Nihonkai Math. J.
Vol.13(2002), 73-120

Geometry Of Geodesics For Convex Billiards And Circular Billiards

NOBUHIRO INNAMI

Dedicated to Professor K. Shiohama on his sixtieth birthday

Abstract

In the present paper circles and ellipses will be characterized by some properties of billiard ball trajectories. Those properties will be discussed in connection with the characterization of flat metrics on tori by some families of geodesics and tori of revolution. The main method is the geometry of geodesics due to H. Busemann which was reconstructed in the configuration space by V. Bangert. In particular, the theory of parallels plays an important role in the present paper.

1 Introduction

Let C be a smooth simple closed and strictly convex curve with length L in the Euclidean plane $\mathbf{E}$ and let $c : \mathbf{R} \longrightarrow \mathbf{E}$ be its representation with arclength, namely $|\dot{c}(t)| = 1$ for any $t \in \mathbf{R}$ where $\mathbf{R}$ is the set of all real numbers. Let $x = (x_j)_{j\in\mathbf{Z}}$ be a sequence of points in C where $\mathbf{Z}$ is the set of all integers. We say that x is a *billiard ball trajectory* if the angle between the tangent vector A to C at x_i and the oriented segment $T(x_{i-1}, x_i)$ from x_{i-1} to x_i is equal to the one between A and $T(x_i, x_{i+1})$ for any $i \in \mathbf{Z}$.

A billiard ball trajectory $x = (x_j)_{j\in\mathbf{Z}}$ in C is represented by a sequence $s = (s_j)_{j\in\mathbf{Z}}$ of real numbers such that $x_j = c(s_j)$ and $s_j < s_{j+1} < s_j + L$ for any $j \in \mathbf{Z}$ and the sequence $s = (s_j)_{j\in\mathbf{Z}}$ will be considered to be a configuration $\{(j, s_j)\}_{j\in\mathbf{Z}}$ in the configuration space $\mathbf{X} = \mathbf{Z} \times \mathbf{R} \subset \mathbf{R}^2$. A configuration $s = (s_j)_{j\in\mathbf{Z}}$ for x is determined uniquely up to the difference pL $(p \in \mathbf{Z})$.

Let $x_0, x_1 \in C$ and (x_0, x_1, x_2) the billiard ball trajectory. Let θ_0 (resp., θ_1) be the angle between the segment $T(x_0, x_1)$ from x_0 to x_1 (resp., $T(x_1, x_2)$) and the tangent vector to C at x_0 (resp., x_1). Set $u_0 = \cos\theta_0$ and $u_1 = \cos\theta_1$. We call $\Omega = C \times (-1, 1)$ the *phase space* which is the set of all pairs (x, u) for $x \in C$ and $u \in (-1, 1)$. Define a *billiard ball map* $\varphi : \Omega \longrightarrow \Omega$ as $\varphi(x_0, u_0) = (x_1, u_1)$. The billiard ball map is an example of a monotone twist map (see [17]). Let $\bar{x} = (x_0, u_0) \in \Omega$ and $\varphi^j(\bar{x}) = (x_j, u_j)$ for all $j \in \mathbf{Z}$. Then, the sequence $x = (x_j)_{j\in\mathbf{Z}}$ is a billiard ball trajectory. Any billiard ball trajectory is given in this way.

In the present paper circles and ellipses will be characterized by some properties of billiard ball trajectories. Those properties will be discussed in connection with the characterization of flat metrics on tori by some families of geodesics and tori of revolution. The main tool is the geometry of geodesics due to H. Busemann (see [6]) which was reconstructed in the configuration space by V. Bangert (see [1],[2]).

We first recall some properties of billiard ball trajectories in circles and ellipses (see [17]).

Billiards in circles : If C is a circle with center z, then its radius is $r = L/2\pi$. Let $x = (x_j)_{j\in\mathbf{Z}}$ be a billiard ball trajectory and $s = (s_j)_{j\in\mathbf{Z}}$ its configuration. Then, $s_{j+1} - s_j$ is constant for any $j \in \mathbf{Z}$. Given $a > 0$ and $y \in C$ there exists a billiard ball trajectory $x = (x_j)_{j\in\mathbf{Z}}$ with a configuration $s = (s_j)_{j\in\mathbf{Z}}$ such that $x_0 = c(s_0) = y$ and all $s_{j+1} - s_j$ are equal to a. Moreover, the envelope curve of initial segments $T(x_0, x_1)$ of these billiard ball trajectories with initial points x_0 is a circle K with center z, and such a billiard ball trajectory $x = (x_j)_{j\in\mathbf{Z}}$ satisfies that the successive point x_{j+1} after x_j in C is given so that the segment $T(x_j, x_{j+1})$ is tangent to K. For any positive integers $q, p \in \mathbf{Z}^+$ with $p/q < 1$ and any point $y \in C$ there exists a periodic billiard ball trajectory $x = (x_j)_{j\in\mathbf{Z}}$ such that $x_{j+q} = x_j$ and $s_{j+q} - s_j = pL$ for any $j \in \mathbf{Z}$ where $s = (s_j)_{j\in\mathbf{Z}}$ is a configuration for x. Let $t = (t_j)_{j\in\mathbf{Z}}$ be a configuration with $t_0 = s_0$ for a billiard ball trajectory $y = (y_j)_{j\in\mathbf{Z}}$. Then, $t_j > s_j$ for any $j > 0$ if $t_1 > s_1$ and $\lim_{j\to\infty} t_j - s_j = \infty$ (the divergence property).

Billiards in ellipses : An ellipse C has the property similar to ones circles have. Let C be an ellipse and for any $x_0 \in C$ let $x = (x_j)_{j\in\mathbf{Z}}$ be a billiard ball trajectory for which the segment $T(x_j, x_{j+1})$ does not pass between the focal points of C for any $j \in \mathbf{Z}$. Then, the segments $T(x_j, x_{j+1})$ are tangent to an ellipse with the same focal points as C. Since x_q depends continuously on

the points x_1 when x_0 is fixed, we can find a periodic billiard ball trajectory $x = (x_j)_{j\in\mathbf{Z}}$ such that $x_{i+q} = x_i$ for any $i \in \mathbf{Z}$ and the sum of the arclengths of subarcs $a(x_j, x_{j+1})$ for $j = i, \cdots, i+q-1$ is pL if $p/q < 1$ and $p/q \neq 1/2$. Notice that $a(x_j, x_{j+1}) = s_{j+1} - s_j$ if $s = (s_j)_{j\in\mathbf{Z}}$ is a configuration for x. Thus, for any point $y \in C$ and any positive integer $q, p \in \mathbf{Z}^+$ with $p/q < 1$ $(p/q \neq 1/2)$ there exists a periodic billiard ball trajectory $x = (x_j)_{j\in\mathbf{Z}}$ with period (q,p) such that $x_{j+q} = x_j$ and $s_{j+q} - s_j = pL$ where $s = (s_j)_{j\in\mathbf{Z}}$ is a configuration for x. Let $x = (x_j)_{j\in\mathbf{Z}}$ and $y = (y_j)_{j\in\mathbf{Z}}$ be billiard ball trajectories such that $x_0 = y_0$ is an endpoint of the diameter (the long axis) of C and $s = (s_j)_{j\in\mathbf{Z}}$ and $t = (t_j)_{j\in\mathbf{Z}}$ their configurations with $t_0 = s_0$. Then, $t_j > s_j$ for all $j \in \mathbf{Z}^+$ and $t_j < s_j$ for all $j \in \mathbf{Z}^-$ if $t_1 > s_1$ where $\mathbf{Z}^+$ (resp., $\mathbf{Z}^-$) is the set of all positive (resp., negative) integers. We will call those points *poles*. The billiard ball trajectories passing between the focal points of C are tangent to a hyperbola with the same focal points.

Taking these properties of billiard ball trajectories in consideration, we propose some characterizations of circles and ellipses.

According to M. Bialy ([4]), the billiard is called *integrable* if a subset of full measure of the phase space is foliated by closed curves invariant under the billiard ball map φ. As was seen in the above the billiards in circles and ellipses are integrable. G. Birkhoff's conjecture is stated in [4] as follows. The only examples of integrable billiards are circular and elliptic billiards. M. Bialy ([4]) has given a partial answer of the conjecture, proving that C is a circle if Ω is foliated by φ-invariant continuous closed curves not null-homotopic in Ω. M. Wojtkowski ([21]) proved that C is a circle if the domain bounded by C is foliated by smooth caustics to which almost every billiard ball trajectories are tangent. As was stated in [4] Bialy's theorem corresponds to a theorem of E. Hopf ([9]) concerning Riemannian metrics on tori without conjugate points. N. Innami ([16]) extended Bialy's theorem to the higher dimensional case and the nonpositive curvature case as L. Green ([7]) did E. Hopf's.

In order to state our contribution we need some words. In this paper a *simple curve* f in Ω means that f with relative topology in Ω is homeomorphic to a point, a bounded closed interval or a circle. We say that a family F of subsets in Ω is a *covering of* Ω *by simple curves* if for any $\bar{x} \in \Omega$ there exixts a simple curve $f \in F$ with $\bar{x} \in f$. We call a covering F of Ω by simple curves a *foliation of* Ω *by simple curves* if for any $\bar{x} \in \Omega$ there exists a unique simple curve $f \in F$ with $\bar{x} \in f$. A covering F of Ω by simple curves is by

definition *element-wise φ-periodic* if for any $f \in F$ there exists an $i \in \mathbf{Z}^+$ such that $f, \varphi(f), \cdots, \varphi^{i-1}(f)$ are mutually disjoint and $\varphi^i(f) = f$. We say that F is *φ-invariant* if $\varphi(f) = f$ for any $f \in F$.

Theorem 1.1. *C is a circle if and only if Ω admits a φ-invariant foliation F of Ω by simple curves such that F is closed in the set of all closed subsets in Ω.*

We have Theorem 1.1 as a consequence of Bialy's theorem, proving that all simple curves $f \in F$ are not null-homotopic in Ω.

As was mentioned by V. Bangert ([1]) we can study the geodesics on tori and billiard ball trajectories in convex curves by using the same idea, the variational method in their configuration spaces. This fact suggests us that there would exist some characterizations of circles and ellipses which correspond to those for flat tori and tori of revolution.

We say that a periodic billiard ball trajectory $x = (x_j)_{j \in \mathbf{Z}}$ with period (q,p) starting at x_0 is *minimal* if $-\sum_{j=0}^{q-1} |x_j - x_{j+1}|$ is minimal in the set of all sequences $x' = (x'_j)_{0 \le j \le q}$ such that $\sum_{j=0}^{q-1} a(x'_j, x'_{j+1}) = pL$ and $x'_0 = x_0$, $x'_q = x_q$, where $a(x'_j, x'_{j+1}) = s'_{j+1} - s'_j$ if $x'_j = c(s'_j)$ with $s'_j < s'_{j+1} < s'_j + L$ for any $j \in \mathbf{Z}$. This is equivalent to the following. If $s = (s_j)_{j \in \mathbf{Z}}$ is a configuration for a minimal periodic billiard ball trajectory x with period (q,p), then $-\sum_{j=0}^{q-1} |c(s_j) - c(s_{j+1})|$ is minimal in the set of all configurations $t = (t_j)_{0 \le j \le q}$ with $t_j < t_{j+1} < t_j + L$ such that $t_0 = s_0$ and $t_q = s_q = s_0 + pL$.

A curve C is of constant width if and only if there ezists a periodic billiard ball trajectory from any point in C with period (2,1) (see [18],[20]). We see the example of a curve C in [15] such that there exists a periodic billiard ball trajectory from any point in C with period (3,1) and it is not an ellipse. These periodic billiard ball trajectories are minimal. These examples show that the assumption of the following corollary is not superfluous.

Corollary 1.2. *Suppose there exists a minimal periodic billiard ball trajectory from any point in C with any period (q,p) such that $p/q < 1$. Then, C is a circle.*

E. Hopf's theorem ([9]) states that a torus T^2 is flat if all points are poles. However, the existence of poles does not imply that the Riemannian metric is flat. In fact, a torus T^2 of revolution has poles and points which are not poles (see [11]). In contrast to this fact a torus T^2 is flat if there is a pole P such that the length of any perpendicular Jacobi vector field $Y(t)$ with

$Y(0) = 0$ along any unit speed geodesic $\gamma(t)$ from $\gamma(0) = P$ is monotone increasing for $t > 0$ (see [11],[13]).

In order to think the similar properties we define poles for convex billiards. Let $s_0 = t_0$ and $x_0 = c(s_0)$. Let $s = (s_j)_{j \in \mathbf{Z}}$ and $t = (t_j)_{j \in \mathbf{Z}}$ be configurations for billiard ball trajectories with $t_1 > s_1$. We say that a point $x_0 \in C$ is a *pole* if t and s do not cross at any other point than s_0, namely, $t_j > s_j$ for $j > 0$ and $t_j < s_j$ for $j < 0$. We say that a pole $x_0 \in C$ has the *monotone property* if $t_j - s_j$ is monotone increasing for $j \in \mathbf{Z}$.

All points in circles are poles with monotone property. In an ellipse the endpoints of long axis are poles without monotone property and other points are not poles.

Corollary 1.3. *C is a circle if and only if all points in C are poles.*

Corollary 1.4. *C is a circle if and only if C has a pole satisfying the monotone property.*

Let M be the universal covering space of a torus T^2 and D the isometry group on M such that $T^2 = M/D$. Let $\tau \in D$. The displacement function d_τ of τ on M is given by $d_\tau(q) = d(q, \tau(q))$ for any point $q \in M$ where d is the natural distance induced from the Riemannian metric of M. A torus T^2 is flat if and only if all displacement functions d_τ for $\tau \in D$ are constant on M (see [12],[2]). If a torus T^2 is a surface of revolution, then there is an element $\eta \in D$ such that the displacement function d_τ is constant on M if $\tau \in D$ is not any power η^m of η. To show the corresponding fact we define a displacement function $D(q,p) : \mathbf{X} \longrightarrow \mathbf{R}$ for any $q, p \in \mathbf{Z}^+$ with $p/q < 1$ as follows. Let $s_j = (j, s_j) \in \mathbf{X}$. Then, $D(q,p)(s_j)$ is by definition the minimum of $-\sum_{k=j}^{j+q} |s_{k+1} - s_k|$ in the set of all configurations from $s_j = (j, s_j)$ to $s_{j+q} = (j+q, s_j + pL)$. In the light of the torus case we will prove the following.

Corollary 1.5. *C is a circle if and only if the displacement functions $D(q,p)$ are constant in $\mathbf{X}$ for all $q, p \in \mathbf{Z}^+$ with $p/q < 1$.*

Let M be the universal covering space of a torus T^2. We say that the geodesics in M satisfy the *divergence property* if $d(\beta(t), \gamma(t)) \longrightarrow \infty$ as $t \longrightarrow \infty$ for any unit speed geodesics $\beta(t)$ and $\gamma(t)$ with $\beta(0) = \gamma(0)$ and $\beta \neq \gamma$. A torus is flat if and only if the geodesics of its universal covering space satisfy the divergence property (see [22]). To prove the corresponding fact we define a divergence property as follows.

Let $s = (s_j)_{j\geq 0}$ and $t = (t_j)_{j\geq 0}$ be configurations of billiard ball trajectories such that $s_0 = t_0$. We define $\mathrm{dis}_\infty(s,t)$ as

$$\mathrm{dis}_\infty(s,t) = \liminf_{j\to\infty} |s_j - t_j|.$$

In contrast with the fact for tori we will prove the following.

Corollary 1.6. *C is a circle if and only if $\mathrm{dis}_\infty(s,t) > L/4$ for the configurations of any billiard ball trajectories $s \neq t$ with $s_0 = t_0$.*

The notion of slope is given as follows in Section 4. Let $x = (x_j)_{j\in\mathbf{Z}}$ be a billiard ball trajectory and let $a(x_j, x_{j+1})$ be the arclength of the subarc of C from x_j to x_{j+1} measured with the positive orientation of C. We define the *slope* of x as

$$\alpha(x) = \liminf_{n\to\infty} \frac{1}{n}\sum_{j=0}^{n-1} a(x_j, x_{j+1}) = \liminf_{n\to\infty} \frac{s_n}{n}.$$

where $s = (s_j)_{j\in\mathbf{Z}}$ is a configuration for x. Let $\alpha(\tilde{x})$ denote the slope of the billiard ball trajectory determined by $\tilde{x}$ for $\tilde{x} \in \Omega$. In addition, the asymptotic billiard ball trajectories and parallel biliard ball trajectories will be defined in the configuration space in Section 5 as in Euclidean geometry.

Let $s = (s_j)_{j\in\mathbf{Z}}$ be a minimal configuration for a billiard ball trajectory $\bar{x} \in \Omega$ with slope $\alpha(\bar{x}) = aL$ where a is an irrational number with $0 < a < 1$. If $\bar{x}$ is contained in a φ-invariant curve f (not a point), then f is a simple closed curve not null-homotopic in Ω (see Lemma 7.1), and, moreover, all points in f corresponds to billiard parallels to s in $\mathbf{X}$ (see Theorem 4.16). Then, we have a caustic in $\mathbf{E}$ which is a closed continuous curve (see Lemma 6.8 and Lemma 6.1). Where we say that a closed continuous curve K is a *caustic* if K has the following property. Let x_0 be an arbitrary point in C and let $T(x_0, x_1)$ be a segment tangent to K. If $x = (x_j)_{j\in\mathbf{Z}}$ is the billiard ball trajectory determined by $T(x_0, x_1)$, then $T(x_j, x_{j+1})$ is a segment tangent to K for all $j \in \mathbf{Z}$. In the light of those results, we will prove the following.

Theorem 1.7. *Suppose there exists a sequence of closed simple curves f_n with slope $\alpha(\bar{x}) = \alpha_n \neq L/2$ for any $\bar{x} \in f_n$ not null-homotopic in Ω such that they make simple caustics in $\mathbf{E}$. If the sequence of slopes α_n converges to $L/2$, then C is an ellipse.*

This theorem should be contrasted with the existence of region free of caustics for an arbitrary convex billiards proved by E. Gutkin and A. Katok ([8]). We state the following which is the main part of the proof of Theorem 1.7.

Corollary 1.8. *Suppose there exists a sequence of simple caustics K_n such that $\alpha(x^n) \neq L/2$ for $x^n = (x^n{}_j)_{j\in\mathbf{Z}}$ being a billiard ball trajectory tangent to K_n and $\alpha(x^n) \longrightarrow L/2$ as $n \longrightarrow \infty$. Then, C is an ellipse.*

2 Billiard geodesic

Let C be a smooth strictly convex simple closed curve in the Euclidean plane $\mathbf{E}$ with length L. Let $\mathbf{X} = \mathbf{Z} \times \mathbf{R} \subset \mathbf{R}^2$ where $\mathbf{Z}$ is the set of all integers and $\mathbf{R}$ is the set of all real numbers. We denote $(i, s_i) \in \mathbf{X}$ by s_i for simplicity. A configuration $s = (s_j)_{i\leq j\leq k}$ makes a broken segment $T(s) = \cup_{j=i}^{k-1} T(s_j, s_{j+1})$ in $\mathbf{R}^2$ where $T(s_j, s_{j+1})$ is the segment from (j, s_j) to $(j+1, s_{j+1})$ in $\mathbf{R}^2$. For $q, p \in \mathbf{Z}$ let $U(q,p)$ be the translation in $\mathbf{X}$ which is given by

$$U(q,p)(s_i) = U(q,p)(i, s_i) = (i+q, s_i + pL)$$

for any $(i, s_i) \in \mathbf{X}$. Let $x = (x_j)_{i\leq j\leq k}$ be a sequence of mutually different points in C. We define a configuration $s = (s_j)_{i\leq j\leq k}$ for x as $x_j = c(s_j)$ and $s_j < s_{j+1} < s_j + L$ for $i \leq j \leq k-1$. We call such a configuration s and a broken segment $T = T(s)$ made of such a configuration s a *C-curve*. We define the negative length of a C-curve $T = T(s)$ as

$$H(s; i, k) = H(s_i, s_{i+1}, \cdots, s_k) = -\sum_{j=i}^{k-1} |c(s_{j+1}) - c(s_j)|$$

where $|\cdot|$ is the natural norm in $\mathbf{E}$ and $c : \mathbf{R} \longrightarrow \mathbf{E}$ is the representation of C by arclength. Let $H(i, k; u, v)$ denote the minimum of $H(s; i, k)$ in the set of all C-curves $s = (s_j)_{i\leq j\leq k}$ from $s_i = (i, u)$ to $s_k = (k, v)$.

We borrows some words from the geometry of geodesics (see [6]).

1. A C-curve $s = (s_j)_{i\leq j\leq k}$ (and $T = T(s)$) is called a *billiard curve* or simply a *b-curve* if $x = (x_j)_{i\leq j\leq k}$ given by $x_j = c(s_j)$ for $i \leq j \leq k$ is a billiard ball trajectory.

2. A b-curve $s = (s_j)_{i \le j \le k}$ (and $T = T(s)$) is called a *billiard geodesic* or simply a *b-geodesic* if $H(s; j, j+2)$ is the minimum in the set of all C-curves from s_j to s_{j+2} for $i \le j \le k-2$, namely $H(s; j, j+2) = H(j, j+2; s_j, s_{j+2})$.

3. A C-curve $s = (s_j)_{i \le j \le k}$ (and $T = T(s)$) is called a *billiard segment* or simply a *b-segment* if $H(s; i, k)$ is the minimum in the set of all C-curves from s_i to s_k, namely $H(s; i, k) = H(i, k; s_i, s_k)$.

4. A b-geodesic $s = (s_j)_{j \ge i}$ (and $s = (s_j)_{j \le i}$) (and $T = T(s)$) is called a *billiard ray from* s_i or simply a *b-ray from* s_i if all sub-b-geodesics are b-segments, namely $H(s; j, k) = H(j, k; s_j, s_k)$ for any $k > j$.

5. A b-geodesic $s = (s_j)_{j \in \mathbf{Z}}$ (and $T = T(s)$) is called a *billiard straight line* or simply a *b-straight line* if all sub-b-geodesics are b-segments, namely $H(s; j, k) = H(j, k; s_j, s_k)$ for any $k > j$.

We will see the difference among those definitions in the following.

Example : Let $a > b > 0$ and let C be an ellipse in $\mathbf{E}$ given by the equation

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1.$$

Let the configuration $s = (s_j)_{j \in \mathbf{Z}}$ correspond to the short axis as a billiard ball trajectory. Then it a b-curve which is not a b-geodesic if $a > \sqrt{2}b$. It is a b-geodesic which is not a b-straight line if $\sqrt{2}b \ge a$. The configuration for the long axis is always a b-straight line. All configurations for billiard ball trajectories passing through focal points are also b-straight lines. There exists no billiard ball trajectory in any ellipse whose configuration is a b-ray which is not a b-straight line. However, the b-rays play important roles in the theory of parallels in the present paper.

The following lemmas are fundamental. The proofs are similar to those ones for geodesics in Riemannian geometry. Most results in Section 2 to 5 are seen in [1] and [2]. The different point is that the configurations in our consideration are only C-curves. It means that there may not exist any C-curve joinning given points. Another different point is that we do not use the notion of circle maps and rotation numbers.

Lemma 2.1. *A C-curve $s = (s_j)_{i \le j \le k}$ is a b-curve if and only if it is a critical C-curve of H in the set of all C-curves from s_i to s_k.*

Proof. Let $H(t_{i+1},\dots,t_{k-1}) = -|c(s_i) - c(t_{i+1})| - \sum_{j=i+1}^{k-2} |c(t_{j+1}) - c(t_j)| - |c(s_k) - c(t_{k-1})|$. If s is a critical C-curve of H, we have that

$$\begin{aligned} 0 =& \frac{\partial H}{\partial t_j}(s_{i+1},\dots,s_{k-1}) \\ =& \left\langle \frac{c(s_j) - c(s_{j-1})}{|c(s_j) - c(s_{j-1})|}, \dot{c}(s_j) \right\rangle - \left\langle \dot{c}(s_j), \frac{c(s_{j+1}) - c(s_j)}{|c(s_{j+1}) - c(s_j)|} \right\rangle \end{aligned}$$

for $j = i+1,\dots,k-1$ where $\dot{c}(s_j)$ is the velocity vector of c at s_j. This means that s is a b-curve. □

Lemma 2.2. *Let $t_i < t_{i+1} < t_i + L$. Then, there exists a unique b-curve $s = (s_j)_{j\in\mathbf{Z}}$ such that $s_i = t_i$ and $s_{i+1} = t_{i+1}$.*

Proof. A billiard ball trajectory $x = (x_j)_{j\in\mathbf{Z}}$ is determined by two successive points x_i and x_{i+1} in C. □

Lemma 2.3. *A b-segment is a b-geodesic.*

Proof. This is a consequence of the definition. □

Lemma 2.4. *H is invariant under any translation $U(q,p)$, namely, if $s = (s_j)_{i\le j\le k}$ and $s' = U(q,p)s = (s'_j)_{i+q\le j\le k+q}$, $s'_j = s_{j-q} + pL$, for any $i+q \le j \le k+q$, then $H(s;i,k) = H(s';i+q,k+q)$. In particular, $s' = U(q,p)s$ is a b-curve (geodesic, segment, ray and straight line, resp.) if s is a b-curve (geodesic, segment, ray and straight line, resp.).*

Proof. s and s' yield the same billiard ball trajectory as a plane figure in $\mathbf{E}$. □

Lemma 2.5. *If $s = (s_{i-1}, s_i, s_{i+1})$ and $s' = (s'_{i-1}, s'_i, s'_{i+1})$ are b-curves with $s_i = s'_i$ and $s \ne s'$, then $(s_{i-1} - s'_{i-1})(s_{i+1} - s'_{i+1}) < 0$. In particular, if $s'_{i-1} \le s_{i-1}$, $s'_i = s_i$ and $s'_{i+1} \le s_{i+1}$, then $s = s'$.*

Proof. Let $x_j = c(s_j)$ and $x'_j = c(s'_j)$ for $j = i-1, i, i+1$. Then, $x_i = x'_i$. If x'_i is in the oriented subarc of C from x_{i-1} to x_i, then x'_{i+1} is in the oriented subarc of C from x_i to x_{i+1}. Therefore, if $s_{i-1} < s'_{i-1} < s_i = s'_i$, then $s_i = s'_i < s'_{i+1} < s_{i+1}$. The other case is proved in the same way. □

We say that s and s' as in the first part of Lemma 2.5 *cross* at i. We sometimes use this lemma for indirect proofs.

Lemma 2.6. *If $T(s_i, s_{i+1})$ intersects $T(s'_i, s'_{i+1})$, namely $s_i < s'_i < s'_{i+1} < s_{i+1} < s_i + L$, then*

$$|c(s_{i+1}) - c(s_i)| + |c(s'_{i+1}) - c(s'_i)| < |c(s'_{i+1}) - c(s_i)| + |c(s_{i+1}) - c(s'_i)|.$$

namely,

$$H(s_i, s_{i+1}) + H(s'_i, s'_{i+1}) > H(s_i, s'_{i+1}) + H(s'_i, s_{i+1}).$$

Proof. Let $A = c(s_i)$, $B = c(s'_i)$, $C = c(s'_{i+1})$, $D = c(s_{i+1})$. For the four-sided figure $ABCD$ the sum of length of diagonal lines $AC + BD$ is greater than the one of a pair of oposite lines $BC + AD$. □

We say that s and s' as in the assumption of Lemma 2.6 *cross* between i and $i + 1$.

Lemma 2.7. *If $s_i < s_k < s_i + L(k - i)$, then there exists a b-segment $s = (s_j)_{i \le j \le k}$ from s_i to s_k, namely $H(s; i, k) = H(i, k; s_i, s_k)$. If $s_k > s_i + L(k-i)$, $k > i$, then there exist no C-curve connecting s_k and s_i.*

Proof. Let $t = (t_j)_{i \le j \le k}$ is defined as

$$t_j = \frac{j - i}{k - i} s_k + \frac{k - j}{k - i} s_i$$

for all j with $i \le j \le k$. Then, t is a C-curve joinning s_i to s_k. The b-segment s mentioned in the statement is given as a C-curve minimizing H in the set of all C-curves connecting s_i to s_k.

If $s = (s_j)_{i \le j \le k}$ is a configuration such that $s_k > s_i + L(k - i)$, then at least one of $s_{j+1} - s_j$ is greater than L. Hence, s is not a C-curve. □

Let $s = (s_j)_{i \le j \le k}$ and $s' = (s'_j)_{i \le j \le k}$ be b-geodesics such that $T(s) \cap T(s') = \emptyset$. Suppose $s_j < s'_j$ for all $i \le j \le k$. Then, we have a strip $[T(s), T(s')]$ in $\mathbf{R}^2$ whose lower boundary is $T(s)$ and upper boundary is $T(s')$. We also denote $[T(s), T(s')] \cap \mathbf{X}$ as $[T(s), T(s')]$.

Proposition 2.8. *Let t_j and t_h $(i \le j < h \le k)$ be in the strip $[T(s), T(s')]$ such that $t_j < t_h < t_j + L(h - j)$. Then there exists a b-geodesic from t_j to t_h in the strip $[T(s), T(s')]$.*

Proof. Let $t' = ({t'}_m)_{j \leq m \leq h}$ be a C-curve from t_j to t_h in $\mathbf{X}$. If t' does not lie in $[T(s), T(s')]$, we deform it as follows. Let $m_0 = \min\{m | s_m \geq {t'}_m, j \leq m \leq h\}$ and $m_1 = \max\{m | s_m \geq {t'}_m, j \leq m \leq h\}$. Set $t'' = ({t''}_m)_{j \leq m \leq h}$ with ${t''}_m = {t'}_m$ for $j \leq m \leq m_0$, ${t''}_m = s_m$ for $m_0 + 1 \leq m \leq m_1$, and ${t''}_m = {t'}_m$ for $m_1 + 1 \leq m \leq h$. The C-curve t'' lies in the upper half plane to $T(s)$. Let ${m_0}' = \min\{m | s_m \leq {t''}_m, j \leq m \leq h\}$ and ${m_1}' = \max\{m | s_m \leq {t''}_m, j \leq m \leq h\}$. By using these ${m_0}'$ and ${m_1}'$ we can get a C-curve t''' which is in the strip $[T(s), T(s')]$. Thus the set of all C-curves from t_j to t_h in the strip $[T(s), T(s')]$ is not empty. Let $t = (t_m)_{j \leq m \leq h}$ be a C-curve at which H assumes its minimum in the set of all C-curves from t_j to t_h in the strip $[T(s), T(s')]$. This C-curve is a b-geodesic in the strip $[T(s), T(s')]$ because of Lemma 2.1, 2.3, 2.5 and 2.6. □

We sometimes need to consider b-segments, b-rays and b-straight lines in a strip $[T(s), T(s')]$ which are by definition *H-minimal* in the strip $[T(s), T(s')]$ but not so possibly in the whole space $\mathbf{X}$. We say that the set W of b-curves $t = (t_j)_{i \leq j \leq k}$ is a *foliation* of the strip $[T(s), T(s')]$ if there exists a unique b-curve in W through any point $t_h \in [T(s), T(s')]$.

Proposition 2.9. *If W is a foliation of the strip $[T(s), T(s')]$ by b-curves, then all b-curves $t = (t_j)_{i \leq j \leq k}$ in the foliation W are b-segments in the strip $[T(s), T(s')]$. Moreover, if t_k and t_m are in a b-curve $t = (t_j)_{i \leq j \leq k} \in W$, then the sub-$b$-curve $t = (t_j)_{h \leq j \leq m}$ of $t = (t_j)_{i \leq j \leq k}$ is the unique b-curve connecting t_h and t_m in the strip $[T(s), T(s')]$.*

Proof. Let $P_W : [T(s), T(s')] \longrightarrow [s_i, {s'}_i]$ be the projection along the foliation W, namely $P_W(t_j) = t_i$ for any $(j, t_j) \in [T(s), T(s')]$ where $t = (t_h)_{i \leq h \leq k} \in W$ is the unique b-curve in W through t_j. Suppose $t = (t_j)_{h \leq j \leq m}$, $i \leq h \leq m \leq k$, is not a b-segment in the strip $[T(s), T(s')]$. Then, there exists a b-segment $t' = ({t'}_j)_{h \leq j \leq m}$ from ${t'}_h = t_h$ to ${t'}_m = t_m$ in the strip $[T(s), T(s')]$ which is different from t, namely $t' \notin W$. Let $M = \max\{P_W({t'}_j) | h \leq j \leq m\}$ and $N = \min\{P_W({t'}_j) | h \leq j \leq m\}$. Then, $M \neq t_i$ or $N \neq t_i$. We will be able to have the same contradiction in either case. So we suppose $M \neq t_i$ and $M = P_W({t'}_{i_0})$ for some $h \leq i_0 \leq m$. Let $t'' = ({t''}_j)_{i \leq j \leq k} \in W$ with ${t''}_i = M$. Then, we have that ${t''}_{i_0 - 1} \geq {t'}_{i_0 - 1}$, ${t''}_{i_0} = {t'}_{i_0}$ and ${t''}_{i_0 + 1} \geq {t'}_{i_0 + 1}$ because of the definition of M. From Lemma 2.5 we see that $t'' = t'$, contradicting that ${t'}_i \neq M = {t''}_i$.

The same argument proves the remaining part of the statement. □

3 Billiard segment

In this section we will show that some properties of billiard segments are similar to those ones of minimizing geodesics in Riemannian geometry.

Proposition 3.1. *Let $t = (t_j)_{h \le j \le m}$ and $u = (u_j)_{h \le j \le m}$ be b-segments with $t \neq u$. Then, $T(t) \cap T(u)$ contains at most two points. If $T(t) \cap T(u) = \{a, b\}$, then a and b are common endpoints of t and u. Furthermore, there exists the unique b-segment from t_i to t_j which is a sub-b-segment of t for any t_j if t_j is not the endpoint (i.e., $j \neq i, m$) of a b-sugment $t = (t_j)_{i \le j \le m}$.*

Proof. Suppose $T(t) \cap T(u) \supset \{a, b\}$. We have to show that the following three cases do not result.

1. There exist h' and m' such that $h' \neq h$ or $m' \neq m$, and $h' < m'$ with $a = (h', t_{h'}) = (h', u_{h'})$, $b = (m', t_{m'}) = (m', u_{m'})$.
2. There exist h' and m' such that $a = (h', t_{h'}) = (h', u_{h'})$ and $b = T(t_{m'}, t_{m'+1}) \cap T(u_{m'}, u_{m'+1})$ $(b \notin \{m', m'+1\} \times \mathbf{R})$.
3. There exist h' and m' such that $a = T(t_{h'}, t_{h'+1}) \cap T(u_{h'}, u_{h'+1})$ and $b = T(t_{m'}, t_{m'+1}) \cap T(u_{m'}, u_{m'+1})$ $(a, b \notin \{h', h'+1, m', m'+1\} \times \mathbf{R})$.

Case (1): Suppose $m' \neq m$ and $t_{m'-1} < u_{m'-1}$. Since

$$\begin{aligned} & H(t_{m'-1}, t_{m'}, t_{m'+1}) + H(u_{m'-1}, u_{m'}, u_{m'+1}) \\ =& H(t_{m'-1}, t_{m'}) + H(t_{m'}, t_{m'+1}) + H(u_{m'-1}, u_{m'}) + H(u_{m'}, u_{m'+1}) \\ =& H(t_{m'-1}, t_{m'}, u_{m'+1}) + H(u_{m'-1}, u_{m'}, t_{m'+1}) \end{aligned}$$

and both $(t_{m'-1}, t_{m'}, u_{m'+1})$ and $(u_{m'-1}, u_{m'}, t_{m'+1})$ are not b-curves, there exist ${t'}_{m'}$ and ${u'}_{m'}$ such that

$$\begin{aligned} & H(t_{m'-1}, t_{m'}, u_{m'+1}) + H(u_{m'-1}, u_{m'}, t_{m'+1}) \\ >& H(t_{m'-1}, {t'}_{m'}, u_{m'+1}) + H(u_{m'-1}, {u'}_{m'}, t_{m'+1}). \end{aligned}$$

Set

$$\begin{aligned} t' &= (t_h, \ldots, t_{h'}, u_{h'+1}, \ldots, u_{m'-1}, {u'}_{m'}, t_{m'+1}, \ldots, t_m) \\ u' &= (u_h, \ldots, u_{h'}, t_{h'+1}, \ldots, t_{m'-1}, {t'}_{m'}, u_{m'+1}, \ldots, u_m) \end{aligned}$$

Then, we have

$$H(t;h,m)+H(u;h,m) > H(t';h,m)+H(u';h,m)$$

and, hence,

$$H(t;h,m) > H(t';h,m) \quad \text{or} \quad H(u;h,m) > H(u';h,m).$$

This contradicts the H-minimal property of t and u.
Case (2): Suppose $h' < m'$. Set

$$t' = (t_h, \ldots, t_{h'}, u_{h'+1}, \ldots, u_{m'}, t_{m'+1}, \ldots, t_m)$$
$$u' = (u_h, \ldots, u_{h'}, t_{h'+1}, \ldots, t_{m'}, u_{m'+1}, \ldots, u_m).$$

By using Lemma 2.6 we have that

$$\begin{aligned} H(t;h,m)+H(u;h,m) &= \sum_{j=h}^{m-1} \{H(t_j,t_{j+1})+H(u_j,u_{j+1})\} \\ &> \sum_{j=h}^{m'-1} \{H(t_j,t_{j+1})+H(u_j,u_{j+1})\} + H(u_{m'},t_{m'+1}) \\ &\quad + H(t_{m'},u_{m'+1}) + \sum_{j=m'+1}^{m-1} \{H(t_j,t_{j+1})+H(u_j,u_{j+1})\} \\ &= H(t';h,m)+H(u';h,m), \end{aligned}$$

and, hence,

$$H(t;h,m) > H(t';h,m) \quad \text{or} \quad H(u;h,m) > H(u';h,m),$$

This is a contradiction.
Case (3): Suppose $h' < m'$. Set

$$t' = (t_h, \ldots, t_{h'}, u_{h'+1}, \ldots, u_{m'}, t_{m'+1}, \ldots, t_m)$$
$$u' = (u_h, \ldots, u_{h'}, t_{h'+1}, \ldots, t_{m'}, u_{m'+1}, \ldots, u_m).$$

In the same way as before we get

$$H(t;h,m) > H(t';h,m) \quad \text{or} \quad H(u;h,m) > H(u';h,m),$$

a contradiction.

Therefore, if $T(t) \cap T(u)$ contains two points, then they are common endpoints of $T(t)$ and $T(u)$.

To prove the last part of the statement, let $t' = (t'_k)_{i \le k \le j}$, $t'_i = t_i$ and $t'_j = t_j$ be a b-segment which is not a sub-b-segment of $t = (t_k)_{i \le k \le m}$. Since (t'_{j-1}, t_j, t_{j+1}) is not a b-curve, there exists a b-segment $(t'_{j-1}, t''_j, t_{j+1})$ with

$$H(t'_{j-1}, t''_j) + H(t''_j, t_{j+1}) < H(t'_{j-1}, t_j) + H(t_j, t_{j+1})$$

Then, we have a C-curve $(t'_i, \ldots, t'_{j-1}, t''_j, t_{j+1})$ which satisfies

$$H(t'_i, \ldots, t'_{j-1}, t''_j, t_{j+1}) < H(t_i, \ldots, t_{j+1}),$$

contradicting that $t = (t_k)_{i \le k \le m}$ is a b-segment. □

By the same reason as Proposition 3.1 we can prove the following.

Lemma 3.2. *Let $s = (s_j)_{j \ge i_0}$ and $t = (t_j)_{j \ge i_0}$ ($s = (s_j)_{j \le i_0}$ and $t = (t_j)_{j \le i_0}$, resp.) be rays with $s_{i_0} \ne t_{i_0}$ and cross. Then,*

$$\liminf_{j \to \infty} |s_j - t_j| > 0, \qquad (\liminf_{j \to -\infty} |s_j - t_j| > 0, resp.)$$

The following is for Lemma 3.4.

Lemma 3.3. *Let $t = (t_j)_{i \le j \le k}$ be a b-segment. Then,*

$$H(t; i, k) \ge -|t_k - u| + H(i, k; t_i, u) \quad \text{if} \quad t_i + L(k - i) > u > t_i$$
$$H(t; i, k) \ge -|t_i - u| + H(i, k; u, t_k) \quad \text{if} \quad t_k - L(k - i) < u < t_k$$

Proof. Let $S = \{v |$ the first inequality is satisfied for any u with $t_i < v < u < t_k\}$. Then, $S \ne \emptyset$. Let $v_k = \inf S$. Suppose $v_k > t_i$ and $v = (v_j)_{i \le j \le k}$, $v_i = t_i$, is a b-segment. Then, for any u with $v_{k-1} < u < v_k$, we have

$$\begin{aligned} H(t; i, k) &\ge -|t_k - v_k| + H(i, k; t_i, v_k) \\ &= -|t_k - v_k| + H(i, k-1; t_i, v_{k-1}) + H(v_{k-1}, v_k) \\ &\ge -|t_k - v_k| - |v_k - u| + H(i.k-1; t_i, v_{k-1}) + H(v_{k-1}, u) \\ &\ge -|t_k - u| + H(i, k; t_i, u), \end{aligned}$$

contradicting the choice of v_k. Thus, $v_k = t_i$. In the same way we can prove the remaining part. □

The following inequality plays an important role in the study of the relation between the straightness of billiard geodesics and the minimum set of displacement functions.

Proposition 3.4. *Let $s = (s_j)_{i\le j\le k}$ and $t = (t_j)_{i\le j\le k}$ be b-segments. Then,*

$$|H(t;i,k) - H(s;i,k)| \le |t_i - s_i| + |t_k - s_k|.$$

Proof. From Proposition 3.1 we have three cases:

(1) $T(s) \cap T(t) = \emptyset$.

(2) $T(s) \cap T(t)$ consists of one point.

(3) $T(s) \cap T(t)$ consists of two points.

Case (1): Suppose $t_i > s_i$ and $t_k > s_k$. In this case we have two possibilities; (i) $t_i < s_k$ and (ii) $t_i > s_k$.

We first suppose (i) $t_i < s_k$. Then,

$$\begin{aligned} H(t;i,k) &\ge -|t_k - s_k| + H(i,k;t_i,s_k) \\ &\ge -|t_k - s_k| - |t_i - s_i| + H(i,k;s_i,s_k) \\ &= -|t_k - s_k| - |t_i - s_i| + H(s;i,k). \end{aligned}$$

In the same way we have

$$\begin{aligned} H(s;i,k) &\ge -|s_i - t_i| + H(i,k;t_i,s_k) \\ &\ge -|t_i - s_i| - |t_k - s_k| + H(t;i,k). \end{aligned}$$

We suppose (ii) $t_i > s_k$ this time. Let ${s^a}_i, {s^a}_k$, for $a = 0,\dots,n$, be such that

(a) $s_i = {s^0}_i < {s^1}_i < \cdots < {s^n}_i = t_i$.

(b) $s_k = {s^0}_k < {s^1}_k < \cdots < {s^n}_k = t_k$.

(c) ${s^{a+1}}_i < {s^a}_k < {s^{a+1}}_i + L(k-i)$ for $a = 0,\dots,n-1$.

From the first conclusion we have

$$|H(i,k;{s^a}_i,{s^a}_k) - H(i,k;{s^{a+1}}_i,{s^{a+1}}_k)| \le |{s^{a+1}}_i - {s^a}_i| + |{s^{a+1}}_k - {s^a}_k|$$

for all $a = 0, \ldots, n-1$. Hence, we have

$$\begin{aligned} |H(t;i,k) - H(s;i,k)| &\leq \sum_{a=0}^{n-1} |H(i,k;{s^a}_i, {s^a}_k) - H(i,k;{s^{a+1}}_i, {s^{a+1}}_k)| \\ &\leq |t_i - s_i| + |t_k - s_k| \end{aligned}$$

Case (2): Let $u = (u_j)_{i \leq j \leq k}$ be a b-segment from $u_i = s_i$ to $u_k = t_k$. As the limit of the Case (1) we have

$$\begin{aligned} |H(t;i,k) - H(u;i,k)| &\leq |t_i - s_i| \\ |H(s;i,k) - H(u;i,k)| &\leq |t_k - s_k|. \end{aligned}$$

Therefore, we get

$$|H(t;i,k) - H(s;i,k)| \leq |t_i - s_i| + |t_k - s_k|.$$

Case (3): In this case $t_i = s_i$ and $t_k = s_k$, and $H(t;i,k) = H(s;i,k)$. Thus, the equality sign holds true. □

4 Displacement function and slope

Let $q, p \in \mathbf{Z}$ with $0 < |p/q| < 1$. We define the displacement function $D = D(q,p) : \mathbf{X} \longrightarrow \mathbf{X}$ as

$$D(s_i) = D(q,p)(s_i) = H(i, i+q; s_i, s_i + pL)$$

for any $s_i = (i, s_i) \in \mathbf{X}$. This is equivalent to that $D(s_i) = H(i, i+q; s_i, U(q,p)(s_i))$ for any $s_i \in \mathbf{X}$. In this section D means $D(q,p)$ unless otherwise stated.

Since C is a strictly convex simple closed curve with length L, we have the periodic property of the displacement functions.

Lemma 4.1. *$D((i,s)) = D((j, s+kL))$ for any integers i, j, k and any $s \in \mathbf{R}$.*

Proposition 3.4 proves that the displacement functions are continuous in $\mathbf{X}$.

Lemma 4.2. $|D(s_i) - D(t_i)| \leq 2|t_i - s_i|$

Let $s = (s_j)_{j\in\mathbf{Z}}$ be a b-curve and $x_j = c(s_j)$ for all $j \in \mathbf{Z}$. We call $s = (s_j)_{j\in\mathbf{Z}}$ a *periodic b-curve with period* (q,p) if $s_{j+q} = s_j + pL$ for all $j \in \mathbf{Z}$. This implies that $x = (x_j)_{i\le j\le i+q}$ is a periodic billiard ball trajectory in C for each $i \in \mathbf{Z}$. The first variation formula and Lemma 2.1 show the following existence lemma of periodic billiard ball trajectories.

Lemma 4.3. *There exist* s_i *in* $\mathbf{X}$ *such that* $D(s_i) = \min\{D(s)|s \in \{i\}\times\mathbf{R}\}$. *If* $s = (s_j)_{i\le j\le i+q}$ *is a b-segment from* s_i *to* $s_{i+q} = s_i + pL$, *then the extension of* $s = (s_j)_{i\le j\le i+q}$ *as a b-curve is with period* (q,p).

Proof. The extension $s = (s_j)_{j\in\mathbf{Z}}$ is given by $s_j = U(q,p)^k(s_h) = (j, s_h + kpL)$ if $j = kq + h$, $0 \le h < k$. In fact, $s = (s_j)_{1\le j\le q+1}$ defined above is a b-curve since $H(s;1,q+1) = H(s;0,q)$. □

The periodic b-geodesic $s = (s_j)_{j\in\mathbf{Z}}$ is said to be *minimal* if $D(s_j) = \min\{D(s)|s \in \{j\}\times\mathbf{R}\}$.

Proposition 4.4. *Suppose* D *assumes its minimum at* s_i. *Then, there exists a unique minimal periodic b-geodesic through* s_i *with period* (q,p). *The minimal periodic b-geodesic is a b-straight line.*

Proof. Let $s = (s_j)_{j\in\mathbf{Z}}$ and $t = (t_j)_{j\in\mathbf{Z}}$ be two minimal periodic b-geodesics through s_i with period (q,p). Suppose $s_{i+1} < t_{i+1}$. Since $s = (s_j)_{i\le j\le i+q}$ and $t = (t_j)_{i\le j\le i+q}$ are b-segments, Proposition 3.1 proves that $T(s)\cap T(t) = \{s_i, s_{i+q}\}$. This means that $s_{i+q-1} < t_{i+q-1}$. However, $s_{i+q} = t_{i+q}$, $s_{i+q+1} < t_{i+q+1}$, since s and t are periodic. This is impossible because $(s_{i+q-1}, s_{i+q}, s_{i+q+1})$ and $(t_{i+q-1}, t_{i+q}, t_{i+q+1})$ are both b-curves and because of Lemma 2.5. Hence, $s_{i+1} = t_{i+1}$ and $s = t$.

By the same reasoning any minimal periodic b-geodesics with period (q,p) do not cross. Since minimality and periodicity are invariant under translations $U(k,l)$ in $\mathbf{X}$, $s' = U(k,l)s$ is also a minimal periodic b-geodesic with period (q,p) for any minimal periodic b-geodesic s with period (q,p). Therefore s' and s do not cross. Suppose a minimal periodic b-geodesic $s = (s_j)_{j\in\mathbf{Z}}$ with period (q,p) is not a b-straight line. Then, there is a positive integer n such that a sub-b-geodesic $s = (s_j)_{0\le j\le nq}$ is not a b-segment, and, in particular, $D(nq,np)(s_0) \neq \min D(nq,np)$. Let t_0 be such that $D(nq,np)(t_0) = \min D(nq,np)$ and $t = (t_j)_{j\in\mathbf{Z}}$ the minimal periodic b-geodesic with period (nq,np). Since $t^k = U(q,p)^k t$ is a minimal periodic b-geodesic with period (nq,np) for each $k \in \mathbf{Z}$, t^k do not intersect each other if $t^1 \neq t$. However, this is impossible because $t^n{}_{nq} = t_0 + npL = t_{nq}$. Therefore, $t^k = t$ for

any $k \in \mathbf{Z}^+$, and, in particular, $H(0, nq; t_0, t_{nq}) = H(t; 0, nq) = nH(t; 0, q)$. Hence, $nH(s; 0, q) = H(s; 0, nq) > H(t; 0, nq) = nH(t; 0, q)$. This is a contradiction to that $\min D(q, p) = H(s; 0, q) \leq H(t; 0, q)$. This states that s is a b-straight line. □

This proves the following.

Theorem 4.5. *If $D(q,p)$ is constant in $\mathbf{X}$, then $\mathbf{R}^2$ is foliated by minimal periodic b-geodesics with period (q,p) which are b-straight lines.*

The following shows the converse of Proposition 4.4.

Lemma 4.6. *If a periodic b-geodesic $s = (s_j)_{j \in \mathbf{Z}}$ with period (q,p) is a b-straight line, then it is minimal.*

Proof. Let $m = \min\{D(q,p)(u) | u \in \{i\} \times \mathbf{R}\}$. Let $t = (t_j)_{j \in \mathbf{Z}}$ be a minimal periodic b-geodesic with period (q,p). By Proposition 3.4 we have

$$\begin{aligned} nH(s; i, i+q) &= H(s; i, i+nq) \\ &< |t_i - s_i| + |t_{i+nq} - s_{i+nq}| + H(t; i, i+nq) \\ &= 2|t_i - s_i| + nH(t; i, i+q) \\ &= 2|t_i - s_i| + nm. \end{aligned}$$

Hence,

$$D(q,p)(s_0) = H(s; i, i+q) \leq \frac{2|t_i - s_i|}{n} + m,$$

and, therefore, we see that $D(q,p)(s_0) \leq m$ as $n \longrightarrow \infty$. Since m is the minimum of $D(q,p)$, it follows that $D(q,p)(s_0) = m$. □

It is stated in Theorem 9.37 of [17] that there exists a periodic b-curve with period (q,p) which is not minimal. The following is a sufficient condition for that there is a "maximal" periodic geodesic with period (q,p).

Proposition 4.7. *Let $s_i \in \{i\} \times \mathbf{R}$ be such that $D(s_i) = \max\{D(s) \mid s \in \{i\} \times \mathbf{R}\}$ and $s = (s_j)_{i \leq j \leq i+q}$ a b-segment from s_i to $s_{i+q} = s_i + pL$. Let $s = (s_j)_{j \in \mathbf{Z}}$ be the extension of s as a b-curve. Suppose a sub-b-segment $s = (s_j)_{i \leq j \leq q+1}$ is a b-segment. Then, $s = (s_j)_{j \in \mathbf{Z}}$ is a periodic b-geodesic with period (q,p).*

Proof. We will prove that $U(q,p)s_{i+1} = s_{i+q+1}$. Suppose $U(q,p)s_{i+1} \neq s_{i+q+1}$. Then, we have

$$H(s_{i+q}, s_{i+q+1}) + H(i+q+1, i+2q+1; s_{i+q+1}, U(q,p)s_{i+q+1}) \\ > H(U(q,p)s; i+q, i+2q+1),$$

and, hence,

$$H(i, i+q; s_{i+q+1}, U(q,p)s_{i+q+1}) \\ = H(i+q+1, i+2q+1; s_{i+q+1}, U(q,p)s_{i+q+1}) \\ > H(U(q,p)s; i+q, i+2q) > H(i, i+q; s_i, U(q,p)s_{i+q}),$$

contradicting the choice of s_i. □

Let $s = (s_j)_{j \geq i_0}$ ($s = (s_j)_{j \leq i_0}$, resp.) be a b-curve. We define the slope $\alpha(s)$ of s as follows.

$$\alpha(s) = \liminf_{j \to \infty} \frac{s_j}{j} \quad \left(\alpha(s) = \liminf_{j \to -\infty} \frac{s_j}{j}, \text{resp.} \right)$$

Proposition 4.8. *Let $s = (s_j)_{j \in \mathbf{Z}}$ be a periodic b-curve with period (q,p). Then,*

$$\alpha(s) = \lim_{j \to \pm\infty} \frac{s_j}{j} = \frac{pL}{q}.$$

Proof. Let $j = iq + k$ $(0 \leq k < q)$. Then,

$$\alpha(s) = \lim_{j \to \pm\infty} \frac{s_{iq+k}}{iq+k} = \lim_{i \to \pm\infty} \frac{s_k + ipL}{iq+k} = \frac{pL}{q}.$$

□

Lemma 4.9. *Let $s = (s_j)_{j \geq i_0}$ ($s = (s_j)_{j \leq i_0}$, resp.) be a ray. Then,*

$$\alpha(s) = \lim_{j \to \infty} \frac{s_j}{j} \quad (\alpha(s) = \lim_{j \to -\infty} \frac{s_j}{j}, \textit{resp.})$$

exists.

Proof. Suppose

$$\liminf_{j \to \infty} \frac{s_j}{j} < \frac{pL}{q} < \limsup_{j \to \infty} \frac{s_j}{j}$$

where $p, q \in \mathbf{Z}^+$. By Proposition 4.4 we have a periodic b-straight line $s' = (s'_j)_{j\in\mathbf{Z}}$ with period (q,p). Since $\alpha(s') = pL/q$ because of Proposition 4.8, the b-ray s intersects the b-straight line s' many times, contradicting Proposition 3.1.

The statement of the other case $s = (s_j)_{j\leq i_0}$ is proved in the same way. □

As an application we can prove the following.

Proposition 4.10. *Let $s = (s_j)_{j\in\mathbf{Z}}$ be a b-straight line. Then,*

$$\alpha(s) = \lim_{j\to\pm\infty} \frac{s_j}{j}$$

exists.

The slope is continuous in the set of all b-straight lines as seen in the following.

Lemma 4.11. *Let $s^n = (s^n{}_j)_{j\in\mathbf{Z}}$ be a sequence of b-straight lines. If the sequence s^n converges to a b-straight line $t = (t_j)_{j\in\mathbf{Z}}$, then $\alpha(s^n) \longrightarrow \alpha(t)$ as $n \longrightarrow \infty$.*

Proof. Suppose for indirect proof that $\limsup_{n\to\infty} \alpha(s^n) = \alpha > \alpha(t)$. Then, there exists a subsequence s^m such that $\alpha(s^m) > \alpha(t)$ converges to $\alpha > \alpha(t)$. We can find a periodic b-straight line $u = (u_j)_{j\in\mathbf{Z}}$ with period (q,p) such that $\alpha > pL/q = \alpha(u) > \alpha(t)$ and $s^m{}_j > u_j$ for any $j \geq 0$ and $t_j < u_j$ for sufficiently large j. The existence of such a b-straight line u implies that s^m cannot converges to t, a contradiction. Therefore, $\limsup_{n\to\infty} \alpha(s^n) \leq \alpha(t)$. In the same way we can prove that $\liminf_{n\to\infty} \alpha(s^n) \geq \alpha(t)$. These inequalities complete the proof. □

The following theorem due to Bangert ([1]) will be needed later. Here we give a proof which is different from his.

Theorem 4.12. (Theorem 3.13 in [1]) *Let $s = (s_j)_{j\in\mathbf{Z}}$ be a b-straight line, then either $U(q,p)s = s$ or $T(U(q,p)s) \cap T(s) = \emptyset$ for any $q, p \in \mathbf{Z}$ with $0 < p/q < 1$.*

Proof. Suppose $q > 0$, $T(U(q,p)s) \cap T(s) \neq \emptyset$ and $T(U(q,p)s) \neq T(s)$. Then, Proposition 3.1 states that $T(U(q,p)s) \cap T(s)$ consists of a single point. Let $i_0 \in \mathbf{Z}$ be such that the intersection is between i_0 and $i_0 + q$. We have two cases (1) $(U(q,p)s)_j > s_j$ for all $j > i_0 + q$ and (2) $(U(q,p)s)_j < s_j$ for all

$j > i_0 + q$. For the case (2) we can have the same contradiction as the case (1), so we take care of (1) only.

Let $s^k = U(q,p)^k s$ for $k \in \mathbf{Z}$. Then, ${s^k}_j = s_{j-kq} + kpL$ for all $j \in \mathbf{Z}$. Recall Lemma 3.2 which states that there exists a positive δ such that ${s^1}_j - s_j > \delta$ for all $j \geq i_0 + q$ and $s_j - {s^1}_j > \delta$ for all $j \leq i_0$. Since, for $n > 0$,

$$\begin{aligned} {s^n}_{i_0+nq} - s_{i_0+nq} &= \sum_{k=1}^{n} {s^k}_{i_0+nq} - {s^{k-1}}_{i_0+nq} \\ &= \sum_{k=1}^{n} {s^1}_{i_0+(n-k+1)q} - s_{i_0+(n-k+1)q} > n\delta, \end{aligned}$$

we have

$$\alpha(s) \leq \frac{p}{q}L - \frac{\delta}{q}.$$

On the other hand, we have, for $-n < 0$,

$$\begin{aligned} s_{i_0-nq} - {s^{-n}}_{i_0-nq} &= \sum_{k=0}^{n-1} {s^{-k}}_{i_0-nq} - {s^{-(k+1)}}_{i_0-nq} \\ &= \sum_{k=0}^{n-1} {s^1}_{i_0-(n-(k+1))q} - s_{i_0-(n-(k+1))q} < -n\delta, \end{aligned}$$

and, hence,

$$\alpha(s) \geq \frac{p}{q}L + \frac{\delta}{q},$$

contradicting the above inequality. □

The *diameter* d of C is by definition

$$d = \max\{|c(s) - c(t)| | s, t \in \mathbf{R}\}.$$

The diameter is characterized by a billiard ball trajectory as follows.

Lemma 4.13. *A b-straight line $s = (s_j)_{j\in\mathbf{Z}}$ is with period $(2,1)$ if and only if $|c(s_{j+1}) - c(s_j)|$ is the diameter of C for all $j \in \mathbf{Z}$.*

Proof. Suppose $|c(t_1) - c(t_0)|$ is the diameter of C where $0 \leq t_0 < L$ and $t_0 < t_1 < t_0 + L$. Let $t = (t_j)_{j\in\mathbf{Z}}$ be a b-curve which is the extension of (t_0, t_1).

Then, $t_{2j} = t_0 + jL$, $t_{2j+1} = t_1 + jL$ for any $j \in \mathbf{Z}$ which are the parameters of the endpoints of a diameter of C. Hence, we have

$$H(i,j;t_i,t_j) \geq \sum_{k=i}^{j-1} H(t_k, t_{k+1}) = H(t;i,j)$$

for all $i < j$. This implies that t is a periodic b-straight line with period (2,1), proving the "if" part.

The "only if" part follows from Lemma 4.6 and the definition of the diameter. In fact, $D(2,1)$ assumes its minimum at t_j for all $j \in \mathbf{Z}$ such that $c(t_j)$ is an endpoint of the diameter. □

Lemma 4.14. *Let $s = (s_j)_{j\in\mathbf{Z}}$ be a b-straight line. Suppose there exists an $i \in \mathbf{Z}$ such that the segment $T(c(s_i), c(s_{i+1}))$ of the billiard ball trajectory is perpendicular to $\dot{c}(s_i)$, namely $s_{i+1} = s_{i-1} + L$. Then, $|c(s_{j+1}) - c(s_j)|$ is the diameter of C for any $j \in \mathbf{Z}$.*

Proof. According to Theorem 4.12, we see $U(2,1)s = s$ since $s_{i+1} = s_{i-1} + L$, and, therefore, the b-straight line s has period (2,1). Lemma 4.13 states that $|c(s_{j+1}) - c(s_j)|$ is the diameter of C. □

The following lemma is obvious because of Proposition 2.9 and the definition of the slope, since the Birkhoff theorem states (Theorem 12.2.13 and Corollary 12.2.14 in [17]) that a closed simpule curve not null-homotopic in Ω which is invariant under φ is the graph of a Lipschitz continuous function $C \longrightarrow (-1,1)$.

Lemma 4.15. *Let f be a closed simple curve not null-homotopic in Ω invariant under the billiard ball map φ and let W be a foliation of $\mathbf{X}$ by b-curves which are determined by all points $\bar{x} \in f$. Then, those b-curves are b-straight lines and have the same slope.*

The following theorem due to Bangert ([1]) plays an important role in the proof of Theorem 1.1. Here M_a denotes the set of all b-straight lines with slope $\alpha = aL$ where $0 < a < 1$.

Theorem 4.16. (Theorem 4.1 in [1]) *If a is an irrational number with $0 < a < 1$, then M_a is totally ordered, namely either $t = s$ or $T(t) \cap T(s) = \emptyset$ holds true for any s and $t \in M_a$.*

Here is the proof of the theorem which is different from Bangert's. Lemmas 4.17 to 4.21 are for the proof of the theorem. Let

$$\Omega(s) = \{s^* \mid s^* = \lim_{n\to\infty} U(-j_n, k_n)s, j_n \to \infty\},$$

$$B(s) = \{U(q,p)s \mid q,p \in \mathbf{Z}\}$$

and $\overline{B(s)}$ its closure.

Lemma 4.17. $\limsup_{j\to\infty} |s_j - t_j| < \infty$.

Proof. Suppose there exists a sequence j_n such that $j_n \to \infty$, $t_{j_n} - s_{j_n} \to \infty$ as $n \to \infty$. If we set $s' = U(0,k)s$, then $s'_j - s_j = kL$. Hence we can find a b-straight line such that $T(s') \cap T(t) \neq \emptyset$ and $\alpha(s') = \alpha(s) = aL$. Since $(U(k,m)t)_j - (U(k,m)s')_j = t_{j-k} - s'_{j-k}$, there exist $k, m \in \mathbf{Z}$ such that $T(U(k,m)t)$ intersects $T(U(k,m)s')$ between 0 and 1. We denote them as t and s again instead of $U(k,m)t$ and $U(k,m)s'$. Let a positive integer q be such that $3L < t_q - s_q$. Let a positive integer $p \in \mathbf{Z}$ be such that $s_q + L < s_0 + pL < s_q + 2L$. Then, the interval $[s_0, s_0 + L]$ in $\{0\} \times \mathbf{R}$ is translated to the interval $[s_0 + pL, s_0 + (p+1)L]$ in $\{q\} \times \mathbf{R}$ by $U(q,p)$. Notice that it is contained in the interval $[s_q, t_q]$ in $\{q\} \times \mathbf{R}$. Let $u = (u_j)_{j\in\mathbf{Z}}$ be a minimal periodic geodesic with period (q,p) such that $u_0 \in [s_0, s_0 + L)$. Then, $T(u)$ intersects both $T(t)$ and $T(U(0,1)s)$ between 0 and q. Hence, we have the ineqality

$$aL = \alpha(t) > \alpha(u) = \frac{p}{q}L > \alpha(U(0,1)s) = \alpha(s) = aL.$$

a contradiction. □

Lemma 4.18. *Let $s, t \in M_a$. Then, there exist $s^* \in \Omega(s)$ and $t^* \in \Omega(t)$ such that $U(q,p)s^*$ does not intersect t^* for all $q, p \in \mathbf{Z}$.*

Proof. Let $j_n, k_n \in \mathbf{Z}$ be a sequence such that $(U(-j_n, k_n)t)_0 = t_{j_n} + k_n L$ converges in $\{0\} \times \mathbf{R}$ as $j_n \to \infty$. Since $|t_j - s_j|$ is bounded for $j > 0$ because of Lemma 4.17, there exist subsequnces j_m and k_m of j_n and k_n such that $s_{j_m} + k_m L$ converges in $\{0\} \times \mathbf{R}$. Then, the intersections $T(U(-j_m, k_m)t) \cap T(U(-j_m, k_m)s)$ go away to the backward infinity. Thus, $s^* = \lim_{m\to\infty} U(-j_m, k_m)s$ does not intersect $t^* = \lim_{m\to\infty} U(-j_m, k_m)t$. The same argument can be applied to $U(q,p)s$. □

Lemma 4.19. *Let $s, t \in M_a$. Assume that $U(q,p)s$ does not intersects t for all $q, p \in \mathbf{Z}$. Then, $\Omega(s) = \Omega(t)$.*

Proof. Let $u \in \Omega(s)$ and $u = \lim_{n\to\infty} U(-j_n, k_n)s$. Set $s^n = U(-j_n, k_n)s$. We can suppose without loss of generality that s^n is monotone in the sense that ${s^n}_j$ is monotone for each $j \in \mathbf{Z}$, say ${s^n}_j < {s^{n+1}}_j$. If we find $q_n, p_n \in \mathbf{Z}$ such that

$$ {s^n}_j \leq (U(q_n, p_n)t)_j \leq {s^{n+1}}_j $$

for all $j \in \mathbf{Z}$, we conclude that $u \in \Omega(t)$, and, hence, $\Omega(s) \subset \Omega(t)$. To prove this inequality, let $s^{n+1} = U(a,b)s^n$ for some $a, b \in \mathbf{Z}$. Let ${I^n}_i = U(-a,-b)^i([{s^n}_{-ai}, {s^{n+1}}_{-ai}])$ where $[{s^n}_{-ai}, {s^{n+1}}_{-ai}] \subset \{-ai\} \times \mathbf{R}$ for a positive i. Then, ${I^n}_i \subset \{0\} \times \mathbf{R}$. Let ${I^n}_i = [x_i, y_i]$ for all $i > 0$. Then, we see

$$ \begin{aligned} y_{i+1} &= (U(-a,-b)^{i+1}s^{n+1})_0 \\ &= {s^{n+1}}_{-a(i+1)} - b(i+1)L \\ &= (U(a,b)s^n)_{-a(i+1)} - b(i+1)L \\ &= {s^n}_{-ai} - biL \\ &= (U(-a,-b)^i s^n)_0 \\ &= x_i \end{aligned} $$

Since $x_i = y_{i+1} > x_{i+1}$ for all $i > 0$, the sequence x_i is monotone decreasing. It is not bounded from below. In fact, if it is not true, then we have a limit point z, and, hence,

$$ z = \lim_{i\to\infty} (U(-a,-b)^i s^n)_0 = U(a,b) \lim_{i\to\infty} (U(-a,-b)^{i-1}s^n)_0 = U(a,b)z, $$

which implies that the limit b-straight line u has slope $\alpha(u) = bL/a$, contradicting the continuity of the slope. So we have

$$ \cup_{i=0}^{\infty} {I^n}_i = (-\infty, s_0^{n+1}]. $$

Thus, there exist k_0 and i_0 such that $t_0 + k_0 L \in {I^n}_{i_0}$. Then,

$$ (U(-a,-b)^{i_0}s^n)_0 \leq t_0 + k_0 L \leq (U(-a,-b)^{i_0}s^{n+1})_0 $$

and, in other words,

$$ {s^n}_j \leq (U(a,b)^{i_0}U(0,k_0)t)_j \leq {s^{n+1}}_j $$

for all $j \in \mathbf{Z}$ because of our assumption. We set $q_n = ai_0$ and $p_n = bi_0 + k_0$.

The other inclusion relation $\Omega(t) \subset \Omega(s)$ is proved by the same way. □

Lemma 4.20. *Let $s, t \in M_a$. Then, $\Omega(s) = \Omega(t)$.*

Proof. Let $s^* \in \Omega(s)$, $t^* \in \Omega(t)$ be as in Lemma 4.19 for s and t. Then, each pair s and s^*, s^* and t^*, t^* and t satisfies the assumption of Lemma 4.17, and, hence, we have

$$\Omega(s) = \Omega(s^*) = \Omega(t^*) = \Omega(t)$$

□

Set $\Omega(a) = \Omega(s)_0$ where $s \in M_a$ and $\Omega(s)_0 = \{s^*{}_0 \in \mathbf{R} \mid s^* \in \Omega(s)\}$. Notice that $\Omega(a) \subset \overline{B(t)}$ for any $t \in \overline{B(s)}$. Let $\cup_\lambda O^\lambda = \mathbf{R} - \Omega(a)$ and $O^\lambda = (x^\lambda, y^\lambda)$. Since $x^\lambda, y^\lambda \in \Omega(a)$, we see that there exist u^λ and v^λ in $\overline{B(t)}$ such that $u^\lambda{}_0 = x^\lambda$ and $v^\lambda{}_0 = y^\lambda$. Notice that there is no b-straight line with slope aL in the strip $[T(u^\lambda), T(v^\lambda)]$.

Lemma 4.21. *$v^\lambda{}_i - u^\lambda{}_j \to 0$ as $j \to \pm\infty$.*

Proof. Since a is irrational, $[U(-j,k)u^\lambda{}_j, U(-j,k)v^\lambda{}_j]$ in $\{0\} \times \mathbf{R}$ are mutually disjoint. Moreover, when they are translated into $\{0\} \times [u^\lambda{}_0, u^\lambda{}_0 + L]$ by some $U(0,m)$, they are also mutually disjoint. Therefore, we have $\sum_{j=-\infty}^{\infty} |u^\lambda{}_j - v^\lambda{}_j| \leq L$. The lemma follows from this. □

Proof of Theorem 4.16: Let $s, t \in M_a$. Suppose $s_0 \in O^\lambda$ and $t_0 \in O^\mu$. If $\lambda \neq \mu$, then $s \in [T(u^\lambda), T(v^\lambda)]$ and $t \in [T(u^\mu), T(v^\mu)]$. Hence $T(s)$ does not intersect $T(t)$, since $O^\lambda \cap O^\mu = \emptyset$. If $\lambda = \mu$, then $|s_j - t_j| \leq |u^\lambda{}_j - v^\lambda{}_j| \to 0$ as $j \to \pm\infty$. It follows from Lemma 3.2 that $T(s)$ does not intersect $T(t)$ if $s \neq t$. □

Since the slope is invariant under all translations, the following is a direct consequence of Theorem 4.16.

Theorem 4.22. *Let $\alpha = aL$ where a is an irrational number with $0 < a < 1$ Suppose there passes a b-straight line through any point $s_0 \in \{0\} \times \mathbf{R}$ with slope α. Then, those b-straight lines yield a foliation of $\mathbf{X}$ which is invariant under all translations, and, therefore, corresponds to a closed simple curve not null-homotopic in Ω invariant under the billiard ball map φ.*

We say that a b-ray $s = (s_j)_{j \geq i_1}$ is *maximal* if any extension $t = (t_j)_{j \geq i_2}$, $i_2 < i_1$, of s as a b-curve is not a b-ray.

Lemma 4.23. *Let $\alpha = aL$ where a is a number with $0 < a < 1$. Suppose there exists a unique b-ray $s = (s_j)_{j \geq i}$ from any point s_i in $\mathbf{X}$ with slope α. Then, any maximal b-ray with slope α is a b-straight line, and, hence, those straight lines yield a foliation of $\mathbf{X}$ which is invariant under all translations.*

Proof. Suppose there exists a maximal b-ray $s = (s_j)_{j \geq i_0}$ which is not a b-straight line. Let k be an integer such that $k \leq s_{i_0} < k + L$. For each t_{i_0-1} with $k - L \leq t_{i_0-1} < k + L$ let $t = (t_j)_{j \geq i_0-1}$ be a unique b-ray from t_{i_0-1} with slope α as in the assumption. Since t_{i_0} continuously depends on t_{i_0-1}, there exists a $t_{i_0-1} = s_{i_0-1}$ such that $t_{i_0} = s_{i_0}$. The uniqueness of b-rays from s_{i_0} with slope α implies that the b-ray $t = (t_j)_{j \geq i_0-1}$ with $t_{i_0-1} = s_{i_0-1}$ is an extension of $s = (s_j)_{j \geq i_0}$, contradicting the maximal property of s. □

5 Billiard asymptote

Let $s = (s_j)_{j \geq i_0}$ be a b-ray. In order to develop the theory of parallels due to H. Busemann, we first define the Busemann function of a b-ray s in the configuration space as

$$B_s(i, t_i) = B_s(t_i) = \lim_{n \to \infty} \{H(i, n; t_i, s_n) - H(s; i_0, n)\}$$

for any $(i, t_i) \in \mathbf{X}$ (see [2], [6], [10]). Lemmas 5.1 to 5.5 prove that the limit of the right-hand side exists, and, hence, the Busemann function is defined. In the same way we define the Busemann function of a b-ray $s = (s_j)_{j \leq i_0}$ as $n \to -\infty$ instead of "$n \longrightarrow \infty$". We states the properties and proofs for only the case $s = (s_j)_{j \geq i_0}$. However, the same properties are true under the suitable change of the expression unless otherwise stated.

Lemma 5.1. $\lim_{j \to \infty} s_j = \infty$.

Proof. If $\{s_j\}_{j \geq i_0}$ is bounded above, then s_j converges to a real number a as $j \longrightarrow \infty$, since $\{s_j\}_{j \geq i_0}$ is monotone increasing. In particular, $|s_j - s_{j+1}| \longrightarrow 0$ as $j \longrightarrow \infty$. Then, $H(s_j, s_{j+1}) \longrightarrow 0$ as $j \longrightarrow \infty$. Let s'_{i_0+1} be a point in $\mathbf{X}$ such that

$$H(s_{i_0}, s_{i_0+1}) > H(s_{i_0}, s'_{i_0+1}) + H(s'_{i_0+1}, s_{i_0+1}).$$

Then, there exists a k_0 such that

$$\begin{aligned} H(s_{i_0}, s_{i_0+1}) - H(s_{i_0}, s'_{i_0+1}) - H(s'_{i_0+1}, s_{i_0+1}) \\ > H(s_{k_0-2}, s_{k_0}) - H(s_{k_0-2}, s_{k_0-1}) - H(s_{k_0-1}, s_{k_0}). \end{aligned}$$

Let $s' = (s'_j)_{j \geq i_0}$ be given by $s'_{i_0} = s_{i_0}$, $s'_{i_0+1} = s'_{i_0+1}$, $s'_j = s_{j-1}$ for $i_0 + 2 \leq j \leq k_0 - 1$ and $s'_j = s_j$ for $k_0 \leq j$. Then, we have

$$\begin{aligned} H(s'; i_0, k_0) &= \sum_{j=i_0}^{k_0-1} H(s'_j, s'_{j+1}) \\ &= H(s_{i_0}, s'_{i_0+1}) + H(s'_{i_0+1}, s_{i_0+1}) \\ &\quad + \sum_{j=i_0+1}^{k_0-3} H(s_j, s_{j+1}) + H(s_{k_0-2}, s_{k_0}) \\ &< H(s_{i_0}, s_{i_0+1}) + \sum_{j=i_0+1}^{k_0-3} H(s_j, s_{j+1}) \\ &\quad + H(s_{k_0-2}, s_{k_0-1}) + H(s_{k_0-1}, s_{k_0}) \\ &= H(s; i_0, k_0). \end{aligned}$$

This contradicts that s is a b-ray. □

Lemma 5.2. $\ell := \limsup_{j \to \infty}(s_{j+1} - s_j) < L$.

Proof. Suppose $\ell = L$. Then, there exists a k_0 such that

$$\begin{aligned} &H(s_{i_0}, s_{i_0+1}) - H(s_{i_0}, s'_{i_0+1}) - H(s'_{i_0+1}, s_{i_0+1}) \\ >&H(s_{k_0-2}, s_{k_0}) - H(s_{k_0-2}, s_{k_0-1}) - H(s_{k_0-1}, s_{k_0}), \end{aligned}$$

since $s_{j+1} - s_j = L$ implies that $H(s_{j+1}, s_j) = 0$. From this we can get the same contradiction as in the proof of Lemma 5.1. □

Lemma 5.3. *There exists a b-segment from (i, t) to (n, s_n) for sufficiently large n.*

Proof. Let $\ell < \ell' < L$. There exists a k_0 such that $s_{k+1} - s_k \leq \ell'$ for all $k \geq k_0$. Hence, for any $(i, t) \in \mathbf{X}$ we have

$$\begin{aligned} s_n - t &= (s_n - s_{n-1}) + \cdots + (s_{k_0+1} - s_{k_0}) + (s_{k_0} - t) \\ &< \ell'(n - k_0) + (s_{k_0} - t) < L(n - i) \end{aligned}$$

for sufficiently large n. Lemma 2.7 proves the existence of a b-segment from (i, t) to (n, s_n). □

Lemma 5.4. *There exists a k_0 such that $H(i,n;t,s_n) - H(s;i_0,n)$ is monotone decreasing in $n \geq k_0$.*

Proof.

$$\begin{aligned}
&H(i,n;t,s_n) - H(s;i_0,n) \\
=&\{H(i,n;t,s_n) + H(s_n,s_{n+1})\} - \{H(s;i_0,n) + H(s_n,s_{n+1})\} \\
\geq&H(i,n+1;t,s_{n+1}) - H(s;i_0,n+1).
\end{aligned}$$

□

Let t^n be a sequence of b-segments from (i,t) to s_n. Choose a subsequence t^m which converges to a b-ray, say $u = (u_j)_{j \geq i}$. Since u is a b-ray, s_0 can be joinned to u_{k_0} by a b-segment for sufficiently large k_0.

Lemma 5.5. *$H(i,n;t,s_n) - H(s;i_0,n)$ is bounded below.*

Proof. Since $H(s;i_0,m) \leq H(i_0,k_0;s_{i_0},{t^m}_{k_0}) + H(k_0,m;{t^m}_{k_0},s_m)$, we have

$$H(i,m;t,s_m) - H(s;i_0,m) \geq H(i,k_0;t,{t^m}_{k_0}) - H(i_0,k_0;s_{i_0},{t^m}_{k_0}).$$

The right-hand side converges to some number as $m \longrightarrow \infty$. Hence, the left-hand side is bounded below for all n because of Lemma 5.4. □

These prove that the Busemann function of a b-ray s can be defined. The following lemma states that Busemann functions are continuous.

Lemma 5.6. *Let $s = (s_j)_{j \geq i_0}$ be a b-ray. Then,*

$$|B_s(i,t_i) - B_s(i,u_i)| \leq |t_i - u_i|$$

for any $(i,t_i),(i,u_i) \in \mathbf{X}$.

Proof. This is because

$$\begin{aligned}
&|\{H(i,n;t_i,s_n) - H(s;i_0,n)\} - \{H(i,n;u_i,s_n) - H(s;i_0,n)\}| \\
=&|H(i,n;t_i,s_n) - H(i,n;u_i,s_n)| \leq |t_i - u_i|.
\end{aligned}$$

The last inequality comes from Proposition 3.4. □

The following lemma shows that a Busemann function is invariant under any translation.

Lemma 5.7. *Let* $s' = U(q,p)s$, *namely,* ${s'}_j = s_{j-q} + pL$. *Then,*

$$B_s(i-q, t-pL) = B_{s'}(i,t)$$

for any $(i,t) \in \mathbf{X}$.

We remember how to find a parallel to a straight line K in the Euclidean geometry. The way of defining the Busemann functions follows this process. A Busemann function is like a height function along a straight line K.

Lemma 5.8. *Let* $t, u \in \mathbf{R}$ *and* $i \in \mathbf{Z}$, $m \in \mathbf{Z}^+$ *such that* $t < u < t + mL$. *Then,*

$$B_s(i,t) \leq H(i, i+m; t, u) + B_s(i+m, u)$$

Proof.

$$\begin{aligned} B_s(i,t) &= \lim_{n\to\infty} \{H(i,n;t,s_n) - H(s;i_0,n)\} \\ &\leq \lim_{n\to\infty} \{H(i,i+m;t,u) + H(i+m,n;u,s_n) - H(s;i_0,n)\} \\ &= H(i,i+m;t,u) + B_s(i+m,u) \end{aligned}$$

□

In the Euclidean geometry the equality sign holds true if u is in the parallel line through t to s. This lemma motivates us to define something just like parallels. Let $t = (t_j)_{j \geq i_1}$ be a C-curve. We say that t is a *co-b-ray* to a b-ray $s = (s_j)_{j \geq i_0}$ if

$$B_s(i,t_i) = H(i, i+m; t_i, t_{i+m}) + B_s(i+m, t_{i+m})$$

for any $i \geq i_1$ and $m > 0$.

We say that a C-curve $t = (t_j)_{j \in \mathbf{Z}}$ is a *b-asymptote* to a b-ray $s = (s_j)_{j \geq i_0}$ if any sub-b-curve $t = (t_j)_{j \geq i}$ of t is a co-b-ray to s. We say that a C-curve $t = (t_j)_{j \in \mathbf{Z}}$ is a *b-parallel* to a b-straight line $s = (s_j)_{j \in \mathbf{Z}}$ if sub-b-curves $(t_j)_{j \geq i}$ and $(t_j)_{j \leq i}$ are co-b-rays to b-rays $(s_j)_{j \geq 0}$ and $(s_j)_{j \leq 0}$ respectively for each $i \in \mathbf{Z}$.

First we should prove the following.

Lemma 5.9. *If* $t = (t_j)_{j \geq i_1}$ *is a co-b-ray to a b-ray* $s = (s_j)_{j \geq i_0}$, *then* t *is a b-ray.*

Proof. For each $j \geq i_1$ and $m > 0$ we have

$$\begin{aligned} H(j, j+m; t_j, t_{j+m}) &= B_s(j, t_j) - B_s(j+m, t_{j+m}) \\ &= \sum_{k=0}^{m-1} \{B_s(j+k, t_{j+k}) - B_s(j+k+1, t_{j+k+1})\} \\ &= \sum_{k=0}^{m-1} H(t_{j+k}, t_{j+k+1}) \\ &= H(t; j, j+m). \end{aligned}$$

□

Lemma 5.10. *Let $s = (s_j)_{j \geq i_0}$ and $t = (t_j)_{j \geq i_1}$ be b-rays. If $\lim_{j \to \infty} |s_j - t_j| = 0$, then $B_s(i, u) = B_t(i, u) - B_t(i_0, s_{i_0})$ for any $(i, u) \in \mathbf{X}$ and they are co-b-rays to each other.*

Proof. For any $(i, u) \in \mathbf{X}$ we have

$$\begin{aligned} B_s(i, u) &= \lim_{n \to \infty} \{H(i, n; u, s_n) - H(s; i_0, n)\} \\ &= \lim_{n \to \infty} \{H(i, n; u, t_n) - H(i_0, n; s_{i_0}, t_n)\} \\ &= B_t(i, u) - B_t(i_0, s_{i_0}) \end{aligned}$$

The following proves the remaining part.

$$\begin{aligned} B_s(i, t_i) &= B_t(i.t_i) - B_t(i_0, s_{i_0}) \\ &= H(i, i+m; t_i, t_{i+m}) + B_t(i+m, t_{i+m}) - B_t(i_0, s_{i_0}) \\ &= H(i, i+m; t_i, t_{i+m}) + B_s(i+m, t_{i+m}). \end{aligned}$$

□

If $t^n = (t^n{}_j)_{i_1 \leq j \leq n}$ be a b-segment from $t^n{}_{i_1}$ to s_n and a sequence $t^n{}_{i_1}$ is bounded, then there exists a sub-sequence t^m which converges a b-ray. It is a parallel in Euclidean geometry. In our geometry we have the same result.

Lemma 5.11. *Let $t^m = (t^m{}_j)_{i_1 \leq j \leq n}$ be a b-segment from $t^m{}_{i_1}$ to s_n. If a sequence t^m converges to a b-ray $t = (t_j)_{j \geq i_1}$, then t is a co-b-ray to s.*

Proof. Since

$$\begin{aligned}&H(i,m;{t^m}_i,s_m)-H(s;i_0,m)\\=&H(i,i+k;{t^m}_i,{t^m}_{i+k})+H(i+k,m;{t^m}_{i+k},s_m)-H(s;i_0,m)\end{aligned}$$

for all $k \in \mathbf{Z}^+$, we have

$$B_s(i,t_i)=H(i,i+k;t_i,t_{i+k})+B_s(i+k,t_{i+k}).$$

This equation states that t is a co-b-ray to s. □

The following shows that sub-b-rays of a co-b-ray t are the unique co-b-rays if the starting point is not the terminal point of t.

Proposition 5.12. *Let $t=(t_j)_{j\geq i_1}$ be a co-b-ray to s and let $i_2 > i_1$. If $u=(u_j)_{j\geq i_2}$ is a co-b-ray to s with $u_{i_2}=t_{i_2}$, then u is a sub-b-ray of t, namely, $u_j=t_j$ for $j \geq i_2$.*

Proof. Since t and u are b-curves we have only to prove that $(t_{i_2-1},t_{i_2},u_{i_2+1})$ is a b-curve (see Lemma 2.2). Since

$$\begin{aligned}&H(i_2-1,i_2+1;t_{i_2-1},u_{i_2+1})\\\geq&B_s(i_2-1,t_{i_2-1})-B_s(i_2+1,u_{i_2+1})\\=&H(t_{i_2-1},t_{i_2})+B_s(i_2,t_{i_2})-B_s(i_2+1,u_{i_2+1})\\=&H(t_{i_2-1},t_{i_2})+H(u_{i_2},u_{i_2+1})\\\geq&H(i_2-1,i_2+1;t_{i_2-1},u_{i_2+1}),\end{aligned}$$

we have

$$H(i_2-1,i_2+1;t_{i_2-1},u_{i_2+1})=H(t_{i_2-1},t_{i_2})+H(t_{i_2},u_{i_2+1}).$$

Combined with Lemma 2.1, this completes the proof. □

We have a property of a b-parallel.

Proposition 5.13. *Let $s=(s_j)_{j\in\mathbf{Z}}$ be a b-straight line. If a b-curve $t=(t_j)_{j\in\mathbf{Z}}$ is a b-parallel to s, then $B_s(j,t_j)+B_{-s}(j,t_j)$ is a constant for $j \in \mathbf{Z}$ where -s is a b-ray $(s_j)_{j\leq i_0}$*

Proof. Since $t = (t_j)_{j\in\mathbf{Z}}$ is a b-parallel to s, we have that

$$B_s(i,t_i) = H(i,i+m;t_i,t_{i+m}) + B_s(i+m,t_{i+m})$$
$$B_{-s}(i+m,t_{i+m}) = H(i,i+m;t_i,t_{i+m}) + B_{-s}(i,t_i)$$

for all $i \in \mathbf{Z}$ and $m \in \mathbf{Z}^+$ Therefore, we see that

$$B_s(i,t_i) + B_{-s}(i,t_i) = B_s(i+m,t_{i+m}) + B_{-s}(i+m,t_{i+m})$$

This completes the proof. □

The following lemma states that we can alter "any m" in the definition of co-b-rays to "some".

Lemma 5.14. *Let $t = (t_j)_{j\geq i_1}$ be a b-curve such that*

$$B_s(i_1,t_{i_1}) = H(i_1,i_1+k;t_{i_1},t_{i_1+k}) + B_s(i_1+k,t_{i_1+k})$$

for some $k > 0$. Then, t is a co-b-ray to s.

Proof. Let $u = (u_j)_{j\geq i_1+k}$ be a co-b-ray to s with $u_{i_1+k} = t_{i_1+k}$. We first claim that u is a sub-b-curve of t. In fact,

$$\begin{aligned}
&H(i_1,i_1+k+1;t_{i_1},u_{i_1+k+1})\\
\geq& -B_s(i_1+k+1,u_{i_1+k+1}) + B_s(i_1,t_{i_1})\\
=& -B_s(i_1+k+1,u_{i_1+k+1}) + B_s(i_1+k,t_{i_1+k}) + H(i_1,i_1+k;t_{i_1},t_{i_1+k})\\
=&H(u_{i_1+k},u_{i_1+k+1}) + H(i_1,i_1+k;t_{i_1},t_{i_1+k})\\
\geq&H(i_1,i_1+k+1;t_{i_1},u_{i_1+k+1}),
\end{aligned}$$

and, in particular, $(t_{i_1},\ldots,t_{i_1+k},u_{i_1+k+1})$ is a b-segment. This proves the first claim.

We will show that the condition on co-b-rays is satisfied. Recall that u is a co-b-ray to s. The equation in the definition of co-b-rays is satisfied for $i \geq i_1 + k$ and $m > 0$. Let $i_1 \leq i < i_1 + k$ and $m > 0$. Then,

$$\begin{aligned}
&B_s(i,t_i)\\
\geq& -H(i_1,i;t_{i_1},t_i) + B_s(i_1,t_{i_1})\\
=& -H(i_1,i;t_{i_1},t_i) + H(i_1,i_1+k;t_{i_1},t_{i_1+k}) + B_s(i_1+k,t_{i_1+k})\\
=&H(i,i_1+k;t_i,t_{i_1+k}) + B_s(i_1+k,t_{i_1+k})\\
=&H(i,i_1+k;t_i,t_{i_1+k}) + H(i_1+k,i_1+k+m;t_{i_1+k},t_{i_1+k+m})\\
&\qquad + B_s(i_1+k+m,t_{i_1+k+m})\\
\geq&H(i,i_1+k+m;t_i,t_{i_1+k+m}) + B_s(i_1+k+m,t_{i_1+k+m})\\
\geq&B_s(i,t_i),
\end{aligned}$$

and, thus,

$$B_s(i,t_i) = H(i,i+m;t_i,t_{i+m}) + B_s(i+m,t_{i+m})$$

for any $i+m \geq i_1+k$. For the remaining case we suppose $i+m < i_1+k$. Then

$$\begin{aligned}
&B_s(i,t_i)\\
\geq& -H(i_1,i;t_{i_1},t_i) + B_s(i_1,t_{i_1})\\
=& -H(i_1,i;t_{i_1},t_i) + H(i_1,i_1+k;t_{i_1},t_{i_1+k}) + B_s(i_1+k,t_{i_1+k})\\
=&H(i,i+m;t_{i_1},t_{i+m}) + H(i+m,i_1+k;t_{i+m},t_{i+k}) + B_s(i_1+k,t_{i_1+k})\\
\geq&H(i,i+m;t_i,t_{i+m}) + B_s(i+m,t_{i+m})\\
\geq&B_s(i,t_i).
\end{aligned}$$

These prove that t satisfies the condition on co-b-rays. □

As an application we prove the following.

Proposition 5.15. *Let $s = (s_j)_{j\in\mathbf{Z}}$ and $s' = (s'_j)_{j\in\mathbf{Z}}$ be periodic b-curves with period (q,p). If s and s' are b-straight lines, then one is a b-parallel to the other.*

Proof. By Lemma 4.6 we have that $H(s;i,i+q) = H(s';i,i+q)$. Hence, we have

$$\begin{aligned}
B_s(i+q,s'_{i+q}) &= \lim_{n\to\infty}\{H(i+q,n;s'_{i+q},s_n) - H(s;0,n)\}\\
&= \lim_{n\to\infty}\{\{H(i,n-q;s'_i,s_{n-q}) - H(s;0,n-q)\}\\
&\qquad + \{H(s;0,n-q) - H(s;0,n)\}\}\\
&= B_s(i,s'_i) - H(s;i,i+q)\\
&= B_s(i,s'_i) - H(s';i,i+q)\\
&= B_s(i,s'_i) - H(i,i+q;s'_i,s'_{i+q}).
\end{aligned}$$

The Lemma follows from Lemma 5.14. □

Proposition 5.16. *Let $s = (s_j)_{j\in\mathbf{Z}}$ be a periodic b-straight line with period (q,p) and $t = (t_j)_{j\in\mathbf{Z}}$ a b-straight line with slope $\alpha(t) = pL/q$ and $T(t)\cap T(s) \neq \emptyset$. Then, t coincides with s.*

Proof. Let $s' = (s'_j)_{j\in\mathbf{Z}}$ where $s'_j = s_j - L$ for all $j \in \mathbf{Z}$. Let t and s cross at j_0 or between j_0 and $j_0 + 1$. Suppose $s_j > t_j$ for all $j > j_0 + 1$ because the statement for the other case is proved in the same way. Combined with Lemma 5.11 and Proposition 5.12, Proposition 5.15 states that a sequence of b-segments $s^n = (s^n{}_j)_{j_0-1\le j\le n}$ from s_{j_0-1} to s'_n converges to a sub-b-ray $s = (s_j)_{j\ge j_0-1}$ of s. Hence, t crosses s', otherwise we have a contradiction that s^n crosses t twice for sufficiently large n. Thus, there exists a $c \in \mathbf{Z}^+$ such that $s'_j > t_j$ for all $j > qc+j_0$. For each $j > qc+j_0$ we see that $(U(qc, pc-1)^k s)_j = s'_j$ is monotone decreasing for $k \in \mathbf{Z}^+$ and $(U(qc, pc-1)^k t)_j$ is monotone increasing for $k \in \mathbf{Z}^+$ because of Theorem 4.12 (see Theorem 3.13 in [1]). Therefore, we have the following.

$$\begin{aligned} s_{kqc+j_0} - t_{kqc+j_0} &\ge s_{j_0} + kpcL - t_{kqc+j_0} \\ &\ge s_{j_0} + k(pc-1)L - (U(qc, pc-1)^{k-1}t)_{kqc+j_0} + kL \\ &= (U(qc, pc-1)^k s)_{kqc+j_0} - (U(qc, pc-1)^{k-1}t)_{kqc+j_0} + kL \\ &= (U(qc, pc-1)s)_{qc+j_0} - t_{qc+j_0} + kL \\ &\ge kL \end{aligned}$$

for all $k \in \mathbf{Z}^+$. Thus, we have

$$\alpha(s) - \alpha(t) \ge \frac{L}{qc},$$

contradicting that $\alpha(t) = pL/q$. Therefore, t coincides with s. □

Combined with Lemma 4.15 and Theorem 4.16 (see Theorem 4.1 in [1]), the following lemma is a simple modification of Theorem 4.12 in [2] which is important in the proof of the differentiability of a leaf f with irrational slope.

Lemma 5.17. *Let a be an arbitrary irrational number with $0 < a < 1$. Suppose there exists a φ-invariant closed simple curve f not null-homotopic in Ω such that $\alpha(\bar{x}) = aL$ for all $\bar{x} \in f$. Let W be a foliation of $\mathbf{X}$ by b-curves which is determined by all points $\bar{x} \in f$. Then, all b-straight lines in W are b-parallels to each other.*

6 Billiard parallel and caustic

The purpose of this section is to prove Theorem 6.10.

Let $0 = (0,0)$ and $A = (0,a)$ $(a > 0)$ in Euclidean plane $\mathbf{E}$. Let $\alpha(s) = (x_1(s), y_1(s))$ and $\beta(s) = (x_2(s), y_2(s))$ be curves in $\mathbf{E}$ such that $\alpha(0) = 0$, $\beta(0) = A$, and α, β are of class C^1 with ${x_1}'(0) > 0$, ${x_2}'(0) < 0$. Let $t(s)$ be the y-coordinate of the intersection point $T(0,A) \cap T(\alpha(s), \beta(s))$ where ${x_i}'$ means the derivative of x_i with respect to its parameter.

Lemma 6.1.

$$t(0) = \frac{\det(\alpha'(0)\,\beta(0))}{{x_1}'(0) - {x_2}'(0)}$$

Proof. We have that

$$t(s) = \frac{x_1(s)y_2(s) - y_1(s)x_2(s)}{x_1(s) - x_2(s)} = \frac{\det(\alpha(s)\,\beta(s))}{x_1(s) - x_2(s)}$$

for any s sufficiently close to 0. This proves the lemma. □

This lemma shows the following.

Lemma 6.2. *Let $\psi : C \longrightarrow C$ be an orientation preserving homeomorphism such that ψ is of class C^1. Let $g(s) = \psi(c(s))$ for all $s \in \mathbf{R}$. Then, there exists a closed continuous curve K in the domain bounded by C such that any segment $T(c(s), g(s))$ is a tangent line to K.*

Let B be the closed domain bounded by C. Let $x = (x_j)_{j \in \mathbf{Z}}$ be a billiard ball trajectory in C and let $\gamma : (-\infty, \infty) \longrightarrow B$ be the unit speed broken segments such that $\gamma(t_j) = x_j$ for all $j \in \mathbf{Z}$. Let $Q = Q_j$ be the reflection with respect to the tangent line to C at $\gamma(t_j)$ which is by definition $Q(X) = X - 2 < X, N > N$ where X is any vector at $\gamma(t_j)$ and N is the inward unit normal vector to C. Then, $\dot{\gamma}(t_j + 0) = Q(\dot{\gamma}(t_j - 0))$. Let the angle between $\dot{c}(t_j)$ and $T(x_j, x_{j+1})$ be θ_j for any $j \in \mathbf{Z}$. We say that $Y(t)$, $-\infty < t < \infty$, is a *perpendicular Jacobi vector field along* γ if Y satisfies the following (see [16]).

1. Y is of class C^∞, $Y''(t) = 0$ and $\langle \dot{\gamma}(t), Y(t) \rangle = 0$ in each interval $[t_j, t_{j+1}]$.
2. $Y(t_j + 0) = Q(Y(t_j - 0))$ for any $j \in \mathbf{Z}$.
3. $Q(Y'(t_j - 0)) - Y'(t_j + 0) = (2\kappa(t_j)/\sin\theta_j)Y(t_j + 0)$ where $\kappa(t_j)$ is the geodesic curvature of C at $\gamma(t_j)$ with respect to the inward unit normal vector to C.

Let $\gamma_u : (-\infty, \infty) \longrightarrow B$ be a variation through billiard ball trajectories with unit speed such that $\gamma_0(t) = \gamma(t)$ for any $t \in (-\infty, \infty)$. Let

$$Y(t) = \left.\frac{\partial \gamma_u}{\partial u}\right|_{u=0}(t)$$

for any $t \in (-\infty, \infty)$. If $\langle Y(a), \gamma'(a)\rangle = 0$ for some $a \in \mathbf{R}$, then $Y(t)$ is a perpendicular Jacobi vector field along γ. Any perpendicular Jacobi vector field is given in this way. Let K be the envelope curve of a variation γ_u through billiard ball trajectories with unit speed and let γ be tangent to K at a_λ, $\lambda \in \Lambda$. Then, $\gamma(a_\lambda)$ are conjugate points to each other along γ, since the perpendicular component of the variation vector field Y is a nontrivial perpendicular Jacobi vector field with $Y(a_\lambda) = 0$ for any $\lambda \in \Lambda$. We sometimes call such points *focal points to C* along γ.

We say that the conjugate points of a nontrivial perpendicular Jacobi vector field $Y(t)$, $-\infty < t < \infty$, along γ *separate the boundary* if there exists a sequence $\{a_j\}_{j \in \mathbf{Z}}$ such that $\gamma(a_j)$ lie in $T(x_j, x_{j+1})$ and $Y(a_j) = 0$ for any $j \in \mathbf{Z}$. Let B be the domain surrounded by C in $\mathbf{E}$. Let $\gamma_u : (-\infty, \infty) \longrightarrow B$ be a variation through billiard ball trajectories such that the b-curves in $\mathbf{X}$ for all γ_u correspond to b-asymptotes to the b-straight line in $\mathbf{X}$ for $\gamma = \gamma_0$. Then, $T(x(u)_j, x(u)_{j+1})$ intersects $T(x_j, x_{j+1})$ for any $j \in \mathbf{Z}$ where $x(u)_j = \gamma_u(t_j)$. From this it follows that there exists a nontrivial perpendicular Jacobi vector field along γ whose conjugate points separate the boundary.

Let J be the set of all nontrivial perpendicular Jacobi vector fields along γ whose conjugate points separate the boundary. We prove the following.

Lemma 6.3. *Let $\gamma : (-\infty, \infty) \longrightarrow B$ be a billiard ball trajectory which corresponds to a b-straight line in $\mathbf{X}$. Then, $J \neq \emptyset$.*

Proof. Suppose for indirect proof that $J = \emptyset$. Then, we have a perpendicular Jacobi vector field $Y(t)$, $t \in (-\infty, \infty)$, along γ such that there exist i_0 and $j_0 \geq i_0 + 2$ with $Y(i_0) = 0$ and $Y(t) \neq 0$ for all $t \in (t_{j_0}, t_{j_0+1}]$ where $\gamma(t_j) \in C$. Let $\gamma_u : (-\infty, \infty) \longrightarrow B$ be the variation through billiard ball trajectories such that $\gamma_0 = \gamma$, $\gamma_u(t_{j_0}(u)) \in C$, $\gamma_u(t_0(u)) = c(u)$, $\gamma_u(t_{i_0}(u)) = \gamma(t_{i_0})$ and its variation vector field is Y. Then, $\gamma_u((t_{j_0}(u), t_{j_0+1}(u)])$ do not cross to one another for sufficiently small $u \geq 0$. Let $\theta(u)$ be the angle between $\dot{c}(u)$ and the oriented segment $T(\gamma_u(t_{j_0}(u)), \gamma_u(t_{j_0+1}(u)))$ and let $\theta_1(u)$ be the angle between $\dot{c}(u)$ and the oriented segment $T(\gamma_u(t_{j_0}(u)), \gamma(t_{j_0+1}))$. Since the neighborhood of $T(\gamma(t_{j_0}), \gamma(t_{j_0+1}))$ is foliated by segments $T(\gamma_u(t_{j_0}(u)),$

$\gamma_u(t_{j_0+1}(u)))$, we see that $\theta(u) > \theta_1(u)$ for any $u < 0$ and $\theta(u) < \theta_1(u)$ for any $u > 0$. Hence, there exists a u_0 such that

$$\sum_{i=i_0}^{j_0} H(\gamma(t_j), \gamma(t_{j+1})) > \sum_{i=i_0}^{j_0-1} H(\gamma_{u_0}(t_j(u_0)), \gamma_{u_0}(t_{j+1}(u_0))) + H(\gamma_{u_0}(t_{j_0}(u_0)), \gamma(t_{j_0+1})),$$

contradicting the straightness of γ. □

Assume that $J \neq \emptyset$ and $Y \in J$ such that $\{a_j\}_{j\in\mathbf{Z}}$ is the set of all parameters for its conjugate points with $t_j < a_j < t_{j+1}$ for any $j \in \mathbf{Z}$. Let Y_m be a perpendicular Jacobi vector field along γ such that $Y_m(t_0 + 0)$ is constant for $m \in \mathbf{Z}$ with $Y_m(t_0 + 0) \perp \gamma'(t_0 + 0)$, $\|Y_m(t_0 + 0)\| = 1$ and $Y_m(t_m) = 0$. Let $S_m = \{b(m)_j \mid Y_m(b(m)_j) = 0, t_j < b(m)_j < t_{j+1}\}$.

To continue the discussion we need a lemma concerning the distribution problem of conjugate points (see [16] and [18]).

Lemma 6.4. (Separation property) *Suppose $\gamma(b)$ is the first conjugate point to $\gamma(a)$ with $a < b$. Any nontrivial perpendicular Jacobi vector field Y along γ with $Y(a) \neq 0$ or $Y(b) \neq 0$ has a unique zero point $\gamma(t_0)$ at $t_0 \in (a, b)$.*

Proof. Let $e(t)$, $t \in \mathbf{R}$, be a vector field along γ such that $\langle \dot{\gamma}(t), e(t)\rangle = 0$ and $\|e(t)\| = 1$ for each interval $[t_j, t_{j+1}]$ and $e(t_j + 0) = Q(e(t_j - 0))$. Any perpendicular Jacobi vector field Y along γ is written in such a way that $Y(t) = y(t)e(t)$ for any $t \in \mathbf{R}$. Then, $y(t)$ is continuous for $t \in \mathbf{R}$ and it satisfies

$$y'(t_j + 0) = y'(t_j - 0) - \frac{2\kappa(t_j)}{\sin\theta_j} y(t_j)$$

for all $j \in \mathbf{Z}$. Thus, if Y and Z are perpendicular Jacobi vector fields along γ, then $y'(t)z(t) - y(t)z'(t)$ is constant for all $t \in \mathbf{R}$. By the assumption there exists a nontrivial perpendicular Jacobi vector field Y along γ such that $y(a) = y(b) = 0$, $y'(a) = 1$ and $y(t) > 0$ for any $t \in (a, b)$. Since $\gamma(b)$ is the first conjugate point to $\gamma(a)$, we have $y'(b) < 0$. Let Z be a nontrivial perpendicular Jacobi vector field along γ such that $Z(t) = z(t)e(t)$ and $Z(a) \neq 0$, say $z(a) > 0$. Since $y'(b)z(b) = z(a)$, we have $z(b) < 0$. Therefore, there exists a $t_0 \in (a, b)$ such that $z(t_0) = 0$, proving the existence of zeros.

Suppose there exists another zero point of z. Let $\gamma(t_1)$ is the first conjugate point $\gamma(t_0)$ with $t_0 < t_1 \leq b$. Since $z'(t_0)z'(t_1) < 0$ and $y(t_0)z'(t_0) =$

$y(t_1)z'(t_1)$, we have a zero point of Y between t_0 and t_1, a contradiction. This proves the uniqueness of existence of zeros. □

The separation property of conjugate points shows that $t_j < a_j < b(m)_j < t_{j+1}$ for all $j \leq m-2$. The sequence $\{b(m)_0\}_{m>2}$ is monotone decreasing and bounded. Therefore, the limit $Y_f = \lim_{m\to\infty} Y_m$ exists and $Y_f \in J$. In the same manner the limit $Y_b = \lim_{m\to-\infty} Y_m$ exists and $Y_b \in J$. Let $\{\underline{b}_j\}_{j\in\mathbf{Z}}$ and $\{\bar{b}_j\}_{j\in\mathbf{Z}}$ be the sequence of parameters such that $Y_b(\underline{b}_j) = 0$, $Y_f(\bar{b}_j) = 0$ and $t_j < \underline{b}_j \leq \bar{b}_j < t_{j+1}$ for any $j \in \mathbf{Z}$. We notice the following.

Lemma 6.5. *$\underline{b}_j$ and $\bar{b}_j$ continuously depend on the billiard ball trajectories γ.*

Lemma 6.6. *Let $J \neq \emptyset$. Then, J is a one-dimensional vector space if and only if $Y_f = Y_b$.*

The following is a condition that $Y_f = Y_b$.

Lemma 6.7. *Let $x = (x_j)_{j\in\mathbf{Z}}$ be a billiard ball trajectory in C which corresponds to γ in $\mathbf{E}$ and a b-straight line $s = (s_j)_{j\in\mathbf{Z}}$ in $\mathbf{X}$. Suppose there exists a variation through b-parallels in $\mathbf{X}$ for $x(u) = (x(u)_j)_{j\in\mathbf{Z}}$ to a b-straight line in $\mathbf{X}$ for $x = (x_j)_{j\in\mathbf{Z}}$ such that $x(0) = x$. Then, J is a one-dimensional vector space where J is defined as above.*

Proof. Let $\gamma_u : (-\infty, \infty) \longrightarrow B$ be billiard ball trajectories corresponding to $x(u)$ with $x(u)_0 = \gamma_u(0)$, $t_0 = 0$. Let $0 < \epsilon < \min\{\underline{b}_0, \bar{b}_0 - \underline{b}_0\}$ and S the $\epsilon/2$-ball around $\gamma(\epsilon)$. Then S is foliated by γ_u. We define a function $F_{\pm s}$ on S as

$$F_{\pm s}(\gamma_u(t)) = B_{\pm s}(0, s_0(u)) \pm t$$

where $s(u) = (s_j(u))_{j\in\mathbf{Z}}$ is b-parallels in $\mathbf{X}$ corresponding to $x(u)$. Suppose $Y_f(\epsilon) = Y_b(\epsilon)$, $Y'_f(\epsilon) \neq Y'_b(\epsilon)$. If $Y_f(t) = y_f(t)e$, $Y_b(t) = y_b(t)e$ where e is the unit vector perpendicular to $\dot{\gamma}_0(\epsilon)$, then $y_b'(\epsilon) < y_f'(\epsilon) < 0$. Roughly speaking, $y_b'(\epsilon)$ and $y_f'(\epsilon)$ are the geodesic curvature of ${F_{-s}}^{-1}(\epsilon)$ and ${F_s}^{-1}(\epsilon)$, respectively. Hence, it is impossible that $F_s + F_{-s}$ is constant on S (see Proposition 5.12). □

We will show a differentiability condition of invariant circles of the billiard ball map φ.

Lemma 6.8. *Let* $\Gamma(s) = (x(s)_0, u(s)_0)$, $(x(s)_0 = c(s))$, *be a closed simple curve not null-homotopic in* Ω *which is invariant under* φ *and let* $x(s) = (x(s)_j)_{j \in \mathbf{Z}}$ *be the sequence of billiard ball trajectories such that* $\varphi(x(s)_0, u(s)_0) = (x(s)_1, u(s)_1)$ *for any* $s \in \mathbf{R}$. *Suppose the family of b-straight lines for* $x(s)$ *are b-parallels to each other. Then,* $x(s)_1$ *is of class* C^1 *for* $s \in \mathbf{R}$.

Proof. Let Y^s be the unique nontrivial perpendicular Jacobi vector field along γ_s whose cojugate points separate the boundary. Then, Y^s is continuous for $s \in \mathbf{R}$ because of Lemma 6.6 and Lemma 6.7. Let $W_s(t_1 + 0)$ be a tangent vector to C such that $W_s(t_1 + 0) = Y^s(t_1 + 0) + r(s)\dot{\gamma}_s(t_1 + 0)$ for some $r(s) \in \mathbf{R}$. Then, W_s is continuous for $s \in \mathbf{R}$ also. Since

$$x(s)_1 = \int_0^s W_s(t_1 + 0)\, ds + x(0)_1,$$

we see that $x(s)_1$ is of class C^1 for $s \in \mathbf{R}$. □

The following is the remark of a role of diameter.

Lemma 6.9. *Let* $\psi : (-\infty, \infty) \longrightarrow \Omega$ *be a closed simple curve not null-homotopic in* Ω *which is invariant under* φ. *Let* $g(s) = \pi(\varphi(\psi(s)))$ *for* $s \in \mathbf{R}$ *where* $\pi : \Omega \longrightarrow C$ *is the natural projection. Let* $\theta(s)$ *be the angle between* $T(c(s), g(s))$ *and* $\dot{c}(s)$ *for any* $s \in \mathbf{R}$. *Suppose* $T(c(s), g(s))$ *is not a diameter of* C *for any* $s \in \mathbf{R}$. *Then, either* $\theta(s)$ *is always greater than* $\pi/2$ *or less that* $\pi/2$.

Proof. Notice that all billiard ball trajectories determined by $T(c(s), g(s))$ are b-straight lines in the configuration space (see Birkhoff's theorem mentioned preceding Lemma 4.15). Suppose there exists an s_0 such that the angle between $T(c(s_0), g(s_0))$ and $\dot{c}(s_0)$ is equal to $\pi/2$. Then, the b-curve for the billiard ball trajectory determined by $T(c(s_0), g(s_0))$ is a periodic b-straight line with period (2,1) in the configuration space because of Lemma 4.13, and, hence, $T(c(s_0), g(s_0))$ is a diameter of C, a contradiction. Since $\theta(s)$ is continuous for $s \in \mathbf{R}$, the lemma is proved. □

The following theorem is a condition that a coustic is a continuous curve.

Theorem 6.10. *Let* $\psi : (-\infty, \infty) \longrightarrow \Omega$ *be a closed simple curve not null-homotopic in* Ω *which is invariant under* φ. *Let* $g(s) = \pi(\varphi(\psi(s)))$ *for any* $s \in \mathbf{R}$ *where* $\pi : \Omega \longrightarrow C$ *is the natural projection. Suppose the b-straight*

lines for billiard ball trajectories determined by $T(c(s), g(s))$ are b-parallels to each other. Then, there exists a closed continuous curve K in the domain B bounded by C such that any segment $T(c(s), g(s))$ is a tangent line to K for any $s \in \mathbf{R}$.

Proof. Since the b-straight lines for billiard ball trajectories determined by $T(c(s), g(s))$ are b-parallels to each other for all $s \in \mathbf{R}$, $g(s)$ is of class C^1 because of Lemma 6.8. Lemma 6.2 shows that there exists a closed continuous curve K in the domain bounded by C such that any segment $T(c(s), g(s))$ is a tangent line to K. □

7 Proof of Theorem 1.1

Let Σ be the space of all b-curves in $\mathbf{X}$ with natural topology. For any $s = (s_j)_{j \in \mathbf{Z}} \in \Sigma$ and any $a \in \mathbf{R}$ let $s + a = (s_j + a)_{j \in \mathbf{Z}}$. We define a equivalence relation $s \sim s'$ in Σ as $s' = s + pL$ for aome $p \in \mathbf{Z}$. Then, there exists the natural homeomorphism from Ω to $\Sigma/\sim$. Let $\sigma : \Sigma \longrightarrow \Sigma/\sim \approx \Omega$ be the natural projection.

Let $k : [0, a] \longrightarrow \Omega$ be a curve with $k(0) = \bar{x} \in \Omega$ and $s = (s_j)_{j \in \mathbf{Z}}$ a b-curve in $\mathbf{X}$ for $\bar{x}$. Then, there exists a unique curve $\ell : [0, a] \longrightarrow \Sigma$ such that $\sigma(\ell(t)) = k(t)$ for any $t \in [0, a]$ and $\ell(0) = s$. We say that the curve ℓ is the *lift of a curve* k into Σ.

Lemma 7.1. *Let $\bar{x} \in \Omega$ be such that the corresponding b-curve s in $\mathbf{X}$ is a b-straight line with its slope $\alpha = aL$ where a is an irrational number with $0 < a < 1$. Let f be a closed simple curve in Ω containing $\bar{x}$ (not a point). Suppose there exists an $i \in \mathbf{Z}$ such that f, $\varphi(f)$, $\cdots$, $\varphi^{i-1}(f)$ are mutually disjoint and $\varphi^i(f) = f$. Then, f is a closed simple curve not null-homotopic in Ω and invariant under φ.*

Proof. Let $\Gamma = \{U(q,p)s \,|\, q, p \in \mathbf{Z}\}$ and $\bar{\Gamma}$ the closure of Γ in $\mathbf{X}$. Let $A = \{u_0 \,|\, u \in \bar{\Gamma}\}$ and $M = \cup_{j=0}^{i-1} \varphi^j(f)$. Then, the number of connected components of M is finite. All elements $\varphi^j(\bar{x})$ contained in M. If $\varphi^j(\bar{x})$ is contained in a connected component f_1 of M, then $\varphi^k(\bar{x})$ which is sufficiently close to $\varphi^j(\bar{x})$ is also contained in f_1. Let $\mu : A \longrightarrow \bar{\Gamma}$ be a map given by sending u_0 to the b-straight line u through u_0. The map μ is well-defined and a homeomorphism because of Theorem 3.13 in [1] (see Theorem 4.12). Let $\mathbf{R} - A = \cup_{\lambda \in \Lambda} O^\lambda$ where O^λ is an open interval which is a connected component of $\mathbf{R} - A$ for each $\lambda \in \Lambda$ and Λ is a countable set.

The points u_0 in A are devided into two types.

1. There exists a sequence of points $(u^n)_0$ in A which is monotone decreasing and conveges to u_0.

2. u_0 is the lower bound of O^λ for some $\lambda \in \Lambda$.

Let g be the lift of f into Σ containing s. In order to find a curve in Σ which connects s and $s+L$ we have only to prove that $s+L \in g$ because $s \in g$. Let $S = \{u_0 \,|\, u = (u_j)_{j\in\mathbf{Z}} \in \bar{\Gamma} \cap g\} \subset A$. Then, S contains s. Set $u_0 = \max S \cap [s_0, s_0+L]$. For indirect proof we suppose $u_0 < s_0 + L$.

Suppose u_0 is of type (1). Let $(u^n)_0 \in A$ be a sequence which is monotone decreasing and converges to u_0. Then, $\mu((u^n)_0)$ tends to $\mu(u_0)$. Hence, $\sigma(\mu(u_0))$ and $\sigma(\mu((u^n)_0))$ are contained in the same connected component of M for sufficiently large n. This contradicts that $u_0 = \max S \cap [s_0, s_0+L]$ because $\mu(u_0)$ is connected to s by a subarc of g and $\mu((u^n)_0)$ is connected to $\mu(u_0)$ by a subarc of g.

Suppose u_0 is of type (2). Namely, u_0 is the lower bound of O^λ for some $\lambda \in \Lambda$, say $O^\lambda = (u_0, v_0)$. Since $\mu(u_0) = u = (u_j)_{j\in\mathbf{Z}} \in \bar{\Gamma}$ and $\mu(v_0) = v = (v_j)_{j\in\mathbf{Z}} \in \bar{\Gamma}$ are not periodic, it follows that $u_j - u_k \neq 0 \pmod{L}$ and $v_j - v_k \neq 0 \pmod{L}$ for any $j \neq k$. Therefore, $\sum_{j=-\infty}^{\infty} v_j - u_j < L$. Hence, $v_j - u_j \longrightarrow 0$ as $j \longrightarrow \infty$. Since $v_j = (U(-j,0)v)_0$, $u_j = (U(-j,0)u)_0$, $\sigma(U(-j,0)v) = \varphi^j(\sigma(v))$ and $\sigma(U(-j,0)u) = \varphi^j(\sigma(u))$, it holds that $\mathrm{d}_\Omega(\varphi^j(\sigma(v)), \varphi^j(\sigma(u))) \longrightarrow 0$ as $j \longrightarrow \infty$. Thus, we see that $\varphi^j(\sigma(v))$ and $\varphi^j(\sigma(u))$ are contained in the same connected component f_2 of M, and, hence, $\sigma(v) \in \varphi^{-j}(f_2)$, $\sigma(u) \in \varphi^{-j}(f_2)$ for suffciently large j. This implies that $v_0 \in S \cap [s_0, s_0+L]$, a contradiction.

Since f is a closed simple curve not null-homotopic in Ω and φ preserves the natural measure of Ω (see [17]), it is impossible that $f \cap \varphi(f) = \emptyset$. Therefore, $\varphi(f) = f$ by assumption. □

Proof of Theorem 1.1 : Let $\alpha = aL$ where a is an irrational number with $0 < a < 1$. Let s^α be a b-straight line in $\mathbf{X}$ with slope α. By assumption and Lemma 7.1 we can get a φ-invariant closed simple curve f_α containing $\sigma(s^\alpha)$ which is not null-homotopic in Ω. Since F is closed in the set of all closed subsets in Ω, all $f \in F$ are closed simple curves not null-homotopic in Ω. Theorem 1.1 follows from this and Bialy's theorem.

8 Convex billiards having poles

We begin with the following lemma.

Lemma 8.1. *Suppose there exists a pole $x \in C$. Then, for any (q,p), $q,p \in \mathbf{Z}^+$, $p/q < 1$, there exists a periodic b-straight line $s = (s_j)_{j\in\mathbf{Z}}$ with period (q,p) such that the strip $[T(s), T(\bar{s})]$ is foliated by b-straight lines where $\bar{s} = (\bar{s}_j)_{j\in\mathbf{Z}}$, $\bar{s}_j = s_j + L$ for any $j \in \mathbf{Z}$.*

Proof. Let $t = (t_j)_{j\in\mathbf{Z}}$ be a periodic b-geodesic with period (q,p) such that $H(t; i, i+q) = \min D(q,p)$. Let $\bar{t} = (\bar{t}_j)_{j\in\mathbf{Z}}$, $\bar{t}_j = t_j + L$. Let $r_m \in [T(t), T(\bar{t})]$ be a sequence of poles such that $r_m \longrightarrow \infty$ as $m \longrightarrow \infty$. Since any point (i,r) with $t_i \le r \le \bar{t}_i$ can be connected by the unique b-curve in the strip $[T(t), T(\bar{t})]$ to r_m for sufficiently large m, we have a family of b-straight lines $r = (r_j)_{j\in\mathbf{Z}}$ passing through the points (i,r) with $t_i \le r \le \bar{t}_i$. It remains to prove that the family of b-straight lines $r = (r_j)_{j\in\mathbf{Z}}$ is a foliation of the strip $[T(t), T(\bar{t})]$. Define a map $\Psi_i : [t_i, \bar{t}_i] \longrightarrow [t_{i+q}, \bar{t}_{i+q}]$ as $\Psi_i(r_i) = r_{i+q}$. From Proposition 5.15 we see that $\Psi_i(t_i) = t_{i+q}$ and $\Psi_i(\bar{t}_i) = \bar{t}_{i+q}$. This fact and Proposition 5.12 show that $\Psi_i([t_i, \bar{t}_i]) = [t_{i+q}, \bar{t}_{i+q}]$ and all b-straight lines r are contained in the strip $[T(t), T(\bar{t})]$. □

The following shows a distinguished role of poles.

Theorem 8.2. *Suppose there is a pole $x \in C$. Then, for any (q,p), $q,p \in \mathbf{Z}^+$, $p/q < 1$, there passes a minimal periodic billiard ball trajectory through x with period (q,p).*

Proof. Let $s = (s_j)_{j\in\mathbf{Z}}$ be a periodic b-straight line with period (q,p) and $\bar{s} = (\bar{s}_j)_{j\in\mathbf{Z}}$ with $\bar{s}_j = s_j + L$. If s_0 is a pole, we have nothing to prove. Suppose u_0 with $s_0 < u_0 < s_0 + L = \bar{s}_0$ is a pole. Let $u = (u_j)_{j\in\mathbf{Z}}$ be a b-geodesic with $u_q = u_0 + pL$, namely it is the extension of the b-segment connecting u_0 and u_q. We first prove that $(u_j)_{j\le 0}$ or $(u_j)_{j\ge 0}$ stays in the strip $[T(s), T(\bar{s})]$. Suppose both $(u_j)_{j\le 0}$ and $(u_j)_{j\ge 0}$ do not lie in the strip $[T(s), T(\bar{s})]$. Suppose $(u_j)_{j\ge 0}$ cross $s = (s_j)_{j\ge 0}$. Then, $(u_j)_{j\le 0}$ cannot cross $(\bar{s}_j)_{j\le 0}$. In fact, set $\bar{u} = (\bar{u}_j)_{j\in\mathbf{Z}}$, $\bar{u}_j = u_{j-q} + pL$. If $u = (u_j)_{j\in\mathbf{Z}}$ crosses $\bar{s}$ at j or between j and $j+1$, then $\bar{u}$ crosses $\bar{s}$ at $j+q$ or between $j+q$ and $j+q+1$. Similarly, if $u = (u_j)_{j\in\mathbf{Z}}$ crosses s at k or between k and $k+1$, then $\bar{u}$ crosses s at $k+q$ or between $k+q$ and $k+q+1$. Since $\bar{u}_q = u_0 + pL = u_q$, we see that $\bar{u}_j > u_j$ for $j < q$ and $\bar{u}_j > u_j$ for $j > q$. However, this is impossible because u and $\bar{u}$ are b-geodesics. Hence, the only possible case is

that $(u_j)_{j\le 0}$ crosses $(s_j)_{j\le 0}$ also. However, we see a contradiction as follows. Let $W=\{v\}$ be a foliation of the strip $[T(s),T(\bar{s})]$ by b-straight lines v and let $P_W : \cup_{j=-\infty}^{\infty}\{j\}\times[s_j,\bar{s}_j] \longrightarrow \{0\}\times[s_0,\bar{s}_0]$ be a projection along the foliation W. Namely, $P_W(k,w)$ is given as follows. Take $v=(v_j)_{j\in\mathbf{Z}}\in W$ such that $v_k=w$ and $P_W(k,w)=v_0$. Since $u=(u_j)_{j\in\mathbf{Z}}$ crosses s twice when it goes out the strip $[T(s),T(\bar{s})]$, the set $\{P_W(j,u_j)|\ u_j$ lies in the strip $[T(s),T(\bar{s})]\}$ has its maximum v_0. This contradicts that both $v=(v_j)_{j\in\mathbf{Z}}\in W$ passing through v_0 and u are b-geodesics (see Lemma 2.5).

Suppose $(u_j)_{j\ge 0}$ stays in the strip $[T(s),T(s')]$. We will prove that $u=(u_j)_{j\in\mathbf{Z}}$ is a periodic b-ray with period (q,p). Let $u^1=U(q,p)u=({u^1}_j)_{j\ge q}$ where ${u^1}_j=u_{j-q}+pL$. Suppose $u^1=U(q,p)u\neq u$, say ${u^1}_{2q}>u_{2q}$. Set $v=\cup_{k=0}^{\infty}\left((U(q,p)^k u)_j\right)_{kq\le j\le (k+1)q}$. For any $m\in\mathbf{Z}^+$ set $u^k=U(q,p)^k u$ and $(U(q,p)^k u)_{mq}=:{u^k}_{mq}$ for $k=0,\cdots,m-1$. Then, $u_{mq}<(U(q,p)u)_{mq}<\cdots<(U(q,p)^{m-1}u)_{mq}$. By Lemma 3.3 and Proposition 3.4,

$$\begin{aligned}|u_0-s_0|+|u_{mq}-s_{mq}|+H(s;0,mq)&\ge H(0,mq;u_0,u_{mq})\\&=H(0,q;u_0,u_q)+H(q,mq;u_q,u_{mq})\end{aligned}$$

$$\begin{aligned}|u_{mq}-{u^1}_{mq}|+H(u;q,mq)&\ge H(q,mq;{u^1}_q,{u^1}_{mq})\\&=H(q,2q;{u^1}_q,{u^1}_{2q})+H(2q,mq;{u^1}_{2q},{u^1}_{mq})\end{aligned}$$

$$\begin{aligned}|{u^1}_{mq}-{u^2}_{mq}|+H(u^1;2q,mq)&\ge H(u^2;2q,mq)\\&=H(u^2;2q,3q)+H(u^2;3q,mq)\end{aligned}$$

$$\begin{aligned}|{u^{m-2}}_{mq}-{u^{m-1}}_{mq}|+H(u^{m-2};(m-1)q,mq)&\\\ge H(u^{m-1};(m-1)q,mq).&\end{aligned}$$

Therefore, we have

$$\begin{aligned}|u_0-s_0|+|u_{mq}-s_{mq}|+|u_{mq}-{u^1}_{mq}|+\cdots&\\+|{u^{m-2}}_{mq}-{u^{m-1}}_{mq}|+H(s;0,mq)&\\\ge H(u;0,q)+H(u^1;q,2q)+\cdots+H(u^{m-1};(m-1)q,mq).&\end{aligned}$$

and, hence,

$$|u_0-s_0|+|{u^{m-1}}_{mq}-s_{mq}|+mH(s;0,q)\ge mH(u;0,q)$$

Dividing both sides by m and taking m to the infinity, we have

$$\min D(q,p)=H(s;0,q)\ge H(u;0,q)$$

From this we see that $H(u;0,q) = \min D(q,p)$, and, hence, u is a minimal periodic b-geodesic with period (q,p).

□

Lemma 8.3. *Suppose there exists a pole $x \in C$. Then, for any (q,p), $q,p \in \mathbf{Z}^+$, $p/q < 1$, and any s_0 corresponding to a pole, there passes a minimal periodic b-geodesic $s = (s_j)_{j\in\mathbf{Z}}$ with period (q,p) such that the strip $[T(s), T(\bar{s})]$ is foliated by b-straight lines and the foliation W corresponds to a not null-homotopic φ-invariant closed curve in the phase space Ω, where $\bar{s}_j = s_j + L$ for all $j \in \mathbf{Z}$.*

Proof. As was seen in Theorem 8.2, there exists a minimal periodic b-geodesic $s = (s_j)_{j\in\mathbf{Z}}$ containing a pole s_0. Let $\bar{s} = (s_j)_{j\in\mathbf{Z}}$, $\bar{s}_j = s_j + L$. We have a foliation W of the strip $[T(s), T(\bar{s})]$ as the limit of b-segments emanating from poles $\bar{s}_{mq} = s_{mq} + L$, and, therefore, any element of W is a co-b-ray of $\bar{s}$. We have to prove that it corresponds to a not null-homotopic φ-invariant closed curve in Ω. We define a map $\beta : [s_0, \bar{s}_0] \longrightarrow [s_1, \bar{s}_1]$ as $\beta(u_0) = u_1$ where $u = (u_j)_{j\in\mathbf{Z}} \in W$. We will prove that $\beta(u_1) = u_2$ if $u_1 \in [s_0, \bar{s}_0]$ and $\beta(u_1 - L) = u_2 - L$ if $u_1 \in [s_1, \bar{s}_1] - [s_0, \bar{s}_0]$. Suppose $u_1 \in [s_0, \bar{s}_0]$. Let $v = (v_j)_{j\in\mathbf{Z}}$ such that $v_j = \bar{s}_{j+1}$ for all $j \in \mathbf{Z}$, namely, $v = U(-1,0)\bar{s}$.. Then, v and $\bar{s}$ are co-b-rays to each other. Let $\bar{v} = (\bar{v}_j)_{j\geq 0}$ be the co-b-ray to v from $(0, u_1)$. Then, $\bar{v}$ is a co-b-ray to $\bar{s}$, since $\bar{s}$ is the unique co-b-ray to v passing through $\bar{s}_0$. This implies that $\beta(\bar{v}_0) = \bar{v}_1$. Since the pair $(0, u_1)$ and v were translated from $(1, u_1)$ and $\bar{s}$ by $U(-1,0)$, we see that $\bar{v}$ is translated to the co-b-ray to $\bar{s}$ from $(1, u_1)$ by $U(1,0)$. Therefore, $\beta(u_1) = \beta(\bar{v}_0) = \bar{v}_1 = u_2$.

Suppose $u_1 \in [s_1, \bar{s}_1] - [s_0, \bar{s}_0]$. Then, by using $U(-1,-1)$ instead of $U(-1,0)$, we can prove the remaining part. □

9 Proof of Corollaries 1.2 to 1.5

Proof of Corollaries 1.2, and 1.5 : By Theorem 4.5 and Lemma 4.6 we see that the assumptions of Corollary 1.2 and 1.5 are equivalent. Let $W_{(q,p)}$ be a foliation of $\mathbf{X}$ by minimal periodic b-geodesics with period (q,p). Then, $W_{(q,p)}$ corresponds to a simple closed curve not null-homotopic in Ω. Those φ-invariant closed curves yield a foliation F of Ω by simple curves. Corollaries follow from Theorem 1.1.

Proof of Corollary 1.3 : Theorem 8.2 and Lemma 4.6 state that $D(q,p)$ is constant in $\mathbf{X}$ for all $q,p \in \mathbf{Z}^+$. Corollary 1.3 follows from Corollary 1.5.

Proof of Corollary 1.4 : Lemma 8.3 shows that for any (q,p) with $p/q < 1$ there is a foliation W of $\mathbf{X}$ by b-straight lines such that W contains b-straight lines with period (q,p). Since those b-straight lines are given as the limit of b-rays emanating from divergent poles, the monotone property implies that all b-straight lines in W are with period (q,p). The assumption of Corollary 1.5 is satisfied.

10 Divergence property

In this section we prove Corollary 1.6.

Proof of Corollary 1.6: Let $q,p \in \mathbf{Z}^+$ with g.c.d$(q,p) = 1$, $p/q < 1$. We prove that the configuration space $\mathbf{X}$ is foliated by minimal periodic b-geodesics with period (q,p), which is equivalent to that all displacement functions $D(q,p)$ are constant in $\mathbf{X}$. Then, Corollary 1.5 proves Corollary 1.6.

Suppose $D(q,p)$ is not constant in $\mathbf{X}$. Theorem 9.3.7 in [17] states that there exist at least two b-curves with period (q,p) one of which is not minimal if $D(q,p)$ is not constant in $\mathbf{X}$. Let $s = (s_j)_{j\in\mathbf{Z}}$ be a minimal periodic b-geodesic with period (q,p) and let $t = (t_j)_{j\in\mathbf{Z}}$ be not a minimal b-curve with period (q,p) such that $|t_0 - s_0| \leq L/2q$. Because of how to make t we know that t does not cross s. Since t is not a b-ray, we have a co-b-ray $t' = (t'_j)_{j\geq 0} \neq t$ to s with $t'_0 = t_0$ which lies between t and s. Then, $\mathrm{dis}_\infty(t,t') \leq L/2q$, and, hence, by assumption, we have that $t = t'$, a contradiction. This implies that $D(q,p)$ is constant.

11 Proof of Theorem 1.7

Let f be a φ-invariant simple closed curve not null-homotopic in Ω. The Birkhoff theorem states that f is a graph of a Lipschiz continuous function $C \longrightarrow (-\infty,\infty)$. From this it follows that f yields a foliation of b-straight lines in $\mathbf{X}$ which is invariant under all translations. Therefore, the slope $\alpha(\bar{x})$ is constant for any $\bar{x} \in f$. Hence, we are allowed to write $\alpha(f)$. We begin with the following lemma.

Lemma 11.1. *Let f be a φ-invariant simple closed curve not null-homotopic in Ω with slope $\alpha(f) \neq L/2$. Suppose f makes a caustic K which lies in C. If K is simple, then K is a convex curve.*

Proof. Let $x = (x_j)_{j\in\mathbf{Z}}$ denote the billiard ball trajectory in $\mathbf{E}$ determined by $\bar{x} \in f$ and $\psi(x_0)$ the point in K where $T(x_0, x_1)$ is tangent to K. Since K is a simple curve, the map $\psi : C \longrightarrow K$ is such that $\psi^{-1}(y)$ is a connected set for any $y \in K$. In order to see the convexity of K it is sufficient to prove that ψ preserves the orientation because the slope of $T(x_0, x_1)$ is monotone increasing with respect to the usual xy-coordinate in $\mathbf{E}$. Suppose for indirect proof that $\psi(y)$ moves for $y \in C$ along the reverse orientation of K. Then, K is the union of locally concave curves. We assume without loss of generality that $\psi(y_1)$ lies in the left side of the line containing $T(y_0, y_1)$ at $y_0 \in C$. Let $z(x_0) \in K$ be a point through which $T(x_0, x_1)$ goes across K from the inside to the outside. As x_0 moves along C until x_1 arrives at x_2, the point $z(x_0)$ moves and passes through $\psi(y_1)$ in K. Since $\psi(x_1)$ moves from $\psi(y_1)$ along the reverse orientation of C, there exists an x_0 such that $\psi(x_1) = z(x_0)$. Then, the segment $T(x_2, x_1)$ coinsides with $T(x_0, x_1)$. Hence, $T(x_1, x_2)$ is perpendicular to C at x_1, and, therefore, $\alpha(f) = L/2$ because of Lemma 4.14, 4.13, and 4.15. It contradicts the assumption $\alpha(f) \neq L/2$. □

We will prove Theorem 1.7 and Corollary 1.8.

Proof of Theorem 1.7: Let $s^n = (s^n{}_j)_{j\in\mathbf{Z}}$ be the configuration for $\bar{x} \in f_n$ such that $c(s^n{}_0)$ is an endpoint of the diameter of C. Each caustic $K(f_n)$ lies in the sector made of segments $T(c(s^n{}_{-1}), c(s^n{}_0))$ and $T(c(s^n{}_0), c(s^n{}_1))$, since $K(f_n)$ is a simple closed convex curve (see Lemma 11.1). Let $t = (t_j)_{j\in\mathbf{Z}}$ be the minimal periodic b-geodesic with period $(2, 1)$ such that $t_0 = s_0$, namely $T(c(t_0), c(t_1))$ is a diameter of C. Since $s^n{}_{-1} \longrightarrow t_{-1}$, $s^n{}_0 = t_0$, $s^n{}_1 \longrightarrow t_1$ as $n \longrightarrow \infty$, the sequence of caustics $K(f_n)$ converges to a segment $T = T(a, b)$ which lies in a diameter $T(c(t_0), c(t_1))$. Let $x = c(u_0)$ be any point in C such that $x \neq c(t_0)$ and $x \neq c(t_1)$. Let $u^n = (u^n{}_j)_{j\in\mathbf{Z}}$ be the sequence of b-straight lines such that $T(c(u^n{}_0), c(u^n{}_1))$ are tangent to $K(f_n)$ and $u^n{}_0 = u_0$. Then, the segments $T^n(x)_- = T(c(u^n{}_{-1}), c(u^n{}_0))$ and $T^n(x)_+ = T(c(u^n{}_0), c(u^n{}_1))$ from x are tangent to $K(f_n)$. Since $K(f_n)$ are simple convex closed curves, the sequence of segments $T^n(x)_-$ and $T^n(x)_+$ converges to segments $T(x)_-$ and $T(x)_+$ which pass through the endpoints a and b of T. The length of $T(a, x) \cup T(b, x)$ is constant in $x \in C$, since $T(x)_-$ and $T(x)_+$ lie in a billiard trajectory, so that the angles of $T(x)_+$ and $T(x)_-$ with the tangent line of C are equal. This states that C is an ellipse. This completes the proof of Theorem 1.7.

Proof of Corollary 1.8 : It is enough to note that the assumption is the same as in Theorem 1.7.

References

[1] V. Bangert. Mather sets for twist maps and geodesics on tori, In U. Kirchgraber, H. O. Walther (eds.) Dynamics Reported, Vol 1, pp.1–56, John Wiley & Sons and B. G. Teubner, 1988.

[2] V. Bangert. Geodesic rays, Busemann functions and monotone twist maps, Calc. Var., Vol 2, 49–63 (1994).

[3] M. Berger. *Geometry I, II*, Berlin Heidelberg New York London Paris Tokyo, Springer-Verlag, 1987.

[4] M. Bialy. Convex billiards and a theorem of E. Hopf, Math. Z., Vol 214, 147–154 (1993).

[5] G. D. Birkhoff. Surface translations and their dynamical applications, Acta Math., Vol 43, 1–119 (1922).

[6] H. Busemann. *The geometry of geodesics*, New York, Academic Press, 1955.

[7] L. Green. A theorem of E. Hopf, Michigan Math. J., Vol 5, 31–34 (1958).

[8] E. Gutkin and A. Katok. Caustics for inner and outer billiards, Commun. Math. Phys., Vol 173, 101–133 (1995).

[9] E. Hopf. Closed surfaces without conjugate points, Nat. Acad. Sci. U.S.A., Vol 34, 47–51 (1948).

[10] N. Innami. Differentiability of Busemann functions and total excess, Math. Z., Vol 47, 235–247 (1982).

[11] N. Innami. On tori having poles, Invent. math., Vol 84, 437–443 (1986).

[12] N. Innami. Families of geodesics which distinguish flat tori, Math. J. Okayama Univ., Vol 28, 207–217 (1986).

[13] N. Innami. A note on nonfocality properties in compact manifolds, Archiv der Math., Vol 48, 277–280 (1987).

[14] N. Innami. A characterization of flat metrics on tori by ergodicity, Ergod Th. Dynam. Sys., Vol 7, 197–201 (1987).

[15] N. Innami. Convex curves whose points are vertices of billiard triangles, Kodai Math. J., Vol 11, 17–24 (1988).

[16] N. Innami. Integral formulas for polyhedral and spherical billiards, J. Math. Soc. Japan, Vol 50, No 2, 339–357 (1998).

[17] A. Katok and B. Hasselblatt. *Introduction to the modern theory of dynamical systems*, New York, Cambridge University Press, 1995.

[18] W. Klingenberg. *A course in differential geometry*, Berlin, Springer-Verlag, 1978.

[19] H. Poritsky. The billiard ball problem on a table with a convex boundary – An illustrative dynamical problem, Ann. Math., Vol 51, No 2, 446–470 (1950).

[20] S. Tanno. C^∞-approximation of continuous ovals of constant width, J. Math. Soc. Japan, Vol 28, 384–395 (1976).

[21] M. P. Wojtkowski. Two applications of Jacobi fields to the billiard ball problems, J. Diff. Geom., Vol 40, 155–164 (1994).

[22] E. M. Zaustinsky. Extremals on compact E-surfaces, Trans. Amer. Soc., Vol 102, 433–445 (1962).

DEPARTMENT OF MATHEMATICS
FACULTY OF SCIENCE, NIIGATA UNIVERSITY
NIIGATA, 950-2181, JAPAN

Received December 10, 2001 Revised May 2, 2002