 0$, respectively.

Proposition 4. *Let $x \in M$, i_k, d_k and $\hat{d}_k$, $k = 0, 1, 2$ be the corresponding quantities of x given in §3.2, and $Fx = \mathcal{F}^{i_0+i_1+i_2+2}x$ be the first return map of $\mathcal{F}$ on M. Then we have*

(I). $0 < \mathcal{B}(DF(V_x^d)) < 1/d_2$ *if one of the following conditions holds:*

- (**D1**) $\tau_1 - i_1\hat{d}_1 - d_2 + [-\frac{2}{\hat{d}_1}, \tau_0 - d_0 - i_1\hat{d}_1] > 0$,
- (**F1**) $\tau_1 - i_1\hat{d}_1 - d_2 + [-\frac{2}{\hat{d}_1}, \tau_0 - d_0 - i_1\hat{d}_1] < -\frac{d_2}{2i_2}$.

(II). $0 < \mathcal{B}(DF(V_x^p)) < 1/d_2$ *if one of the following conditions holds:*

- (**D2**) $\tau_1 - i_1\hat{d}_1 - d_2 + [-\frac{2}{\hat{d}_1}, \tau_0 - (1 - \frac{1}{2(1+i_0)})d_0 - i_1\hat{d}_1] > 0$,
- (**F2**) $\tau_1 - i_1\hat{d}_1 - d_2 + [-\frac{2}{\hat{d}_1}, \tau_0 - (1 - \frac{1}{2(1+i_0)})d_0 - i_1\hat{d}_1] < -\frac{d_2}{2i_2}$.

(III). $0 < \mathcal{B}(DF(V_x^p)) < \mathcal{B}(DF(V_x^d)) < 1/d_2$ *if one of the following conditions holds:*

(**P1**) *One of the paired conditions*[4] *(that is,* ***(D1)-(D2)*** *or* ***(F1)-(F2)****) holds and*

$$[\tau_1 - i_1\hat{d}_1 - d_2, -\frac{2}{\hat{d}_1}, \tau_0 - (1 - \frac{1}{2(1+i_0)})d_0 - i_1\hat{d}_1] > [\tau_1 - i_1\hat{d}_1 - d_2, -\frac{2}{\hat{d}_1}, \tau_0 - d_0 - i_1\hat{d}_1]. \tag{4.2}$$

(**P2**) ***(D1)-(F2)*** *hold.*

[4]In the following we will use the term **(D1)-(D2)**, which is short for "both conditions **(D1)** and **(D2)**". Similarly, we use the term **(D1)-(D2)-(P1)**, which is short for "all three conditions **(D1)**, **(D2)** and **(P1)**"

The statements of Proposition 4 is rather technical, but the proof is quite straightforward. Geometrically, it gives different criterions when the cone bounded by V_x^p and V_x^d is mapped under DF to the cone bounded by V_{Fx}^p and V_{Fx}^d, see Proposition 5 for more details.

Proof. Note that $\mathcal{B}(V_x^d) = 1/d_0(x)$ and $\mathcal{B}(V_x^p) = 0$. So the curvatures of $DF(V_x^d)$ and $DF(V_x^p)$ are given by

$$\begin{aligned}\mathcal{B}(DF(V_x^d)) &= [d_2, \frac{2i_2}{d_2}, \bar{\tau}_1, -\frac{2}{\hat{d}_1}, \hat{\tau}_0, -\frac{2}{\hat{d}_0}, -i_0\hat{d}_0 + d_0] = [d_2, \frac{2i_2}{d_2}, \bar{\tau}_1, -\frac{2}{\hat{d}_1}, \hat{\tau}_0 - \hat{d}_0] \\ &= [d_2, \frac{2i_2}{d_2}, \bar{\tau}_1, -\frac{2}{\hat{d}_1}, \tau_0 - d_0 - i_1\hat{d}_1]; \\ \mathcal{B}(DF(V_x^p)) &= [d_2, \frac{2i_2}{d_2}, \bar{\tau}_1, -\frac{2}{\hat{d}_1}, \hat{\tau}_0, -\frac{2}{\hat{d}_0}] = [d_2, \frac{2i_2}{d_2}, \bar{\tau}_1, -\frac{2}{\hat{d}_1}, \hat{\tau}_0 - \frac{\hat{d}_0}{2}] \\ &= [d_2, \frac{2i_2}{d_2}, \bar{\tau}_1, -\frac{2}{\hat{d}_1}, \tau_0 - pd_0 - i_1\hat{d}_1], \text{ where } p = 1 - \frac{1}{2(1+i_0)}.\end{aligned}$$

Here we use the facts that $\hat{\tau}_0 = \tau_0 - i_0\hat{d}_0 - i_1\hat{d}_1$ and $[\cdots, a, b, 0] = [\cdots, a]$ whenever $b \neq 0$. Then it is easy to see that (I) $0 < \mathcal{B}(DF(V_x^d)) < 1/d_2$ is equivalent to

$$[\frac{2i_2}{d_2}, \bar{\tau}_1, -\frac{2}{\hat{d}_1}, \tau_0 - d_0 - i_1\hat{d}_1] > 0. \tag{4.3}$$

Moreover, (4.3) holds if and only if one of the following conditions holds:

- $[\bar{\tau}_1, -\frac{2}{\hat{d}_1}, \tau_0 - d_0 - i_1\hat{d}_1] > 0$, which corresponds to **(D1)**;
- $0 > [\bar{\tau}_1, -\frac{2}{\hat{d}_1}, \tau_0 - d_0 - i_1\hat{d}_1] > -\frac{2i_2}{d_2}$, which corresponds to **(F1)**.

We can derive **(D2)** and **(F2)** from (II) in the same way.

Condition (III) $0 < \mathcal{B}(DF(V_x^p)) < \mathcal{B}(DF(V_x^d)) < 1/d_2$ is equivalent to

$$[\frac{2i_2}{d_2}, \bar{\tau}_1, -\frac{2}{\hat{d}_1}, \tau_0 - pd_0 - i_1\hat{d}_1] > [\frac{2i_2}{d_2}, \bar{\tau}_1, -\frac{2}{\hat{d}_1}, \tau_0 - d_0 - i_1\hat{d}_1] > 0. \tag{4.4}$$

Then (4.4) holds if and only if one of the following conditions holds:

- $0 < [\bar{\tau}_1, -\frac{2}{\hat{d}_1}, \tau_0 - pd_0 - i_1\hat{d}_1] < [\bar{\tau}_1, -\frac{2}{\hat{d}_1}, \tau_0 - d_0 - i_1\hat{d}_1]$, which is **(D1)-(D2)-(P1)**;
- $[\bar{\tau}_1, -\frac{2}{\hat{d}_1}, \tau_0 - pd_0 - i_1\hat{d}_1] < [\bar{\tau}_1, -\frac{2}{\hat{d}_1}, \tau_0 - d_0 - i_1\hat{d}_1] < -\frac{d_2}{2i_2}$, which is **(F1)-(F2)-(P1)**;
- $[\bar{\tau}_1, -\frac{2}{\hat{d}_1}, \tau_0 - d_0 - i_1\hat{d}_1] > 0$ and $[\bar{\tau}_1, -\frac{2}{\hat{d}_1}, \tau_0 - pd_0 - i_1\hat{d}_1] < -\frac{d_2}{2i_2}$, which is **(D1)-(F2)**, or equivalently, **(P2)**.

This completes the proof of the proposition. □

Remark 5. It is worth pointing out the following observations:

(1). The conditions **(F1)** and **(F2)** are empty when $i_2 = 0$.

(2). The arguments in the proof of Proposition 4 apply to general billiards with several circular boundary components. That is, for any consecutive circular components that a billiard orbit passes, say Γ_k, $k = 0, 1, 2$, let i_k be the number of reflections of the orbit on Γ_k. Then we have the same characterizations for $Fx = \mathcal{F}^{i_0+i_1+i_2+2}x$.

(3). In the case when all arcs Γ_k lie on the same circle, the left-hand sides of **(D1)** and **(F1)** are zero, and (I) always fails. This is quite natural, since the circular billiard is well-known to be not hyperbolic, but parabolic.

The following is an application of Proposition 4 to asymmetric lemon tables satisfying Assumptions **(A0)**–**(A3)**.

Proposition 5. *Let $Q(b, R)$ be an asymmetric lemon billiard satisfying the assumptions **(A0)**–**(A3)**. Then for a.e. $x \in M$,*

$$0 = \mathcal{B}(V^p_{Fx}) < \mathcal{B}(DF(V^p_x)) < \mathcal{B}(DF(V^d_x)) < \mathcal{B}(V^d_{Fx}) = 1/d(Fx). \tag{4.5}$$

In comparing to Proposition 4, the statement of Proposition 5 is very clear, but the proof is quite technical and a little bit tedious. From the proof, one can see that $\frac{d_0}{1+i_0} < d_1$ is a very subtle assumption to guarantee $d_0 < pd_0 + \frac{d_1}{2}$. One can check that (4.5) may fail if $d_0 > pd_0 + \frac{d_1}{2}$. This is why we need $\frac{d_0}{1+i_0} < d_1$ in the assumption (**A2**).

Proof. It suffices to show that for a.e. point $x \in M$, one of the following combinations holds: **(D1)-(D2)-(P1)**, or **(F1)-(F2)-(P1)**, or **(D1)-(F2)**.

Case 1. Let $x \in X_0$. Then Eq. (4.1) implies that

$$\tau_0 - d_0 - i_1\hat{d}_1 < \tau_0 - (1 - \frac{1}{2(1+i_0)})d_0 - i_1\hat{d}_1 < 0, \quad \tau_1 - i_1\hat{d}_1 - d_2 < -\frac{d_2}{2i_2}.$$

Therefore, **(F1)-(F2)-(P1)** hold.

Case 2. Let $x \in X_1$. Note that $i_1 = 0$ and $\hat{d}_1 = d_1$. We denote $p = 1 - \frac{1}{2(1+i_0)}$ and $q = 1 - \frac{1}{2i_2}$, and rewrite the the corresponding conditions using $i_1 = 0$:

(**D1′**). $\tau_1 - d_2 + [-\frac{2}{d_1}, \tau_0 - d_0] > 0$, or equivalently, $\tau_1 - d_2 + \dfrac{1}{-\frac{2}{d_1} + \frac{1}{\tau_0 - d_0}} > 0$;

(**F1′**). $\tau_1 - d_2 + [-\frac{2}{d_1}, \tau_0 - d_0] < -\frac{d_2}{2i_2}$, or equivalently, $\tau_1 - qd_2 + \dfrac{1}{-\frac{2}{d_1} + \frac{1}{\tau_0 - d_0}} < 0$;

(**D2′**). $\tau_1 - d_2 + [-\frac{2}{d_1}, \tau_0 - pd_0] > 0$, or equivalently, $\tau_1 - d_2 + \dfrac{1}{-\frac{2}{d_1} + \frac{1}{\tau_0 - pd_0}} > 0$;

(**F2′**). $\tau_1 - d_2 + [-\frac{2}{d_1}, \tau_0 - pd_0] < -\frac{d_2}{2i_2}$, or equivalently, $\tau_1 - qd_2 + \dfrac{1}{-\frac{2}{d_1} + \frac{1}{\tau_0 - d_0}} < 0$;

(**P1′**). $[\tau_1 - d_2, -\frac{2}{d_1}, \tau_0 - pd_0] < [\tau_1 - d_2, -\frac{2}{d_1}, \tau_0 - d_0]$.

Note that $0 < \tau_0 < d_0 + d_1$. There are several subcases when τ_0 varies in $(0, d_0 + d_1)$, see Fig. 4.

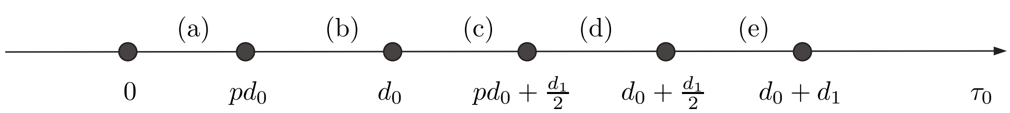

FIGURE 4. Subcases of Case 2 according to the values of τ_0. Note that $d_0 < pd_0 + \frac{d_1}{2}$, which follows from the assumption that $\frac{d_0}{1+i_0} < d_1$.

Subcase (a). Let $\tau_0 < pd_0$. Hence $\tau_1 > d_2$ and $0 < \frac{1}{\tau_1 - d_2} < \frac{1}{d_0 - \tau_0} < \frac{1}{pd_0 - \tau_0}$. Then we claim that **(D1′)-(D2′)-(P1′)** hold.

Proof of Claim. Note that $\tau_0 + \tau_1 > d_0 + d_2$, or equally $\tau_1 - d_2 > d_0 - \tau_0 > 0$. So $\frac{1}{\tau_1 - d_2} < \frac{1}{d_0 - \tau_0} < \frac{2}{d_1} + \frac{1}{d_0 - \tau_0}$ which implies (**D1′**). Similarly, $\frac{1}{\tau_1 - d_2} < \frac{1}{pd_0 - \tau_0} < \frac{2}{d_1} + \frac{1}{pd_0 - \tau_0}$ implies (**D2′**). If both terms in (**P1′**) are positive, then (**P1′**) is equivalent to $[-\frac{2}{d_1}, \tau_0 - pd_0] > [-\frac{2}{d_1}, \tau_0 - d_0]$. Analogously, if both terms in (**P1′**) are negative, then (**P1′**) is equivalent to $[\tau_0 - pd_0] < [\tau_0 - d_0]$, or equivalently, $\tau_0 - pd_0 > \tau_0 - d_0$, which holds trivially. This completes the proof of the claim.

Subcase (b). Let $pd_0 \le \tau_0 < d_0$. The proof of (**D1′**) in the previous case is still valid. For (**D2′**), we note that $\tau_0 < d_0 < pd_0 + \frac{d_1}{2}$ and hence $-\frac{2}{d_1} + \frac{1}{\tau_0 - pd_0} > 0$. So (**D2′**) follows. Since both terms are positive, (**P1′**) follows from $[-\frac{2}{d_1}, \tau_0 - pd_0] > 0 > [-\frac{2}{d_1}, \tau_0 - d_0]$.

Subcase (c). Let $d_0 \le \tau_0 < pd_0 + d_1/2$. There are two subcases:

- $\tau_1 > d_2$. The proof is similar to Case (b), and **(D1′)-(D2′)-(P1′)** follows.
- $\tau_1 \le d_2$. Note that $0 < \frac{1}{\tau_0 - pd_0} - \frac{2}{d_2} < \frac{1}{\tau_0 - d_0} - \frac{2}{d_2} < \frac{1}{d_2 - \tau_1}$, from which **(D1′)-(D2′)-(P1′)** follows.

Subcase (d). Let $pd_0 + d_1/2 \le \tau_0 < d_0 + d_1/2$, which is equivalent to $\frac{1}{\tau_0 - pd_0} - \frac{2}{d_1} < 0 < \frac{1}{\tau_0 - d_0} - \frac{2}{d_1}$. Using Assumption **(A2)** that $\tau_0 + \tau_1 > d_0 + d_2$, we see that (**D1′**) always holds (as in Subcase 2(c)). Using the condition $\tau_0 + \tau_1 < pd_0 + d_1 + qd_2$, we see that $\tau_1 - qd_2 < \frac{d_1}{2}$. There are two subcases:

- $\tau_1 - qd_2 < 0$. Then (**F2′**) holds trivially.
- $\tau_1 - qd_2 > 0$. Then $\frac{1}{\tau_1 - qd_2} > \frac{2}{d_1}$ implies (**F2′**).

Subcase (e). Let $\tau_0 \ge d_0 + d_1/2$. There are also two subcases:

- $\tau_1 - qd_2 < 0$. Then $\frac{1}{\tau_1 - qd_2} < 0 < \frac{2}{d_1} - \frac{1}{\tau_0 - d_0} < \frac{2}{d_1} - \frac{1}{\tau_0 - pd_0}$, and **(F1′)-(F2′)-(P1′)** hold.
- $\tau_1 - qd_2 > 0$. Then $\frac{1}{\tau_1 - qd_2} > \frac{2}{d_1}$ implies **(F1′)-(F2′)-(P1′)**.

Case 3. Let $x \in X_2$. Then $\tau_0 < d_1/2$ by assumption. The proof of Case 2 (a)-(c) also works this case, and **(D1′)-(D2′)-(P1′)** hold. □

4.2. **Proof of Theorem 1.** Let $\mathcal{S}_\infty = \bigcup_{n \in \mathbb{Z}} \mathcal{F}^{-n}\mathcal{S}_1$. Note that $\mu(\mathcal{S}_\infty) = 0$. Let $x \in M \backslash \mathcal{S}_\infty$. The push-forward $V_n^p(x) = DF^n(V^p_{F^{-n}x})$ and $V_n^d(x) = DF^n(V^d_{F^{-n}x})$ are well defined for all $n \ge 1$, and together generate two sequences of tangent vectors in $T_x\mathcal{M}$. By Proposition 3 and Lemma 5, we get the following relation inductively:

$$0 = \mathcal{B}(V_x^p) < \cdots < \mathcal{B}(V_n^p(x)) < \mathcal{B}(V_{n+1}^p(x)) \\ < \mathcal{B}(V_{n+1}^d(x)) < \mathcal{B}(V_n^d(x)) < \mathcal{B}(V_x^d) = 1/d(x).$$

Therefore $\mathcal{B}^u(x) := \lim_{n\to\infty} \mathcal{B}(V_n^d(x))$ exists, and $0 < \mathcal{B}^u(x) < 1/d(x)$. Let $E_x^u = \langle V_x^u \rangle$ be the corresponding subspace in $T_x\mathcal{M}$ for all $x \in M \backslash S_\infty$.

Let $\Phi : \mathcal{M} \to \mathcal{M}, (s, \varphi) \mapsto (s, -\varphi)$ be the time reversal map on $\mathcal{M}$. Then we have $\mathcal{F} \circ \Phi = \Phi \circ \mathcal{F}^{-1}$, and $F \circ \Phi = \Phi \circ F^{-1}$. In particular, $V_x^s = D\Phi(V^u_{\Phi x})$ satisfies $-1/d(\Phi x) < \mathcal{B}(V_x^s) = -\mathcal{B}(V^u_{\Phi x}) < 0$, which corresponds to a stable vector for each $x \in M \backslash \mathcal{S}_\infty$. Therefore, $T_x\mathcal{M} = E_x^u \oplus E_x^s$ for every point $x \in M \backslash \mathcal{S}_\infty$, and such a point x is a hyperbolic point of the induced billiard map F.

Next we show that the original billiard map $\mathcal{F}$ is hyperbolic. It suffices to show that the Lyapunov exponent $\chi^+(\mathcal{F}, x) > 0$ for a.e. $x \in M$, since $(\mathcal{M}, \mathcal{F}, \mu)$ is a (discrete) suspension over (M, F, μ_M) with respect to the first return time function, say ξ_M, which is given by $\xi_M(x) = i_0 + i_1 + i_2 + 2$ (see § 3.2). Note that $\int_M \xi_M d\mu_M = \frac{1}{\mu(M)}$. So the averaging return time $\overline{\xi}(x) = \lim_{k\to\infty} \frac{\xi_k(x)}{k}$ exists for a.e. $x \in M$, where $\xi_k(x) = \xi_M(x) + \cdots + \xi_M(F^{k-1}x)$ be the k-th return time of x to M. Moreover, $1 \le \overline{\xi}(x) < \infty$ for a.e. $x \in M$. Then for a.e. $x \in M$, we have

$$\chi^+(\mathcal{F}, x) = \lim_{n\to\infty} \frac{1}{n} \log \|D_x\mathcal{F}^n\| = \lim_{k\to\infty} \frac{k}{\xi_k(x)} \cdot \frac{1}{k} \log \|D_x F^k\| = \frac{1}{\overline{\xi}(x)} \cdot \chi^+(F, x) > 0.$$

This completes the proof. □

4.3. **Proof of Theorem 2.** Let Γ_1 be a major arc of the unit circle with endpoints A, B satisfying $|AB| < 1$. For $R > 1$, it is easy to see that $b = (R^2 - |AB|^2/4)^{1/2} - (1 - |AB|^2/4)^{1/2} > R - 1$.

Moreover, $b > 1$ as long as $R > 2$. In the following, we will assume $R > 2$. Therefore, (**A0**) holds. Then it suffices to show that the assumptions (**A1**)–(**A3**) hold for the tables $Q(R) = Q(b, R)$ considered in Theorem 2.

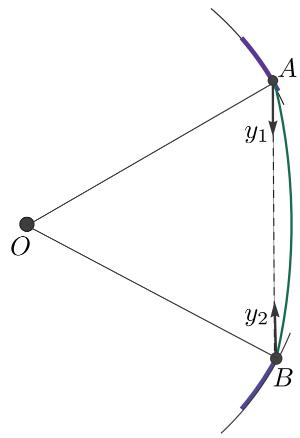

FIGURE 5. First restriction on R. The thickened pieces on Γ_1 are related to U.

Let $y_1 = \overrightarrow{AB}$ and $y_2 = \overrightarrow{BA}$ be the two points in the phase space $\mathcal{M}$ moving along the chord AB, see Fig. 5. Note that $y_i \in \mathcal{S}_0$. Since $|AB| < 1$, we have $\angle AOB < \frac{\pi}{3}$, or equally, $n^* := \lfloor \frac{2\pi}{\angle AOB} \rfloor \geq 6$, where $\lfloor t \rfloor$ is the largest integer smaller than or equal to t. Then for each $i = 1, 2$, there is a small neighborhood $U_i \subset \mathcal{M}$ of y_i, such that for any point $x \in U_i \cap \mathcal{M}_1 \backslash \mathcal{F}^{-1}\mathcal{M}_1$, one has

- $\mathcal{F}x$ enters $\mathcal{M}_R$ and stays on $\mathcal{M}_R$ for another i_1 iterates, where $i_1 = \eta(\mathcal{F}x)$,
- $\mathcal{F}^{i_1+2}x \in M_n$ for some $n \geq n^* - 1 = 5$.

See Section 3 for the definitions of $\eta(\cdot)$ and M_n. Let $U = (U_1 \cup U_2) \cap \mathcal{M}_1 \backslash \mathcal{F}^{-1}\mathcal{M}_1$. See Fig. 5, where the bold pieces on Γ_1 indicate the bases of U. Note that the directions of vectors in U are close to the vertical direction and are pointing to some points on Γ_R.

Before moving on to the verification of the assumptions (**A1**)–(**A3**), we need the following lemma. Note that for each $x \in M$, $\mathcal{F}^{i_0}x$ is the last reflection of $x \in M$ on Γ_1 before hitting Γ_R. That is, $\mathcal{F}^{i_0}x = \mathcal{F}^{-1}x_1$. See Fig. 6 and 7 for two illustrations.

Lemma 5. *For any $\epsilon > 0$, there exists $R_0 = R(\epsilon) > 2$ such that for any $R \geq R_0$, the following hold for the induced map $F : M \to M$ of the billiard table $Q(R)$. That is, for any point $x \in M$ with $d_1 \leq 4$,*

1. *$\mathcal{F}^{i_0}x \in U$ and $\mathcal{F}^{i_0+i_1+2}x \in \bigcup_{n\geq 5} M_n$;*
2. *the reflection points of $\mathcal{F}^{i_0}x$ and of $\mathcal{F}^{i_0+i_1+2}x$, both on Γ_1, are ϵ-close to the corners $\{A, B\}$;*
3. *the total length of the trajectory from $\mathcal{F}^{i_0}x$ to $\mathcal{F}^{i_0+i_1+2}x$ is bounded from above by $|AB| + \epsilon$.*

Note that the number 4 in '$d_1 \leq 4$' is chosen to simplify the presentation of the proof of Theorem 2. The above lemma holds for any number larger than $2|AB|$. In the following we will use $d(x) = \rho \cdot \sin\theta$, where $\theta = \frac{\pi}{2} - \varphi$ is the angle from the direction of x to the tangent line of Γ_ρ at x.

Proof of Lemma 5. Let p_1 be the reflection point of x_1 on Γ_R, and let θ_1 be the angle between the direction of x_1 with the tangent line L of Γ_R at p_1. Then $\sin\theta_1 \leq \frac{4}{R}$, since $d(x_1) = d_1 \leq 4$. Consider the vertical line that passes through p_1. Clearly the length of the chord on Γ_R cut out by this vertical line is less than $|AB|$, and hence less than 1. Therefore, the angle θ_2 between the vertical direction with the tangent direction of Γ_R at p_1 satisfies $\sin\theta_2 < \frac{1}{2R}$.

(1). Note that the angle θ between $\mathcal{F}^{i_0}x$ and the vertical direction is $\theta_1 \pm \theta_2$, which satisfies $|\sin\theta| \le \sin\theta_1 + \sin\theta_2 < \frac{5}{R}$. Then $\mathcal{F}^{i_0}x \in U$ if R is large enough. Moreover, $\mathcal{F}^{i_0+i_1+2}x = \mathcal{F}^{i_1+1}x_1 \in \bigcup_{n\ge 5} M_n$ by our choice of U.

(2). Let P be a point on Γ_1, and $\theta(A)$ be the angle between the vertical direction with the line PA. Similarly we define $\theta(B)$. Let θ_* be the minimal angle of $\theta(A)$ and $\theta(B)$ among all possible choices of P on Γ_1 that is not ϵ-close to either A or B. Then the angle between the vertical direction with the line connecting P to any point on Γ_R is bounded from below by θ_*. Let $x_1 \in \mathcal{M}_R$. The angle $\theta(x_1)$ between the direction of x_1 with the tangent direction of Γ_R at the base point of x_1 satisfies $\theta(x_1) \ge \theta_* - \arcsin\frac{1}{2R}$. Then $d_1 = R\cdot\sin\theta(x_1) \ge R\sin\theta_* - \frac{1}{2}$. So it suffices to assume $R_0 = \frac{5}{\sin\theta_*}$.

(3). By enlarging R_0 if necessary, the conclusion follows from (2), since the two reflections are close to the corners and the arc Γ_R is almost flat. □

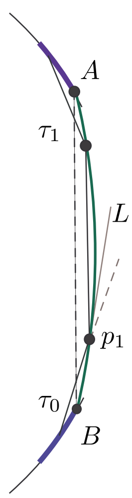

FIGURE 6. The case that there are multiple reflections on Γ_R.

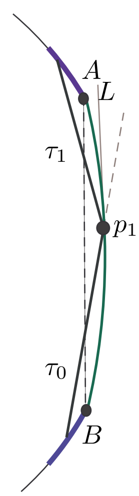

FIGURE 7. The case that there is a single reflection on Γ_R.

Now we continue the proof of Theorem 2.

(1). To verify (**A1**), we assume $x \in X_0$, which means $i_1 \ge 1$ (see Fig. 6 for an illustration). Then we have $d_1 < |AB| < 1$, which implies $i_0 \ge 2$ and $i_2 \ge 3$. Then we have $1 - \frac{1}{2(1+i_0)} \ge \frac{5}{6}$, and $1 - \frac{1}{2i_2} \ge \frac{5}{6}$. Therefore, a sufficient condition for (**A1**) is

$$\tau_0 < \frac{5}{6}d_0 + \frac{1}{2}d_1, \quad \tau_1 < \frac{5}{6}d_2 + \frac{1}{2}d_1. \tag{4.6}$$

Since $\tau_0, \tau_1 < 2d_1$, we only need to show that $\frac{3}{4}\tau_0 < \frac{5}{6}d_0$ and $\frac{3}{4}\tau_1 < \frac{5}{6}d_2$, or equivalently, $\tau_0 < \frac{10}{9}d_0$ and $\tau_1 < \frac{10}{9}d_2$. To this end, we first set $R_1 = 100$: then for each $R \ge R_1$, the angle θ of the vertical direction (assume that AB is vertical) and the tangent line of any point on Γ_R satisfies $\sin\theta < \frac{1}{2R} \le \frac{1}{200}$. Then the angle θ_0 between the vertical direction and the free path corresponding to τ_0 satisfy $\sin\theta_0 \le \sin\theta + \frac{d_1}{R} \le \frac{1}{100}$ (since $2d_1 < 1$). This implies $2d_0 \ge (1-0.05)\cdot|AB|$. On the other hand, by making R_1 even larger if necessary, we can assume $\tau_0 + \tau_1 + 2i_1 d_1 \le (1+0.05)\cdot|AB|$

for $x \in X_0$ (see Lemma 5). Then we have

$$\tau_0 < \frac{\tau_0 + 2d_1}{2} < \frac{1.05}{2} \cdot |AB| \leq \frac{1.05}{2} \cdot \frac{2d_0}{0.95} < \frac{10}{9} d_0.$$

Similarly, we have $\tau_1 < \frac{10}{9} d_2$.

(2). To verify (**A2**), we assume $x \in X_1$, which means $i_1 = 0$ and $i_2 \geq 1$ (see Fig. 7). Both relations in (**A2**) are trivial if $d_1 \geq 4$. In the case $d_1 < 4$, one must have $i_0 \geq 2$, $i_2 \geq 3$ (by the assumption $R > R_0$). Note that $|AB| < \tau_0 + \tau_1 < 4d_1$. Pick R_2 large enough, such that for any $R \geq R_2$, for any point $x_1 \in \mathcal{M}_R$ with $d_1 < 4$, the angle θ of the direction corresponding to x_1 with the vertical direction is bounded by 0.05 (recall that AB is vertical). Then we have[5] $2d_0 \leq \frac{4}{3} \cdot |AB|$. This implies

$$\frac{d_0}{1 + i_0} \leq \frac{1}{3} d_0 \leq \frac{2}{9} \cdot |AB| < \frac{8}{9} d_1 < d_1.$$

For the second inequality in (**A2**), by making R_2 even larger if necessary (see Lemma 5), we can assume that $\tau_0 + \tau_1 \leq 1.04 \cdot |AB|$, $2d_0, 2d_2 \geq 0.95 \cdot |AB|$. Combining with the fact that $d_1 > |AB|/4$, we have

$$\frac{5}{6} d_0 + d_1 + \frac{5}{6} d_2 > \frac{5}{6} \cdot \frac{0.95}{2} \cdot |AB| + \frac{1}{4} \cdot |AB| + \frac{5}{6} \cdot \frac{0.95}{2} \cdot |AB| > 1.04 \cdot |AB|.$$

Therefore, $\tau_0 + \tau_1 < \frac{5}{6} d_0 + d_1 + \frac{5}{6} d_2$, and (**A2**) follows.

(3). The condition (**A3**) on X_2 holds trivially, since $i_2 = 0$ implies that $d_1 \geq 4 > 2\tau_0$.

Then we set $R_* = \max\{R_i : i = 0, 1, 2\}$. This completes the proof of Theorem 2. □

Appendix A. A Detailed condition

Now we give an equivalent version of Proposition 4, whose formulation is longer but easier to check when proving the hyperbolicity of billiard systems. The statement of the proposition is technical, but the proof is straightforward.

Proposition 6. *Denote* $G_0 = \tau_0 - d_0 - \frac{i_1}{1+i_1} d_1$, $G_1 = \tau_1 - \frac{i_1}{1+i_1} d_1 - d_2$. *Then*

(I) $0 < \mathcal{B}(DF(V_x^d)) < 1$ *if one of the following holds:*

- **(D1a)** $G_1 > 0$ *and* $\frac{1}{G_0} > 2\frac{1+i_1}{d_1}$;
- **(D1b)** $G_1 > 0$ *and* $\frac{1}{G_0} + \frac{1}{G_1} < 2\frac{1+i_1}{d_1}$;
- **(D1c)** $G_1 < 0$ *and* $\frac{1}{G_0} + \frac{1}{G_1} < 2\frac{1+i_1}{d_1} < \frac{1}{G_0}$;
- **(F1a)** $G_1 + \frac{d_2}{2i_2} < 0$ *and* $\frac{1}{G_0} < 2\frac{1+i_1}{d_1}$;
- **(F1b)** $G_1 + \frac{d_2}{2i_2} < 0$ *and* $\frac{1}{G_0} + \frac{1}{G_1 + \frac{d_2}{2i_2}} > 2\frac{1+i_1}{d_1}$;
- **(F1c)** $G_1 + \frac{d_2}{2i_2} > 0$ *and* $\frac{1}{G_0} + \frac{1}{G_1 + \frac{d_2}{2i_2}} > 2\frac{1+i_1}{d_1} > \frac{1}{G_0}$;

(II) $0 < \mathcal{B}(DF(V_x^p)) < 1$ *if one of the following holds:*

- **(D2a)** $G_1 > 0$ *and* $\frac{1}{G_0 + \frac{d_0}{2(1+i_0)}} > 2\frac{1+i_1}{d_1}$;
- **(D2b)** $G_1 > 0$ *and* $\frac{1}{G_0 + \frac{d_0}{2(1+i_0)}} + \frac{1}{G_1} < 2\frac{1+i_1}{d_1}$;
- **(D2c)** $G_1 < 0$ *and* $\frac{1}{G_0 + \frac{d_0}{2(1+i_0)}} + \frac{1}{G_1} < 2\frac{1+i_1}{d_1} < \frac{1}{G_0 + \frac{d_0}{2(1+i_0)})}$;
- **(F2a)** $G_1 + \frac{d_2}{2i_2} < 0$ *and* $\frac{1}{G_0 + \frac{d_0}{2(1+i_0)}} < 2\frac{1+i_1}{d_1}$;
- **(F2b)** $G_1 + \frac{d_2}{2i_2} < 0$ *and* $\frac{1}{G_0 + \frac{d_0}{2(1+i_0)}} + \frac{1}{G_1 + \frac{d_2}{2i_2}} > 2\frac{1+i_1}{d_1}$;

[5] This is a very rough estimate, but is sufficient for our need.

(F2c) $G_1 + \frac{d_2}{2i_2} > 0$ *and* $\frac{1}{G_0 + \frac{d_0}{2(1+i_0)}} + \frac{1}{G_1 + \frac{d_2}{2i_2}} > 2\frac{1+i_1}{d_1} > \frac{1}{G_0 + \frac{d_0}{2(1+i_0)}}$;

(III) *Eq.* (4.2) *holds if and only if one of the following holds:*

(P1a) $\tau_0 < (1 - \frac{1}{2(1+i_0)})d_0 + (1 - \frac{1}{2(1+i_1)})d_1$;

(P1b) $\tau_0 > d_0 + (1 - \frac{1}{2(1+i_1)})d_1$.

Proof. (1). To show **(D1)**, we let $Y = \frac{1}{\tau_0 - d_0 - \frac{i_1}{1+i_1}d_1} - 2\frac{1+i_1}{d_1}$. Then **(D1)** can be rewritten as $G_1 + \frac{1}{Y} > 0$, which holds if and only if one of the following holds:

(1.1) $G_1 > 0$ and $Y > 0$;
(1.2) if $G_1 > 0$ and $Y < 0$, then $\frac{1}{G_1} < -Y$;
(1.3) if $G_1 < 0$ and $Y > 0$, then $Y < -\frac{1}{G_1}$.

The three conditions **(D1a)**–**(D1c)** are evidently equivalent to three items (1.1)–(1.3), respectively. The verifications of **(D2)** and **(F1)**-**(F2)** are reduced to the similar calculations and hence are omitted here.

(2). Let $X = \frac{1}{G_1} - 2\frac{1+i_1}{d_1}$ and

$$Y = \frac{1}{\tau_0 - (1 - \frac{1}{2(1+i_0)})d_0 - \frac{i_1}{1+i_1}d_1} - 2\frac{1+i_1}{d_1}. \tag{A.1}$$

Then Eq. (4.2) can be rewritten as $\frac{1}{X} < \frac{1}{Y}$, which is true if and only if one of the following holds:

(2.1) if $X < 0$ and $Y < 0$, then $X > Y$;
(2.2) if $X > 0$ and $Y > 0$, then $X > Y$;
(2.3) $X < 0$ and $Y > 0$.

Then plugging in the formulas $X = \frac{1}{G_1} - 2\frac{1+i_1}{d_1}$ and Eq. (A.1) for Y, we get that

(2.1.1) $\tau_0 < (1 - \frac{1}{2(1+i_0)})d_0 + \frac{i_1}{1+i_1}d_1$;
(2.1.2) $\tau_0 > d_0 + (1 - \frac{1}{2(1+i_1)})d_1$;
(2.2′) $d_0 + \frac{i_1}{1+i_1}d_1 < \tau_0 < (1 - \frac{1}{2(1+i_0)})d_0 + (1 - \frac{1}{2(1+i_1)})d_1$;
(2.3′) $(1 - \frac{1}{2(1+i_0)})d_0 + \frac{i_1}{1+i_1}d_1 < \tau_0 < \min\{d_0 + \frac{i_1}{1+i_1}d_1, (1 - \frac{1}{2(1+i_0)})d_0 + (1 - \frac{1}{2(1+i_1)})d_1\}$.

Note that Condition (2.2′) is nonempty if and only if $\frac{d_0}{1+i_0} < \frac{d_1}{1+i_1}$. However, we can always combine (2.1.1)-(2.2′)-(2.3′) into one condition:

$$\tau_0 < (1 - \frac{1}{2(1+i_0)})d_0 + (1 - \frac{1}{2(1+i_1)})d_1.$$

Therefore we get **(P1a)**-**(P1b)**. □

Acknowledgements

L.B. is supported in part by NSF grant DMS-1265883. H.Z. is supported in part by NSF CAREER grant DMS-1151762. H.Z. would like to thank U. Houston, at where part of the work was done. P.Z. would like to thank UMass Amherst, at where part of the work was done. The authors are grateful to the referees for their careful readings, useful comments and suggestions which helped them to improve the readability of this paper significantly.

References

[1] L. A. Bunimovich. *On billiards closed to dispersing.* Matem. Sbornik **95** (1974), 49–73.

[2] L. A. Bunimovich. *On ergodic properties of certain billiards.* Funct. Anal. Appl. **8** (1974), 254–255.

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

[4] L.A. Bunimovich, G. Del Magno. *Track billiards.* Comm. Mathh. Phys. **13** (2009) 699–713,

[5] J. Chen, L. Morh, H.-K. Zhang, P. Zhang. *Ergodicity and coexistence of elliptic islands in a family of convex billiards.* Chaos **23** (2013), 043137.

[6] N. Chernov, R. Markarian. *Chaotic billiards.* Math. Surv. Monogr. **127**, AMS, Providence, RI, 2006.

[7] N. Chernov, H.-K. Zhang. *Billiards with polynomial mixing rates.* Nonlinearity **18** (2005), 1527–1553.

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

[9] R. Douady. *Applications du théorème des tores invariants.* Univ. Paris 7, Thèse 3-eme cycle, 1982.

[10] E. Heller, S. Tomsovic. *Postmodern quantum mechanics.* Physics Today **46** (1993), 38–46.

[11] A. Ya. Khinchin. *Continued fractions.* The University of Chicago Press, Chicago, 1964.

[12] V. F. Lazutkin. *Existence of a continuum of closed invariant curves for a convex billiard.* Math. USSR Izvestija **7** (1973), 185–214.

[13] V. Lopac, I. Mrkonjic, D. Radic. *Classical and quantum chaos in the generalized parabolic lemon-shaped billiards.* Phys. Rev. E 59 (1999), 303–311.

[14] V. Lopac, I. Mrkonjic, D. Radic. *Chaotic behavior in lemon-shaped billiards with elliptical and hyperbolic boundary arcs.* Phys. Rev. E 64 (2001) 016214.

[15] R. Markarian. *Billiards with polynomial decay of correlations.* Ergod. Th. Dynam. Sys. **24** (2004), 177–197.

[16] H. Makino, T. Harayama, Y. Aizawa. *Quantum-classical correspondences of the Berry-Robnik parameter through bifurcations in lemon billiard systems.* Phys. Rev. E **63** (2001), 056203.

[17] S. Oliffson Kamphorst, S. Pinto-de-Carvalho. *The first Birkhoff coefficient and the stability of 2-periodic orbits in billiards.* Exp. Math. **14** (2005), 299–306.

[18] S. Ree, L. E. Reichl. *Classical and quantum chaos in a circular billiard with a straight cut.* Phys. Rev. E **60** (1999), 1607 (9).

[19] Ya. Sinaĭ. *Dynamical systems with elastic reflections. Ergodic properties of diepersing billiards.* Russian Math. Surveys, **25** (1970) 137–189.

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

School of Mathematics, Georgia Institute of Technology, Atlanta, GA 30332

E-mail address: bunimovh@math.gatech.edu

Department of Mathematics and Statistics, UMass Amherst, Amherst, MA 01003

E-mail address: hongkun@math.umass.edu

Department of Mathematics, University of Houston, Houston, TX 77004

E-mail address: pzhang@math.uh.edu