TSUKUBA J. MATH.
Vol. 33 No. 1 (2009), 147–160

CIRCULAR BILLIARDS AND PARALLEL AXIOM IN CONVEX BILLIARDS

By

Shinetsu Tamura and Nobuhiro Innami

Abstract. Circles will be characterized by some properties of billiard ball trajectories. The theory of parallels and the parallel axiom play important roles in the geometry of the configuration space. Those characterizations are concerned with Bialy's theorem which is a partial answer to Birkhoff's conjecture.

1. Introduction

Let C be a smooth simple closed and strictly convex curve with length L in the Euclidean plane $\mathbf{E}$ and let $c : \mathbf{R} \to \mathbf{E}$ be its representation by arclength, namely $|\dot{c}(t)| = 1$ for any $t \in \mathbf{R}$ where $\mathbf{R}$ is the set of all real numbers. The orientation of C is assumed to be anti-clockwise. Let $x = (x_j)_{j \in \mathbf{Z}}$ be a sequence of points in C where $\mathbf{Z}$ is the set of all integers. Let $T(x) = \bigcup_{j=-\infty}^{\infty} T(x_j, x_{j+1})$ where $T(x_j, x_{j+1})$ is the oriented segment from x_j to x_{j+1} for each $j \in \mathbf{Z}$. We say that x (and $T(x)$) 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 $i \in \mathbf{Z}$. The convex billiard has been investigated in its phase space and its configuration space.

We call $\Omega = C \times (-1, 1)$ the *phase space* which is the set of all pairs (x, u) for $x \in C$ and $u \in (-1, 1)$. 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 $u_1 = \cos\theta_1$. Define a *billiard ball map* $\varphi : \Omega \to \Omega$ as $\varphi(x_0, u_0) = (x_1, u_1)$. The billiard ball map is an example of a monotone twist map (see [10]). If $\bar{x} = (x_0, u_0) \in \Omega$ and $(x_j, u_j) = \varphi^j(\bar{x})$ 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.

The billiard is said to be *integrable* if a subset of full measure of the phase

Received September 22, 2008.

space Ω is foliated by closed curves invariant under the billiard ball map φ. The billiards in circles and ellipses are integrable. Birkhoff's conjecture is stated in [3] as follows. The only examples of integrable billiards are circular and elliptic billiards. Bialy ([3]) has given a partial answer to the conjecture, proving that C is a circle if Ω is foliated by φ-invariant continuous closed curves not null-homotopic in Ω. Wojtkowski ([11]) 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 [3] Bialy's theorem corresponds to a theorem of Hopf ([7]) concerning Riemannian metrics on tori without conjugate points. Innami ([8]) extended Bialy's theorem to the higher dimensional case and the nonpositive curvature case as Green ([5]) did.

A sequence of points $x = (x_j)_{j \in \mathbf{Z}}$ 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 all $j \in \mathbf{Z}$ 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 uniquely determined up to the difference pL $(p \in \mathbf{Z})$. The theory of parallels for billiard trajectories in the configuration space has been developed in [1], [2] and [9]. We define the *slope* $\alpha(x)$ of x as

$$\alpha(x) = \liminf_{n \to \infty} \frac{s_n}{n}$$

where $s = (s_j)_{j \in \mathbf{Z}}$ is a configuration for x. Let $\alpha(\bar{x})$ denote the slope of the billiard ball trajectory determined by $\bar{x}$ for $\bar{x} \in \Omega$. It is known that all points $\bar{x}$ which are in a φ-invariant closed curve f not null-homotopic in Ω have the same slopes ([1], [9]). We define the slope $\alpha(f)$ of any φ-invariant closed curve f not null-homotopic in Ω as $\alpha(f) = \alpha(\bar{x})$ for any $\bar{x}$.

In the present paper we prove the following theorem which improves Biary's theorem.

Theorem 1.1. *Let C be a strictly convex closed curve of class C^1 with length L and with constant width. Suppose there exists a sequence of φ-invariant closed curves f_n not null-homotopic in Ω whose slopes $\alpha_n = \alpha(f_n)$ converge to $L/2$. Then, C is a circle.*

Let f be a φ-invariant closed curve not null-homotopic in Ω and f^- the curve consisting of the points $\bar{x}^-$ which correspond to the reversed billiard ball trajectories to $\bar{x} \in f$. Then f^- is also a φ-invariant closed curve not null-homotopic in Ω with slope $\alpha(f^-) = L - \alpha(f)$ (see Section 5).

Corollary 1.2. *Let C be a strictly convex closed curve of class C^1. Suppose there exists a sequence of φ-invariant closed curves f_n not null-homotopic in Ω such that* $\lim_{n\to\infty} f_n = \lim_{n\to\infty} f_n^-$ *as $n \to \infty$. Then, C is a circle.*

The corollary shows that our theorem betters Biary's theorem.

Corollary 1.3. *Let C be a strictly convex closed curve of class C^1 with length L. Suppose the slope function α is continuous in Ω and $\alpha^{-1}(L/2)$ has no interior points. Then C is a circle.*

We define poles for convex billiards as follows. 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 the 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 $j < 0$. All points in circles are poles. The endpoints of long axis in an ellipse are poles and other points are not poles.

Corollary 1.4. *Let C be a strictly convex closed curve of class C^1 with constant width. Suppose there exists a pole in C. Then C is a circle.*

2. Preliminaries

Details of theorems introduced in this section can be seen in [9]. 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 $s = (s_j)_{i\le j\le k}$ makes a broken segment $T(s) = \bigcup_{j=i}^{k-1} T(s_j, s_{j+1})$ 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 $x = (x_j)_{i\le j\le k}$ be a sequence of points in C with $x_j \ne x_{j+1}$ for any j. We define a configuration $s = (s_j)_{i\le j\le k}$ for x as $x_j = c(s_j)$ and $s_j < s_{j+1} < s_j + L$ for $i \le j \le k-1$. 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

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

where $|\cdot|$ is the natural norm in $\mathbf{E}$ and $c : \mathbf{R} \to \mathbf{E}$ 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 $s = (s_j)_{i \le j \le k}$ from $s_i = (i,u)$ to $s_k = (k,v)$.

A C-curve $s = (s_j)_{i \le j \le k}$ (and $T = T(s)$) is called a *billiard curve* or simply a *b-curve* if $x = (x_j)_{i \le j \le k}$ given by $x_j = c(s_j)$ for $i \le j \le k$ is a billiard ball trajectory. The b-curves are the critical points of the function H in the set of all C-curves connecting given endpoints. 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 $H(s;j,j+2) = H(j,j+2;s_j,s_{j+2})$. 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 $H(s;i,k) = H(i,k;s_i,s_k)$. A b-geodesic $s = (s_j)_{j \ge i}$ (resp., $s = (s_j)_{j \le i}$) (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 $k > j \ge i$ (resp., $j < k \le i$). 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 $k > j$.

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

PROPOSITION 2.1. *If W is a foliation of the strip $[T(s), T(s')]$ by b-curves, then all b-curves $t = (t_j)_{i \le j \le k}$ 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 $t = (t_j)_{i \le j \le k} \in W$, then the sub-b-curve $t = (t_j)_{h \le j \le m}$ of $t = (t_j)_{i \le j \le k}$ is the unique b-curve connecting t_h and t_m in the strip $[T(s), T(s')]$.*

Let f be a φ-invariant closed curve which is not null-homotopic in Ω. Then, the set of all configurations for all points $\bar{x} \in f$ makes a foliation of $\mathbf{X}$ which is invariant under all translations. Proposition 2.1 implies that those configurations are b-straight lines in $\mathbf{X}$.

PROPOSITION 2.2. *Let $t = (t_j)_{h \le j \le m}$ and $u = (u_j)_{k \le j \le m}$ be b-segments with $t \ne u$. 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 if at least one of t_i and t_j is not an endpoint of the segment t.*

Let $q, p \in \mathbf{Z}$ with $0 < |p/q| < 1$. The *displacement function* $D = D(q, p) : \mathbf{X} \to \mathbf{R}$ is given by

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

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 $s_i \in \mathbf{X}$.

We say that a b-curve $s = (s_j)_{j \in \mathbf{Z}}$ is *with period* (q, p) if $s_{j+q} = s_j + pL$ for any $j \in \mathbf{Z}$. The periodic b-geodesic $s = (s_j)_{j \in \mathbf{Z}}$ is said to be *minimal* if $D(s_j) = \min\{D(s) \mid s \in \{j\} \times \mathbf{R}\}$.

Proposition 2.3. *Suppose $D(q, p)$ 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 whose slope is pL/q.*

The *diameter* d of C is by definition $d = \max\{|c(s) - c(t)| \mid s, t \in \mathbf{R}\}$. The diameter is characterized by a billiard ball trajectory as follows.

Lemma 2.4. *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 $j \in \mathbf{Z}$.*

The following proposition helps us improve the assumption in Bialy's theorem, combined with Theorem 4.1.

Proposition 2.5. *C is with constant width if and only if $\mathbf{X}$ is foliated by periodic b-straight lines with period $(2, 1)$.*

Let $s = (s_j)_{j \geq i_0}$ be a b-ray. We 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], [4], [9]). In the same way we define the Busemann function of a b-ray $s = (s_j)_{j \leq i_0}$ by using $n \to -\infty$ instead of "$n \to \infty$". 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.

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 $i \in \mathbf{Z}$.

LEMMA 2.6. *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.*

If $t^n = (t_j^n)_{i_1 \leq j \leq n}$ be a b-segment from $t_{i_1}^n$ to s_n and a sequence $t_{i_1}^n$ is bounded, then there exists a subsequence t^m which converges to a b-ray.

LEMMA 2.7. *Let $t^n = (t_j^n)_{i_1 \leq j \leq n}$ be a b-segment from $t_{i_1}^n$ to s_n. If a sequence t^n converges to a b-ray $t = (t_j)_{j \geq i_1}$, then 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 2.8. *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 $j \geq i_2$.*

PROPOSITION 2.9. *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.*

PROPOSITION 2.10. *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.*

LEMMA 2.11. *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-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 and the foliation W corresponds to a φ-invariant closed curve not null-homotopic in the phase space Ω, where $\bar{s}_j = s_j + L$ for all $j \in \mathbf{Z}$.*

3. Convex Parts of Caustics

Let $x, y \in C$. The orientation of C is assumed to be anti-clockwise. Let $T(x, y)$ be the oriented segment from x to y and $S(x, y)$ the oriented straight line through x and y. Let $H(x, y)$ be the closed half plane which is in the left side of $S(x, y)$ in Euclid plane. Let M_a be the set of all b-straight lines with slope aL. Suppose M_a is a foliation of $\mathbf{X}$ in this section. Let $(s_0)_1$ is a function defined on $[0, L]$ which is given by $(s_0)_1 = s_1$ where $s = (s_j)_{j \in \mathbf{Z}}$ is the unique b-straight line in M_a through s_0. Let

$$C_a(r) = \bigcap_{0 \le s_0 \le r} H(c(s_0), c((s_0)_1))$$

where $0 \le r \le L$. Obviously $C_a(r) \subset C_a(r')$ if $0 < r' < r < L$. We call $C_a = C_a(L)$ the *convex part of caustic with slope aL*.

Lemma 3.1. *Assume that M_a is a foliation of $\mathbf{X}$. Then, C_a is a nonempty convex set and all billiard ball trajectories x intersecting C_a are with slopes $\alpha(x)$ greater than or equal to aL if $\alpha(x) < L/2$.*

In the following proof it is important that the tangent line of C at $c(s_0)$ intersect $T(c(s_0), c((s_0)_1))$ with an angle less than $\pi/2$, and that $(s_0)_1$ is monotone and continuous in $s_0 \in [0, L)$ ([9]).

Proof. We will prove that $C_a(L)$ is not empty. Let $a_0 \in [0, L)$ be the number such that $(a_0)_1 \le L$. Then, $c((a_0)_1) \in C_a(a_0)$, since $(s_0)_1$ is monotone increasing in $s_0 \in [0, L)$. Let $y_0 \in T(c(0), c(0_1))$ be the nearest point from 0 in the set of all points $T(c(0), c(0_1)) \cap T(c(s_0), c((s_0)_1))$ for $0 < s_0 \le 0_1$. Let $b_0 = 0_{-1} + L$, namely, $(b_0)_1 = L$. The boundary $\partial C_a(b_0)$ of $C_a(b_0)$ is a convex curve which consists of $T(c(0), y_0)$, a convex curve K from y_0 to a point w_0 in $T(c(b_0), c(L))$ and $T(w_0, c(L))$. Let $u \in C$ be the point at which the oriented tangent line to K with right derivative at y_0 intersects C and let d_0 be the parameter such that $0_1 < d_0 < L$ and $u = c(d_0)$.

We have two cases; $d_0 \le b_0$ and $d_0 > b_0$. If $d_0 \le b_0$, then there exists the smallest parameter $b_1 > b_0$ such that $T(c(b_1), c((b_1)_1))$ passes through y_0. More precisely, $c((s_0)_1)$ is between $c(b_0)$ and $c(0) = c(L)$ for $s_0 \in [d_0, b_0]$ and $T(c(s_0), c((s_0)_1))$ intersects $T(c(0), c(0_1))$ at a point between $c(0)$ and y_0 for $s_0 \in [b_0, b_1]$. Hence, it follows that $y_0 \in C_a(b_1)$. If $d_0 > b_0$, then $L \le (s_0)_1 \le (d_0)_1$ for $s_0 \in [b_0, d_0]$, and, hence, y_0 is in the left side of $T(c(s_0), c((s_0)_1))$. Therefore, there exists a parameter b_1 with $b_1 > d_0 > b_0$ such that $T(c(b_1), c((b_1)_1))$ passes through y_0. It follows that $y_0 \in C_a(b_1)$. The convex curve $\partial C_a(b_1)$ consists of a convex curve from y_0 to a point w_1 in the segment $T(c(b_1), c((b_1)_1))$ and $T(w_1, y_0)$. Let $U(b_1)$ be the supporting line to $C_a(b_1)$ through $c(b_1)$ which is not the segment $T(c(b_1), c((b_1)_1))$ and let p_1 be a point $U(b_1) \cap C_a(b_1)$. Then, p_1 is in the left side of $S(c(b_1), c((b_1)_1))$. If the supporting line $U(b_1)$ does not intersect the segment $T(c(0), c(0_1))$, then p_1 is in the left side of $S(c(s_0), c((s_0)_1))$ for any $s_0 \in (b_1, L]$, and, hence, $p_1 \in C_a(L)$. If the supporting line $U(b_1)$ intersect $T(c(0), c(0_1))$ at a point y_1, then we can find b_2 with $b_1 < b_2 \leq L$ such that $T(c(s_0), c((s_0)_1))$ intersects $T(c(0), y_1)$ for any $s_0 \in [b_1, b_2]$. Thus, p_1 is in the left side of $S(c(s_0), c((s_0)_1))$ for any $s_0 \in [b_1, b_2]$, and, hence, $p_1 \in C_a(b_2)$. By using $c(b_2)$ and the supporting line $U(b_2)$ to $C_a(b_2)$ through $c(b_2)$ instead of $c(b_1)$ and $U(b_1)$, we find a point $p_2 \in C_a(b_2)$ such that $p_2 \in C_a(L)$ or there exist a parameter b_3 with $b_2 < b_3 \leq L$ and $y_2 \in T(c(0), c(0_1))$ such that p_2 is in the left side of $S(c(s_0), c(s_0)_1))$ for any $s_0 \in [b_2, b_3]$ and $T(c(s_0), c((s_0)_1))$ intersect $T(c(0), c(0_1))$ at some point between $c(0)$ and y_2 on $T(c(0), c(0_1))$, and, hence, $p_2 \in C_a(b_3)$. This is a process of making $b_1 < b_2 < \cdots < b_n < L$ and a sequence of points $p_1, p_2, \ldots, p_n$ in $C_a(b_n)$. Since $p_1, p_2, \ldots, p_n$ are in this order on $C_a(b_n)$, the sequence b_i is a finite sequence. Thus, we have $C_a = C_a(L)$ which is not empty.

By construction of C_a, we easily see that all billiard ball trajectories x intersecting C_a have slopes $\alpha(x)$ greater than or equal to aL if $\alpha(x) < L/2$. □

The following lemma is obvious from the proof of Lemma 3.1.

LEMMA 3.2. *Assume that M_a and $M_{a'}$ are foliations of* $\mathbf{X}$ *with $a < a' < L/2$. Then, $C_{a'} \subset C_a$.*

4. Parallel Axiom and Periodic Trajectory

Let M_a be the set of all points $\bar{x} \in \Omega$ whose configuration is a b-straight line in $\mathbf{X}$ with slope $\alpha(\bar{x}) = aL$ where $0 < a < 1$. We also denoted the set of those b-straight lines in $\mathbf{X}$ as M_a for convenience. We say that M_a satisfies the *parallel axiom* if given two b-straight lines in M_a are b-parallel to each other.

THEOREM 4.1. *Let $a = p/q$ be a rational number with $0 < a < 1$. Assume that M_a is a totally ordered set and satisfies the parallel axiom. Then, all b-straight lines in M_a are with period (q, p).*

We need two lemmas to prove the theorem. Let $u = (u_j)_{j \in \mathbf{Z}}$ be a periodic b-straight line with period (q, p) and let $s = (s_j)_{i_0 \le j \le i_0 + mq}$ be a b-segment. Let $f_s(i) = s_{i_0 + iq} - u_{i_0 + iq}$ for any $i \in I[0, m]$ where $I[a, b]$ is the set $\{a, a+1, \dots, b\}$ of integers. We say that $I[a, b]$ is a *maximal monotone interval* for f_s in $I[0, m]$ if $f_s(i)$ is a monotone sequence in $i \in I[a, b]$ and is not a monotone sequence in any interval of integers containing $I[a, b]$ as a proper subset.

Lemma 4.2. *Let $I[a_1, b_1]$ and $I[a_2, b_2]$ be maximal monotone intervals. If $I[a_1, b_1] \cap I[a_2, b_2]$ contains at least three numbers, then s is a sub-b-segment of a periodic b-geodesic with period (q, p).*

Proof. Let $I[a_1, b_1] \cap I[a_2, b_2] \ni i_1 - 1, i_1, i_1 + 1$. Then, $f_s(i_1 - 1) = f_s(i_1) = f_s(i_1 + 1)$, namely,

$$\begin{aligned} s_{i_0 + (i_1 - 1)q} - u_{i_0 + (i_1 - 1)q} &= s_{i_0 + i_1 q} - u_{i_0 + i_1 q} \\ &= s_{i_0 + (i_1 + 1)q} - u_{i_0 + (i_1 + 1)q}. \end{aligned}$$

Since

$$u_{i_0 + (i_1 - 1)q} + pL = u_{i_0 + i_1 q} = u_{i_0 + (i_1 + 1)q} - pL,$$

we have

$$s_{i_0 + (i_1 - 1)q} + pL = s_{i_0 + i_1 q} = s_{i_0 + (i_1 + 1)q} - pL,$$

Hence, $U(q, p)(s_{i_0 + (i_1 - 1)q}) = s_{i_0 + i_1 q}$, $U(q, p)(s_{i_0 + i_1 q}) = s_{i_0 + (i_1 + 1)q}$. This implies that s is a sub-b-segment of a periodic b-geodesic with period (q, p). $\square$

Lemma 4.3. *Let $I[a_1, b_1], \dots, I[a_n, b_n]$ be maximal monotone intervals with $a_1 < \cdots < a_n$ and $I[a_k, b_k] \cap I[a_{k+1}, b_{k+1}] \neq \emptyset$ for $k = 1, \dots, n-1$. Then, n is less than or equal to 2.*

Proof. Suppose without loss of generality that $f_s(i)$ is monotone nonincreasing in $i \in I[a_1, b_1]$, monotone nondecreasing in $i \in I[a_2, b_2]$, monotone nonincreasing in $i \in I[a_3, b_3]$, and so on. Suppose $n \ge 3$. It follows from Lemma 4.2 that we find $a, b \in I[1, n-1]$ such that $a < b$ and

$$f_s(a-1) \ge f_s(a), \quad f_s(a) < f_s(a+1)$$

and

$$f_s(b-1) \le f_s(b), \quad f_s(b) > f_s(b+1),$$

namely,

$$s_{i_0+(a-1)q} \geq s_{i_0+aq} - pL, \quad s_{i_0+aq} < s_{i_0+(a+1)q} - pL$$

and

$$s_{i_0+(b-1)q} \leq s_{i_0+bq} - pL, \quad s_{i_0+bq} > s_{i_0+(b+1)q} - pL.$$

Let $\bar{s} = U(q,p)(s)$, namely, $\bar{s}_j = s_{j-q} + pL$ for all $j \in I[i_0+q, i_0+(m+1)q]$. Then, we have

$$\bar{s}_{i_0+aq} = s_{i_0+(a-1)q} + pL \geq s_{i_0+aq}$$

$$\bar{s}_{i_0+(a+1)q} = s_{i_0+aq} + pL < s_{i_0+(a+1)q},$$

and

$$\bar{s}_{i_0+bq} = s_{i_0+(b-1)q} + pL \leq s_{i_0+bq}$$

$$\bar{s}_{i_0+(b+1)q} = s_{i_0+bq} + pL > s_{i_0+(b+1)q}.$$

This implies that $T(\bar{s})$ crosses $T(s)$ at least two times and one of their intersection is not any endpoint of $T(\bar{s})$ and $T(s)$. It contradicts to Proposition 2.2, so we have $n \leq 2$. $\square$

PROOF OF THEOREM 4.1. Suppose $s = (s_j)_{j \in \mathbf{Z}}$ in M_a is not with period (q,p). Let $u = (u_j)_{j \in \mathbf{Z}}$, $t = (t_j)_{j \in \mathbf{Z}}$ in M_a such that $u_0 < s_0 < t_0$, u and t are with period (q,p) and any b-straight line $v = (v_j)_{j \in \mathbf{Z}}$ in M_α with $u_0 < v_0 < t_0$ is not with period (q,p) (see Proposition 2.3). Then, it follows that either $\lim_{j\to\infty}|s_j - u_j| = 0$ and $\lim_{j\to-\infty}|s_j - t_j| = 0$ or $\lim_{j\to\infty}|s_j - t_j| = 0$ and $\lim_{j\to-\infty}|s_j - u_j| = 0$. In fact, if $\lim_{j\to\infty}|s_j - u_j| \neq 0$, for example, then $\bar{s} = \lim_{n\to-\infty} U(q,p)^n(s)$ is a periodic b-straight line with period (q,p) which is between u and t, contradicting to the choice of u and t. We assume without loss of generality that the former case occurs. Let $s^n_{nq} = s_0 + npL$ for each n and let $s^n = (s^n_j)_{0\leq j\leq mq}$ be a b-segment with $s^n_0 = s_0$ and $s^n_{nq} = s_0 + npL$. Then, there exists a subsequence s^m of s^n which converges to a b-ray $w = (w_j)_{0\leq j}$ from $w_0 = s_0$. It follows from the property of the strip $[T(u), T(t)]$ that either $\lim_{j\to\infty}|w_j - u_j| = 0$ or $\lim_{j\to\infty}|t_j - w_j| = 0$.

If $\lim_{j\to\infty}|t_j - w_j| = 0$, then it follows from Lemma 2.6 that w is a co-b-ray from s_0 to t. Since s is the asymptote through s_0 to t, this contradicts that the unique co-b-ray from s_0 to t is a sub-b-ray of s (see Proposition 2.8). So we assume that $\lim_{j\to\infty}|u_j - w_j| = 0$. Let $f_{s^n}(i) = s^n_{iq} - u_{iq}$ for $i \in I[0,n]$. Let $I[0,a_n]$

and $I[b_n, n]$ be maximal monotone intervals in $I[0, n]$ for f_{s^n}. It follows from Lemma 4.3 that $f_{s^n}(i)$ is monotone nonincreasing in $i \in I[0, a_n]$ and monotone nondecreasing in $i \in I[b_n, n]$, and, $f_{s^n}(b_n) \to 0$ as $n \to \infty$. Set $\bar{s}^n = U(-nq, -np)(s^n)$. Then, $\bar{s}_0^n = s_0$ and $\bar{s}_{-nq}^n = s_0 - npL$. Let w' be a b-ray which is a limit of converging subsequence of $\bar{s}^n$. Then, w' is a co-b-ray to $-u$, since $f_{\bar{s}^n}(b_n - n) \to 0$ as $n \to \infty$. However, this is impossible, since $-s$ is the unique co-b-ray to $-u$ through s_0 where $-u = (u_j)_{j \leq 0}$. This contradiction comes from the original assumption in the way of our proof. □

5. Examples

In this section we show some examples of foliations of $\mathbf{X}$ by b-straight lines which satisfy the parallel axiom.

Example 5.1. *Let a be an arbitrary irrational number with $0 < a < 1$. Suppose there exists a φ-invariant closed curve f not null-homotopic in Ω such that $\alpha(\bar{x}) = aL$ for all $\bar{x} \in f$. Then, $M_a = f$ and it satisfies the parallel axiom.*

This example was stated in [9], although the parallel axiom was not proved. We give a proof here.

Proof. We have only to prove that any b-ray $s = (s_j)_{j \geq i_0}$ from s_{i_0} with slope $\alpha(s) = aL$ is a sub-b-ray of the unique b-straight line $s' = (s_j')_{j \in \mathbf{Z}}$ passing through $s_{i_0}' = s_{i_0}$ in M_a. In order to prove this we assume without loss of generality that $0 \leq s_{i_0} < L$ and $s_{i_0+1}' < s_{i_0+1}$. Let $q, p \in \mathbf{Z}^+$ with $p/q > a$. Let $D(q,p)(t_{i_0}) = \min D(q,p)$ with $0 \leq t_{i_0} < L$ and $t = (t_j)_{j \in \mathbf{Z}}$ the minimal periodic b-straight line with period (q,p). Then, the set $\{c(t_j) \mid j \in \mathbf{Z}\}$ consists of q points. Let $u_0 < u_1 < \cdots < u_{q-1}$ with $0 \leq u_i < L$ for any $i = 0, \dots, q-1$ be the parameters of such points with respect to the boundary curve c. Let a number k be such that $u_k \leq s_{i_0} < u_{k+1}$ and let $v = (v_j)_{j \in \mathbf{Z}}$ and $v' = (v_j')_{j \in \mathbf{Z}}$ be minimal periodic b-straight lines with period (q,p) and with $v_{i_0} = u_k$, $v_{i_0}' = u_{k+1}$. Since $p/q > a$, the b-ray $T(s)$ is under $T(v')$ and intersects $T(v)$ just once. There exists a subsequence of v (resp., v') such that it converges to a b-straight line $w = (w_j)_{j \in \mathbf{Z}}$ (resp., $w' = (w_j')_{j \in \mathbf{Z}}$) with slope aL as $p/q \to a$. From the construction of w and w' it follows that s and s' are in the strip $[T(w), T(w')]$, and $|w_j' - w_j| \to 0$ as $j \to \infty$. In particular, we see that $s_j - s_j' \to 0$. Lemma 2.6 shows that s is a co-b-ray to s', contradicting to Proposition 2.8. □

Example 5.2. *Let f be a φ-invariant closed curve not null-homotopic in Ω with slope aL $(0 < a < 1)$. Let g_n and h_n be sequences of closed curves not null-homotopic in Ω with slope $\alpha(g_n) < aL$ and $\alpha(h_n) > aL$. If they converges to the closed curve f, then M_a satisfies the parallel axiom.*

Proof. We first prove that $M_a = f$. Suppose for indirect proof that $M_a \neq f$, namely, there exists a point $\bar{x} = (c(s_0), u_0) \in M_a$ with $\bar{x} \notin f$. Then, there exist b-straight lines $v = (v_j)_{j \in \mathbf{Z}}$ and $w = (w_j)_{j \in \mathbf{Z}}$ with $v_0 = w_0 = s_0$ and slope aL. Assume without loss of generality that $v_1 < w_1$. Let $v^n = (v_j^n)_{j \in \mathbf{Z}}$ (resp., $w^n = (w_j^n)_{j \in \mathbf{Z}}$) be the b-straight line corresponding to the unique point in g_n (resp., h_n) with $v_0^n = s_0$ (resp., $w_0^n = s_0$). Then, $v_j^n < v_j$ and $w_j < w_j^n$ for all $j > 0$. This means that $\lim_{n \to \infty} v^n \neq \lim_{n \to \infty} w^n$, contradicting to $g_n, h_n \to f$ as $n \to \infty$.

We will prove that M_a satisfies the parallel axiom. As was seen in the above M_a gives a foliation of $\mathbf{X}$. Let $s = (s_j)_{j \in \mathbf{Z}}$ and $t = (t_j)_{j \in \mathbf{Z}}$ be in M_a with $s_0 < t_0$. Let s and t correspond to points $\bar{x}$ and $\bar{y}$ in Ω, respectively. Let $\bar{x}_n \in h_n$ (resp., $\bar{y}_n \in g_n$) be such that the first coordinates of $\bar{x}_n$ (resp., $\bar{y}_n$) and $\bar{x}$ (resp., $\bar{y}$) are equal. Let $s^n = (s_j^n)_{j \in \mathbf{Z}}$ (resp., $t^n = (t_j^n)_{j \in \mathbf{Z}}$) be configurations of $\bar{x}_n$ (resp., $\bar{y}_n$) with $s_0^n = s_0$ (resp., $t_0^n = t_0$). Then, s and t are b-straight lines in $\mathbf{X}$. Since $\alpha(s^n) > aL$ (resp., $\alpha(t^n) < aL$), we see that s^n intersects t (resp., t^n intersects s), and $s^n \to s$ (resp., $t^n \to t$) as $n \to \infty$. It follows from Lemma 2.7 that s and t are b-asymptotes to each other. The same argument is valid for the reversed b-straight lines $-\bar{x}$ and $-\bar{y}$. This completes the proof. □

Example 5.3. *Suppose the slope function α in Ω is continuous. Let a be a number with $0 < a < 1$. If $\alpha^{-1}(aL)$ has no interior points, then, M_a satisfies the parallel axiom. In particular, M_a satisfies the parallel axiom if a is an irrational number.*

Proof. Since the set K^n (resp., N^n) of all points $\bar{x}$ in Ω with $\alpha(\bar{x}) < aL - 1/n$ (resp., $\alpha(\bar{x}) > aL + 1/n$) is a φ-invariant open set in Ω, it follows from Birkhoff's theorem (see [10]) that the boundary ∂K^n (resp., N^n) is a φ-invariant closed curve g_n (resp., h_n) not null-homotopic in Ω with slope $\alpha(g_n) = aL - 1/n$ (resp., $\alpha(h_n) = aL + 1/n$). Since $\alpha^{-1}(aL)$ has no interior points, we have $\lim_{n \to \infty} g_n = \lim_{n \to \infty} h_n =: f$. Example 5.2 shows that $M_a = f$ and it satisfies the parallel axiom. □

Example 5.4. *Suppose there exists a pole $x \in C$. Then, M_a satisfies the parallel axiom for any irrational number a with $0 < a < 1$.*

Proof. Let $q, p \in \mathbf{Z}^+$ with $p/q < 1$. Then, it follows from Lemma 2.11 that there exists a foliation W of $\mathbf{X}$ whose b-straight lines are with slope pL/q. For any irrational number a with $0 < a < 1$ we have the foliation of $\mathbf{X}$ with slope aL as the limit set of W as $p/q \to a$. Example 5.1 shows that M_a satisfies the parallel axiom. $\square$

6. Proofs

Proof of Theorem 1.1. Let C_{a_n} be the sequence of convex sets as in Lemma 3.1 and let $C_{L/2}$ be its limit set. Since $C_{L/2}$ is contained in every diameter of C, the set $C_{L/2}$ consists of only one point O. Thus C is a circle with center O. $\square$

Let $x = (x_j)_{j \in \mathbf{Z}}$ be a billiard ball trajectory whose configuration is a b-straight line $s = (s_j)_{j \in \mathbf{Z}}$. Let $x^- = (x_j^-)_{j \in \mathbf{Z}}$ be its reversed billiard trajectory whose configuration is a b-straight line $s^- = (s_j^-)_{j \in \mathbf{Z}}$ with $s_0^- = s_0$. Then, it follows that $s_j^- = jL + s_{-j}$ for all $j \in \mathbf{Z}$. Therefore, we have that $\alpha(x^-) = L - \alpha(x)$.

Proof of Corollary 1.2. Since $\lim_{n\to\infty} f_n = \lim_{n\to\infty} f_n^- =: f$, it follows from $\alpha(f^-) = L - \alpha(f)$ and Example 5.2 that $\alpha(f) = L/2$ and $M_{L/2}$ satisfies the parallel axiom. Theorem 4.1, Proposition 2.5 and Theorem 1.1 prove Corollary 1.2. $\square$

Proof of Corollary 1.3. Since the slope function α is continuous in Ω and $\alpha^{-1}(L/2)$ has no interior points, we can find a sequence of closed curves in Ω as in the assumption of Corollary 1.2. $\square$

Proof of Corollary 1.4. It follows from Lemma 2.11 that there exists a sequence of φ-invariant closed curves not null-homotopic in Ω with slope $(n-2)L/2n$. Theorem 1.1 proves Corollary 1.4. $\square$

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. Bialy. Convex billiards and a theorem of E. Hopf, Math. Z., Vol 214, 147–154 (1993).
[4] H. Busemann. The geometry of geodesics, New York, Academic Press, 1955.

[5] L. Green. A theorem of E. Hopf, Michigan Math. J., Vol 5, 31–34 (1958).
[6] E. Gutkin and A. Katok. Caustics for inner and outer billiards, Commun. Math. Phys., Vol 173, 101–133 (1995).
[7] E. Hopf. Closed surfaces without conjugate points, Nat. Acad. Sci. U.S.A., Vol 34, 47–51 (1948).
[8] N. Innami. Integral formulas for polyhedral and spherical billiards, J. Math. Soc. Japan, Vol 50, No 2, 339–357 (1998).
[9] N. Innami. Geometry of geodesics for convex billiards and circular billiards, Nihonkai Math. J, Vol 13, No 1, 73–120 (2002).
[10] A. Katok and B. Hasselblatt. Introduction to the modern theory of dynamical systems, New York, Cambridge University Press, 1995.
[11] M. P. Wojtkowski. Two applications of Jacobi fields to the billiard ball problems, J. Diff. Geom., Vol 40, 155–164 (1994).

(Shinetsu Tamura)
Graduate School of Science and Technology
Niigata University
Niigata, 950-2181, Japan

(Nobuhiro Innami)
Department of Mathematics
Faculty of Science, Niigata University
Niigata, 950-2181, Japan