JOURNAL OF MODERN DYNAMICS
VOLUME 1, NO. 2, 2007, 155–173 WEB SITE: http://www.math.psu.edu/jmd

ON CONFIGURATION SPACES OF PLANE POLYGONS, SUB-RIEMANNIAN GEOMETRY AND PERIODIC ORBITS OF OUTER BILLIARDS

DANIEL GENIN AND SERGE TABACHNIKOV
(Communicated by Anton Zorich)

ABSTRACT. Following a recent paper by Baryshnikov and Zharnitsky, we consider outer billiards in the plane possessing invariant curves consisting of periodic orbits. We prove the existence and abundance of such tables using tools from sub-Riemannian geometry. We also prove that the set of 3-periodic outer billiard orbits has empty interior.

1. INTRODUCTION

The classical billiard system describes the motion of a point in a plane domain subject to the elastic reflection off the boundary, described by the familiar law of geometric optics: the angle of incidence equals the angle of reflection; see, e.g., [15, 16] for surveys of mathematical billiards.

For every $n \geq 2$, the billiard system inside a circle has a special property: every point of the circle is the starting point of an n-periodic billiard orbit; this orbit is an inscribed regular n-gon. Similarly, one can consider periodic orbits with other rotation numbers corresponding to regular star-shaped polygons. Likewise, an ellipse enjoys the same property for every $n \geq 3$; the orbits are inscribed n-gons of extremal perimeter length (for circles and ellipses, this behavior is due to the complete integrability of the billiard ball map). A billiard table of constant width also has the property that every point on its boundary belongs to a 2-periodic, back-and-forth, billiard trajectory, and this is a dynamic characterization of the curves of constant width. The phase space of the billiard ball map is a cylinder (assuming that the table is simply connected), and the family of n-periodic orbits forms an invariant circle of the billiard ball map consisting of n-periodic points.

How exceptional is this property? More specifically, given $n \geq 3$, one wants to describe plane billiards such that the billiard ball map has an invariant curve consisting of n-periodic points. The first result in this direction was obtained by Innami [9]: he constructed a billiard curve, other than a circle or an ellipse,

---

Received April 17, 2006.
2000 *Mathematics Subject Classification:* Primary 53C17, 37J45.
Key words and phrases: outer billiards, invariant curves, sub-Riemannian geometry.
Partially supported by NSF grant DMS-0555803.

whose every point is a vertex of a 3-periodic billiard trajectory (these billiard triangles are isosceles in his example). A similar result was obtained by M. Berger (unpublished).

In recent papers [1, 2], Baryshnikov and Zharnitsky developed a new elegant approach to this problem.[1] In a nutshell, their innovation was in switching focus from billiard curves to the space of plane n-gons. If such an n-gon, $z_1, \dots, z_n$, is a periodic billiard trajectory then, due to the law "the angle of incidence equals the angle of reflection", we know the directions of the billiard curve at each vertex z_i. This determines an n-dimensional distribution $\mathscr{D}$ in the space of n-gons, called the *Birkhoff distribution*. One can further restrict attention to the hypersurface of n-gons with a fixed perimeter length: the Birkhoff distribution is everywhere tangent to these hypersurfaces.

It turns out that the Birkhoff distribution is nonintegrable. A 1-parameter family of billiard n-gons $z_1(t), \dots, z_n(t)$, $0 \le t \le 1$, is interpreted as a curve in the space of n-gons, tangent to $\mathscr{D}$ and satisfying the "monodromy" condition $z_i(1) = z_{i+1}(0)$, $i = 1, \dots, n-1$, and $z_n(1) = z_1(0)$. The problem of finding such *horizontal* curves belongs to sub-Riemannian geometry (and this is the reason for mentioning it in the title of this paper, although we do not study horizontal curves of extremal length, the main subject of sub-Riemannian geometry).

The main result of [2] is that, for every $n \ge 2$, the space of billiard tables, for which the billiard ball map possesses an invariant curve consisting of n-periodic points, is infinite-dimensional (this space also has an infinite codimension in the space of all billiard tables).

EXAMPLE 1.1. Let us illustrate this approach in the simplest case $n = 2$, that is, curves of constant width (for information on billiards of constant width, see [10]).

Consider the space of oriented segments, say, of length 2. A segment z_1z_2 is "allowed" to move in such a way that the velocities of both endpoints are orthogonal to the segment. Let $O = (x, y)$ be the midpoint of the segment and α its direction. Then the space of segments has coordinates (x, y, α) and

$$z_1 = (x - \cos\alpha, y - \sin\alpha),\ z_2 = (x + \cos\alpha, y + \sin\alpha).$$

We conclude that the velocity of point O is orthogonal to the vector $(\cos\alpha, \sin\alpha)$, and this is the definition of the Birkhoff distribution in this case. Thus $\mathscr{D}$ is the kernel of the 1-form $\lambda = \cos\alpha\ dx + \sin\alpha\ dy$. This 1-form is contact: $\lambda \wedge d\lambda \ne 0$. We have identified our space of segments with the space of cooriented contact elements in the plane, a fundamental example of a contact 3-dimensional manifold (see, e.g., [5]).

If $z_1(t)z_2(t)$ is a curve, tangent to $\mathscr{D}$ and satisfying the monodromy condition $z_1(1) = z_2(0), z_2(1) = z_1(0)$, then the respective curve $(x(t), y(t), \alpha(t))$ is Legendrian (i.e., tangent to the contact distribution) and satisfies $x(1) = x(0), y(1) = y(0)$ and $\alpha(1) = \alpha(0) + \pi$.

---

[1] This approach is also developed in [11]. We learned about this paper after the completion of the present work.

The projection of this Legendrian curve on the (x, y)-plane, that is, the trajectory of the midpoint O, is a closed smooth curve Δ with singularities (generically, semicubical cusps); the total winding of this curve is π. Every such curve Δ uniquely lifts to a Legendrian curve in the space of contact elements: the missing α coordinate is recovered as the slope: $\cot\alpha = -dy/dx$.

In other words, we use Δ as a guide: place a segment of length 2 so that its midpoint O is on the curve Δ and the segment is orthogonal to Δ, and then slide the segment around Δ. If certain convexity conditions hold (for example, Δ should not have inflection points) then the endpoints z_1 and z_2 will describe a curve of constant width.

For example, if the endpoints z_1 and z_2 move along a circle then Δ degenerates to its center, and if z_1 and z_2 describe the Reuleaux triangle then Δ is made of three 60° arcs of a circle, see Figure 1.

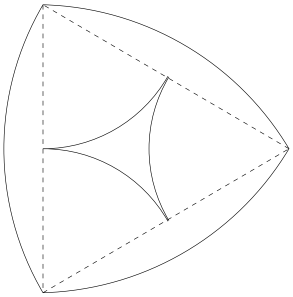

FIGURE 1. Reuleaux triangle and the curve Δ

In this paper we extend the results from [2] to outer (or dual) billiards, a dynamical system that is a close relative of the conventional, inner, billiards; see [4, 15, 16] for surveys and [6] for a recent result concerning hyperbolic behavior of outer billiards.

An outer billiard table is a strictly convex compact plane domain D. Pick a point x outside D. There are two support lines from x to D; choose one of them, say, the right one from x's view-point, and reflect x in the support point. One obtains a new point, y, and the transformation $T\colon x \mapsto y$ is the outer billiard map, see Figure 2. If the boundary of the outer billiard table contains straight segments (for example, is a convex polygon), the map T or its inverse is not defined on the lines containing these segments; this still leaves a set of full measure on which all iterations of T are defined. We will consider only outer billiard tables with strictly convex smooth boundaries.

Let us mention four fundamental properties of outer billiards, important for us. First, the outer billiard map commutes with affine transformations of the plane. Secondly, the outer billiard map preserves the standard area form in the plane (similarly to the billiard ball map which is an area preserving map of a cylinder). Thirdly, the outer billiard map is a twist map of the exterior of the

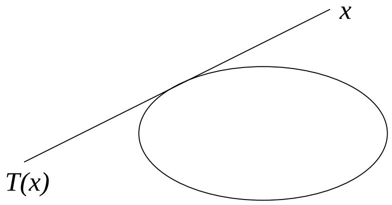

FIGURE 2. Outer billiard map

outer billiard table D, a semi-infinite cylinder, with respect to the vertical foliation consisting of tangent half-lines to the boundary of D. And finally, an n-periodic outer billiard orbit is an n-gon (possibly, star-shaped), circumscribed about the table D and having extremal area. This is an outer analog of a variational property of periodic billiard trajectories: they are inscribed polygons of extremal perimeter length. As a consequence, if an outer billiard has a 1-parameter family of n-periodic orbits then the respective circumscribed n-gons have the same areas.

Let G_n be the *cyclic configuration space* of the plane, that is, the space of polygons $z_1,\dots,z_n$ such that $z_i \neq z_{i+1}$ for $i = 1,\dots,n$ (where the index i is understood cyclically). If $z_1(t),\dots,z_n(t)$ is a 1-parameter family of n-periodic outer billiard orbits with $T(z_i(t)) = z_{i+1}(t)$ then each segment $z_i(t)z_{i+1}(t)$ envelopes the boundary of the outer billiard table and is bisected by the tangency point. This prompts the following definition.

Denote by $\mathscr{F}$ the n-dimensional distribution in G_n determined by the following condition: the velocities of points $z_1,\dots,z_n$ lie in $\mathscr{F}$ if, for each $i = 1,\dots,n$, the motion of the line $z_i z_{i+1}$ is an infinitesimal rotation about the midpoint of the segment $z_i z_{i+1}$. We call $\mathscr{F}$ the *dual Birkhoff distribution*.

The dual Birkhoff distribution is tangent to the level hypersurfaces of the area function, and we can restrict attention to nondegenerate polygons of a fixed nonzero area. On this hypersurface, $\mathscr{F}$ is nonintegrable: the vector fields tangent to $\mathscr{F}$ and their first commutators generate the tangent space at every point (see Theorem 2.5).

We are interested in the outer billiard tables possessing invariant curves consisting of n-periodic orbits. Such an invariant curve gives rise to a horizontal curve $z_1(t),\dots,z_n(t)$, $0 \le t \le 1$, in G_n satisfying the monodromy condition $z_i(1) = z_{i+1}(0)$, $i = 1,\dots,n$. If the outer billiard table is a circle (or an ellipse) then, for every pair (n,k) where $3 \le n, 1 \le k \le n/2$, and k coprime with n, there is an invariant curve consisting of n-periodic orbits with rotation number k (that is, making k turns around D). We show that, for every such pair (n,k), one can perturb a circle so that the resulting outer billiard still has an invariant curve consisting of n-periodic orbits with rotation number k; the space of such deformations is infinite-dimensional and depends on functional parameters (see Theorem 3.1).

In Section 4, we study the simplest case of $n = 3$ in detail; this is the closest outer billiard analog of curves of constant width. This case illustrates the general approach by explicit computations. We also discuss a discretization of the problem in which a smooth outer billiard curve is replaced by a polygon.

We also prove that, for every outer billiard table, the set of 3-periodic orbits has empty interior (see Theorem 5.3). For inner billiards, this is a well-known result having a number of different proofs [13, 14, 18, 19]. We also give different proofs, one of them follows ideas of [19] and another those of [2]. Note that there exist area preserving twist maps of a cylinder whose set of periodic points has nonempty interior.

One is naturally led to the following interesting problem: *can invariant billiard curves (inner or outer) consisting of periodic points with different periods, or the same period but different rotation numbers, coexist, unless the table is an ellipse? In particular, is it true that the only billiard of constant width, having an invariant curve consisting of 3-periodic points, is a circle?*

In conclusion of this introduction, let us mention a recent preprint [17] which is a follow-up to Section 5 of the present paper.

2. CONFIGURATION SPACE OF PLANE POLYGONS AND THE DUAL BIRKHOFF DISTRIBUTION

Let us fix some notations. The determinant of two vectors, that is, the cross-product, is denoted by $[\ ,\ ]$. A generic point in the plane is denoted by $z = (x, y)$. If $z_1 = (x_1, y_1)$ and $z_2 = (x_2, y_2)$ then $z_1 dz_2$ denotes the 1-form $x_1 dx_2 + y_1 dy_2$. Likewise, $[z_1, dz_2]$ denotes the 1-form $x_1 dy_2 - y_1 dx_2$ and $[dz_1, dz_2]$ the 2-form $dx_1 \wedge dy_2 + dx_2 \wedge dy_1$. If $w = (u, v)$ is a vector in $\mathbb{R}^2$ and $z = (x, y)$ is a point then by $w\partial z$ we mean the tangent vector $u\partial x + v\partial y$. When working with n-gons, we always understand the indices cyclically, so that $i = i + n$. For a polygon $z_1, \ldots, z_n$, set

$$A(z_1, \ldots, z_n) = \sum_{i=1}^{n} [z_i, z_{i+1}];$$

this is twice the area of the polygon (counted with appropriate multiplicities and signs).

Let $\mathscr{F}$ be the dual Birkhoff distribution on the space of n-gons defined in Section 1. Let $Z = (z_1, \ldots, z_n) \in G_n$ be an n-gon and $W = (w_1, \ldots, w_n)$ a tangent vector to G_n at Z; the vector w_i is the velocity of the vertex z_i.

LEMMA 2.1. *$\mathscr{F}$ is an n-dimensional distribution. A vector W lies in $\mathscr{F}$ if and only if, for all $i = 1, \ldots, n$, one has: $[w_i + w_{i+1}, z_{i+1} - z_i] = 0$, that is, the vectors $w_i + w_{i+1}$ and $z_{i+1} - z_i$ are collinear.*

Proof. To see that $\mathscr{F}$ is n-dimensional, view a polygon $Z = (z_1, \ldots, z_n) \in G_n$ as an ordered collection of n lines $L = (z_1 z_2, z_2 z_3, \ldots, z_n z_1)$ (the polygon can be reconstructed from these lines only if they are distinct). From this dual viewpoint, G_n becomes an open subset in the space L_n of n-tuples of oriented lines in the plane. An infinitesimal rotation of the i-th line $z_i z_{i+1}$ about the midpoint of the segment $z_i z_{i+1}$ determines a tangent vector to $\mathtt{L}_n$ at point L, and these vectors are clearly independent for $i = 1, \dots, n$. It follows that $\dim \mathcal{F} = n$.

For W to belong to $\mathcal{F}$, the line through points $z_i + \varepsilon w_i$ and $z_{i+1} + \varepsilon w_{i+1}$ must pass through the midpoint $(z_i + z_{i+1})/2$; here ε is infinitesimal. Thus the linear term in ε of the determinant

$$\left[\frac{z_{i+1} - z_i}{2} + \varepsilon w_{i+1}, \frac{z_i - z_{i+1}}{2} + \varepsilon w_i\right]$$

vanishes, which is equivalent to $[w_i + w_{i+1}, z_{i+1} - z_i] = 0$. □

An n-gon is called *nondegenerate* if no three consecutive vertices lie on one line. The set of nondegenerate n-gons is denoted by U_n. Clearly, U_n is an open subset of G_n.

Let us introduce the 1-forms $\alpha_i = [z_{i+1} - z_i, dz_{i+1} + dz_i]$, $i = 1, \dots, n$.

LEMMA 2.2. *Let $Z \in U_n$ be a nondegenerate polygon. Then the 1-forms α_i are linearly independent at Z and the intersection of their kernels is the fiber $F(Z)$ of $\mathcal{F}$ at point Z.*

Proof. First, we claim that each form α_i vanishes on every tangent vector $W \in \mathcal{F}$. Indeed, $\alpha_i(W) = [z_{i+1} - z_i, w_i + w_{i+1}]$, and this is zero, according to Lemma 2.1.

Suppose a nontrivial linear combination $\sum t_i \alpha_i$ vanishes at point Z. Then, for every test vector $W = (w_1, \dots, w_n)$, one has:

$$\begin{aligned} 0 = \sum t_i [z_{i+1} - z_i, w_i + w_{i+1}] &= \sum t_i [z_{i+1} - z_i, w_i] + \sum t_i [z_{i+1} - z_i, w_{i+1}] \\ &= \sum [t_i (z_{i+1} - z_i), w_i] + \sum [t_{i-1}(z_i - z_{i-1}), w_i] \\ &= \sum [t_i (z_{i+1} - z_i) + t_{i-1}(z_i - z_{i-1}), w_i]. \end{aligned}$$

Therefore

$$t_i (z_{i+1} - z_i) + t_{i-1}(z_i - z_{i-1}) = 0, \quad i = 1, \dots, n. \tag{2.1}$$

Since the polygon Z is nondegenerate, the vectors $z_{i+1} - z_i$ and $z_i - z_{i-1}$ are linearly independent, and (2.1) implies that $t_i = 0$ for all i.

Since $\mathcal{F}$ is n-dimensional and the 1-forms α_i are linearly independent, one has: $\bigcap_{i=1}^n \operatorname{Ker} \alpha_i = \mathcal{F}$. □

The next lemma shows that one can restrict attention to the level hypersurfaces of the area function.

LEMMA 2.3. *Every nonzero level hypersurface of the function A is smooth, and the distribution $\mathcal{F}$ is tangent to these hypersurfaces.*

Proof. Let $Q: V \to V^*$ be a linear operator in a vector space V. Consider the quadratic function $F(x) = Q(x) \cdot x$ on V. Then every nonzero level hypersurface of F is smooth. Indeed, by Euler's formula, $\nabla F(x) \cdot x = 2F(x)$, hence dF does not vanish on a nonzero level hypersurface. The function A is a quadratic form on the space of n-gons, and the first claim follows.

One has the equality

$$\sum_{i=1}^n \alpha_i = \sum_{i=1}^n [z_{i+1}, dz_i] - [z_i, dz_{i+1}] = -dA,$$

which implies the second claim. □

Let us construct a system of n linearly independent vector fields tangent to $\mathscr{F}$. Let $Z = (z_1, \dots, z_n)$ be a nondegenerate polygon. Set $a_i = [z_i - z_{i-1}, z_{i+1} - z_i]$, twice the oriented area of the triangle $z_{i-1}z_iz_{i+1}$. Denote by $W_k = (w_{1k}, \dots, w_{nk})$ the tangent vector to U_n at Z whose i-th component is zero, unless $i = k$ or $i = k+1$. For these values of i, one has:

$$w_{k,k} = a_{k+1}(z_k - z_{k-1}), \quad w_{k+1,k} = a_k(z_{k+2} - z_{k+1}),$$

see Figure 3.

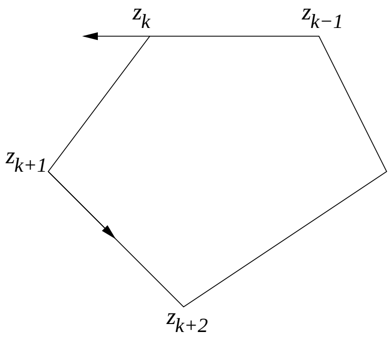

FIGURE 3. Tangent vector W_k

LEMMA 2.4. *The vectors $W_1, \dots, W_n$ are linearly independent at every point $Z \in U_n$ and they span $F(Z)$.*

Proof. To show that W_k lies in $\mathscr{F}$, it suffices to check that the vector $a_{k+1}(z_k - z_{k-1}) + a_k(z_{k+2} - z_{k+1})$ is collinear with the vector $z_{k+1} - z_k$. Indeed,

$$[z_{k+1} - z_k, z_{k+2} - z_{k+1}](z_k - z_{k-1}) + [z_k - z_{k-1}, z_{k+1} - z_k](z_{k+2} - z_{k+1}) \\ = [z_k - z_{k-1}, z_{k+2} - z_{k+1}](z_{k+1} - z_k),$$

which follows from the general identity, valid for any three vectors:

$$[u, v]\, w + [v, w]\, u + [w, u]\, v = 0.$$

Under identification of the space of n-gons with the space of n-tuples of lines L_n, the tangent vector W_k corresponds to a nontrivial infinitesimal rotation of the line z_kz_{k+1} about the midpoint of the segment z_kz_{k+1}. Therefore these vectors are linearly independent and hence span the dual Birkhoff distribution. □

Now we prove that the dual Birkhoff distribution is nonintegrable.

THEOREM 2.5. *Let M^{2n-1} be the hypersurface in U_n consisting of nondegenerate n-gons with a fixed nonzero area. Then the tangent space of M at every point is generated by the vectors fields tangent to $\mathscr{F}$ and their first commutators.*

Using terminology common in sub-Riemannian geometry, see [12], the bracket growth type of $\mathscr{F}$ is $(n, 2n-1)$ (just like that of the Birkhoff distribution, see [2]).

Proof. In what follows, we also use the notation [,] for the commutator of vector fields. We hope that it will be clear from the context when the bracket means the commutator and when a determinant.

Consider the vector fields $\xi_k = [W_{k-1}, W_k]$ (the fields W_i and W_j clearly commute if $|i-j| \geq 2$). We want to show that these fields, along with $\mathscr{F}$, span the tangent space to M at every point.

According to Lemmas 2.2 and 2.3 and the proof of the latter, the 1-forms $\alpha_1, \ldots, \alpha_{n-1}$ constitute a frame in the conormal space to $\mathscr{F}$ at every point $Z \in M$. For $k = 1, \ldots, n$, consider the vector $(\alpha_1(\xi_k), \ldots, \alpha_{n-1}(\xi_k)) \in \mathbb{R}^{n-1}$. It suffices to show that these n vectors span $\mathbb{R}^{n-1}$.

Since the fields W_k belong to the kernel of each form α_i, the Cartan formula implies: $\alpha_i(\xi_k) = -d\alpha_i(W_{k-1}, W_k)$.

Recall that $Z = (z_1, \ldots, z_n)$ with $z_i = (x_i, y_i)$. Let $\omega_i = dx_i \wedge dy_i$, $i = 1, \ldots, n$. We claim that $d\alpha_i = 2(\omega_{i+1} - \omega_i)$. Indeed,

$$d\alpha_i = [dz_{i+1} - dz_i, dz_{i+1} + dz_i] = [dz_{i+1}, dz_{i+1}] - [dz_i, dz_i] = 2(\omega_{i+1} - \omega_i).$$

Thus we want to show that the vectors

$$((\omega_2 - \omega_1)(W_{k-1}, W_k), \ldots, (\omega_n - \omega_{n-1})(W_{k-1}, W_k))$$

span $\mathbb{R}^{n-1}$ for $k = 1, \ldots, n$. This will follow if we show that the vectors

$$(\omega_1(W_{k-1}, W_k), \omega_2(W_{k-1}, W_k), \ldots, \omega_n(W_{k-1}, W_k))$$

span $\mathbb{R}^n$ for $k = 1, \ldots, n$. But $\omega_i(W_{k-1}, W_k) = 0$, unless $i = k$, and $\omega_k(W_{k-1}, W_k) = -a_{k-1}a_k a_{k+1} \neq 0$. Thus $\omega_i(W_{k-1}, W_k)$ is a nondegenerate diagonal matrix, and we are done. □

REMARK 2.6. The dual Birkhoff distribution $\mathscr{F}$ on the cyclic configuration space G_n is Lagrangian with respect to the following symplectic structure. The space of oriented lines in the plane, topologically a cylinder, has a canonical area form (see, e.g., [15, 16]), and the direct product of these forms is a symplectic structure in the space of n-tuples of oriented lines $\mathtt{L}_n$. In the proof of Lemma 2.1, we realized U_n as an open subset in $\mathtt{L}_n$, and this provides a symplectic structure on U_n. Infinitesimal rotations of distinct lines determine symplectically orthogonal tangent vectors to $\mathtt{L}_n$, hence $\mathscr{F}$ is Lagrangian. The Lagrangian property of the dual Birkhoff distribution echoes that of the Birkhoff distribution [2].

3. OUTER BILLIARDS WITH INVARIANT CURVES CONSISTING OF PERIODIC POINTS

In this section we show that, for every pair (n, k) with $3 \leq n, 1 \leq k \leq n/2$ and k coprime with n, there is an abundance of outer billiard tables, possessing invariant curves consisting of n-periodic orbits with rotation number k; such tables are constructed as perturbations of circles. More specifically, the space of such outer billiard tables has "functional dimension" $n-1$, that is, is locally represented by $n-1$ functions of one real variable.

This section relies on the material from [12], especially, Chapter 5 and Appendix D. In particular we freely use the notions of smooth Hilbert manifolds and their smooth submanifolds. Here is a rough outline of the general theory.

Let M be a smooth manifold with a smooth d-dimensional distribution $\mathscr{F}$. Denote by $H^2_{\mathscr{F}}$ the space of parameterized horizontal paths $[0,1] \to M$ whose first 2 derivatives are square-integrable (one can work with a higher number of derivatives as well). For a point $Z \in M$, denote by $H^2_{\mathscr{F}}(Z)$ the space of paths starting at Z. Then $H^2_{\mathscr{F}}(Z)$ is a Hilbert manifold.

To describe the local coordinate charts for $H^2_{\mathscr{F}}(Z)$, choose horizontal vector fields $X_1, \dots, X_d$ that form a local frame for $\mathscr{F}$. If $\gamma(t)$ is a horizontal curve then $\gamma'(t) = \sum f_i(t) X_i(\gamma(t))$ where $f = (f_1, \dots, f_d) \in L^2([0,1], \mathbb{R}^d)$. The functions f_i are the coordinates of the curve γ. Thus the functional dimension of $H^2_{\mathscr{F}}(Z)$ equals d, and the space of nonparameterized oriented horizontal curves with a fixed starting point has functional dimension $d-1$.

Returning to our context, let M be the space of nondegenerate plane n-gons with a fixed nonzero area and $\mathscr{F}$ the dual Birkhoff distribution. Let Z be a regular star-shaped n-gon with rotation number k. We have the following result.

THEOREM 3.1. *There exists a set of (parameterized) outer billiard curves, sufficiently close to a circular one and possessing invariant curves consisting of n-periodic orbits with rotation number k, which is in 1-1 correspondence with points of a smooth Hilbert submanifold of codimension $2n-1$ in $H^2_{\mathscr{F}}(Z)$.*

Proof. To fix ideas, consider the case of simple periodic n-gons, that is, the case of $k = 1$. At the end of the proof, we will indicate the small adjustment to be made in the case of $k > 1$.

Start with a circular outer billiard. One has a 1-parameter family of periodic n-gons

$$Z(t) = (z_1(t), \dots, z_n(t)) \text{ with } z_j(t) = e^{2\pi i (j+t)/n}, \ t \in [0,1] \tag{3.1}$$

circumscribed about the circular table (the table is scaled appropriately). Let $\sigma \colon G_n \to G_n$ be the cyclic permutation of the vertices of a polygon: $\sigma(z_1, \dots, z_n) = (z_2, \dots, z_n, z_1)$. Then $Z(1) = \sigma(Z(0))$.

The family (3.1) determines a horizontal curve $\gamma_0(t)$ in G_n whose endpoints are $Z(0)$ and $Z(1)$. Denote by $H^2_{\mathscr{F}}(Z(0), Z(1)) \subset H^2_{\mathscr{F}}(Z(0))$ the set of horizontal curves with terminal point $Z(1)$. If $\gamma(t) \in H^2_{\mathscr{F}}(Z(0), Z(1))$ and $\gamma(t)$ is sufficiently close to $\gamma_0(t)$ then $\gamma(t)$ corresponds to a 1-parameter family of convex n-gons whose vertices trace n arcs, sufficiently close to the unit circle, that smoothly match together to form an invariant curve of the outer billiard map consisting of n-periodic points. The outer billiard curve is recovered as the envelope of the segments connecting the consecutive vertices.

Thus we wish to describe a neighborhood of γ_0 in $H^2_{\mathscr{F}}(Z(0), Z(1))$, namely, to show that $H^2_{\mathscr{F}}(Z(0), Z(1))$ is a Hilbert submanifold of codimension $2n-1$ in $H^2_{\mathscr{F}}(Z(0))$. Denote by $E \colon H^2_{\mathscr{F}}(Z(0)) \to M$ the endpoint map that assigns the endpoint $\gamma(1)$ to a path γ. This map is smooth, and if E is a submersion at $\gamma_0 \in H^2_{\mathscr{F}}(Z(0))$ then $H^2_{\mathscr{F}}(Z(0), Z(1)) = E^{-1}(Z(1))$ is a codimension $2n-1$ submanifold, see [12]. A curve which is a singular point of the endpoint map is called *singular*.

To prove that γ_0 is not singular we use a criterion due to Hsu [8]. Denote by $\mathscr{F}^\perp \subset T^*M$ the annihilator of the dual Birkhoff distribution and let Ω be the restriction of the canonical symplectic form on the cotangent bundle to $\mathscr{F}^\perp$. Let $\pi\colon T^*M \to M$ be the projection. A characteristic in $\mathscr{F}^\perp$ is a curve that does not intersect the zero section and whose direction at every point lies in the kernel of Ω. By Hsu's criterion, a horizontal curve is singular if and only if it is the projection under π of a characteristic curve in $\mathscr{F}^\perp$.

It is convenient for computations to use the following form of this criterion. The 1-forms α_j, introduced in Section 2, generate $\mathscr{F}^\perp$: every covector in $\mathscr{F}^\perp$ can be written as $\sum \lambda_j \alpha_j$ where the coefficients λ_j are “Lagrange multipliers". There is exactly one relation: $\sum \alpha_j = 0$ (see the proof of Lemma 2.3), and to factor this relation out, we assume that $\sum \lambda_j = 0$. Consider the 2-form $\omega(\lambda) = \sum \lambda_j d\alpha_j$ on M. Then one has the following result, see [12]: the kernel of Ω at a point $(x, \sum \lambda_j \alpha_j) \in \mathscr{F}^\perp$ is isomorphically projected by the differential $d\pi$ onto the kernel of the form $\omega(\lambda)$ restricted to the fiber $F(x)$ of $\mathscr{F}$ at point $x \in M$.

Let us find the 2-form $\omega(\lambda)$. From the proof of Theorem 2.5, we know that $d\alpha_j = 2(\omega_{j+1} - \omega_j)$ where $\omega_j = dx_j \wedge dy_j$. Thus

$$-\frac{1}{2}\omega(\lambda) = \sum_j \lambda_j(\omega_j - \omega_{j+1}) = \sum_j (\lambda_j - \lambda_{j-1})\omega_j. \tag{3.2}$$

Assume that $\dot\gamma_0(t)$ lies in the kernel of the restriction of $\omega(\lambda)$ on the fiber $F(\gamma_0(t))$:

$$\omega(\lambda)(W_k(\gamma_0(t)), \dot\gamma_0(t)) = 0, \quad k = 1, \dots, n. \tag{3.3}$$

It is straightforward to compute that $\omega_j(W_k(\gamma_0(t)), \dot\gamma_0(t)) = 0$ unless $j = k$ or $j = k+1$; in these two cases, one has:

$$\omega_k(W_k(\gamma_0(t)), \dot\gamma_0(t)) = C, \quad \omega_{k+1}(W_k(\gamma_0(t)), \dot\gamma_0(t)) = -C$$

where C is a nonzero constant depending on n only. Thus (3.2) and (3.3) imply that $2\lambda_k - \lambda_{k-1} - \lambda_{k+1} = 0$ for all k. This, in turn, implies that all λ_k are equal (indeed, by induction, $\lambda_{k+1} = k\lambda_2 - (k-1)\lambda_1$ for all k, which implies, for $k = n$, that $\lambda_1 = \lambda_2$). Since $\sum \lambda_j = 0$, all λ_k vanish. It follows that the curve γ_0 is not singular.

Finally, if the rotation number k of a periodic trajectory is greater than 1, one changes the curve (3.1) to

$$Z(t) = (z_1(t), \dots, z_n(t)) \;\text{ with }\; z_j(t) = e^{2\pi i(jk+t)/n}, \; t \in [0,1],$$

and the proof proceeds along the same lines. ☐

REMARK 3.2. To illustrate the importance of checking that the curve γ_0 is not singular, let us give a simple example of a rigid horizontal curve, that is, a curve that does not admit smooth perturbations in the class of horizontal curves with fixed endpoints, see [12]. Consider the curve $\gamma(t) = (t, 0, 0)$, $t \in [0,1]$ tangent to the distribution $dz = y^2 dx$ in 3-space. If $(x(t), y(t), z(t))$ is its perturbation as a horizontal curve with fixed endpoints then

$$0 = z(1) - z(0) = \int_0^1 y^2(t) x'(t)\, dt.$$

Since $x'(t) > 0$, we have $y(t) = 0$, and hence $z(t) = 0$, for all t; thus the perturbed curve is a reparameterization of γ.

4. CASE STUDY: $n = 3$ AND DISCRETIZATION

Three is the smallest possible period of the outer billiard map. In this section we illustrate the general constructions of Sections 2 and 3 in the case $n = 3$.

Let γ be a smooth strictly convex outer billiard curve and $z_1z_2z_3$ a 3-periodic orbit. Let w_1, w_2, w_3 be the midpoints of the sides of the triangle $z_1z_2z_3$.

LEMMA 4.1. *The inscribed triangle $w_1w_2w_3$ has the property that, for every $i = 1,2,3$, the tangent line to γ at vertex w_i is parallel to the side $w_{i+1}w_{i+2}$. Conversely, if an inscribed triangle has this property then the tangent lines to γ at its vertices form a circumscribed triangle whose sides are bisected by the tangency points. Furthermore, this property of an inscribed triangle is equivalent to having extremal area.*

Proof. The first claim holds by definition of the outer billiard map and the second follows from elementary geometry. To prove the last statement, fix points w_2 and w_3 and vary w_1. The area is extremal when the distance from w_1 to the line w_2w_3 has an extremum, and this happens if and only if the tangent line to γ at w_1 is parallel to w_2w_3. ☐

It follows that we can replace circumscribed triangles of extremal areas by inscribed ones. We want to construct smooth strictly convex closed curves such that an inscribed triangle of extremal area can be turned around inside the curve.

Following the approach of Section 2, let M be the 5-dimensional manifold of triangles of area $3/2$. Let $w_1w_2w_3$ be a triangle from M; denote by c its center of mass. Let $u = w_1 - c, v = w_2 - c$; then $[u, v] = 1$ and the frame (u, v) belongs to $SL(2,\mathbb{R})$. The cyclic group $\mathbb{Z}_3$ acts on M by the cyclic permutations of the vertices of a triangle and on $SL(2,\mathbb{R})$ by the transformation $T(u, v) = (v, -u - v)$. Extend the action of $\mathbb{Z}_3$ to $SL(2,\mathbb{R}) \times \mathbb{R}^2$ taking product with the trivial action on the second factor. On M, we have a 3-dimensional distribution $\mathscr{F}$ determined by the condition that the velocity of w_i is parallel to the line $w_{i+1}w_{i+2}$, $i = 1,2,3$.

LEMMA 4.2. *The correspondence $(w_1, w_2, w_3) \mapsto ((u, v), c)$ is a $\mathbb{Z}_3$-equivariant diffeomorphism $M \to SL(2,\mathbb{R})\times\mathbb{R}^2$. The fibers of the projection $SL(2,\mathbb{R})\times\mathbb{R}^2 \to SL(2,\mathbb{R})$ are transverse to the distribution $\mathscr{F}$.*

Proof. The first claim is obvious. To prove the second, note that a vector tangent to the fibers of the projection $SL(2,\mathbb{R}) \times \mathbb{R}^2 \to SL(2,\mathbb{R})$ is induced by a parallel translation of the triangle: all three vertices move with the same velocity. If this vector lies in $\mathscr{F}$ then it must be parallel to all three sides of the triangle, which is impossible. ☐

REMARK 4.3. One may identify M with the group of equiaffine transformations of the plane; then $\mathscr{F}$ becomes invariant distribution on a group. We do not know whether this indicates a deep connection between outer billiards and this group or this is just a coincidence.

We want to construct a closed horizontal curve $\tilde{\gamma}(t),\ t \in [0,2\pi]$ in M satisfying the monodromy condition

$$\tilde{\gamma}(t+2\pi/3) = \sigma(\tilde{\gamma}(t)) \tag{4.1}$$

where σ is the cyclic permutation of the vertices of a triangle. By Lemma 4.2, this curve projects to a smooth curve $\gamma(t)$ in $SL(2,\mathbb{R})$ satisfying $\gamma(t+2\pi/3) = T(\gamma(t))$. Conversely, such a curve $\gamma(t) \subset SL(2,\mathbb{R})$ lifts uniquely to a horizontal curve $\tilde{\gamma}(t) \subset M$, provided a lift of the initial point is chosen, but the monodromy condition (4.1) may fail. But once (4.1) is satisfied – and assuming convexity – we obtain the desired plane curve possessing a 1-parameter family of inscribed triangles of extremal areas.

Let us work out explicit formulas. Denote by $\gamma(t) = (u(t), v(t)),\ t \in [0,2\pi]$ a closed curve in $SL(2,\mathbb{R})$ such that $u(t+2\pi/3) = v(t)$ and $v(t+2\pi/3) = -u(t)-v(t)$. Let $c(t)$ be the curve of the center of mass of the respective triangle, that is, the projection to $\mathbb{R}^2$ of a lifted horizontal curve $\tilde{\gamma}(t)$. Set:

$$[u(t),u'(t)] = p(t), \quad [v(t),v'(t)] = q(t), \quad [u'(t),v(t)] = [v'(t),u(t)] = r(t).$$

PROPOSITION 4.4. *The velocity of the center of mass is given by the formula:*

$$c'(t) = \frac{1}{3}(p(t) - 2q(t) + 2r(t))\ v(t) - \frac{1}{3}(q(t) - 2p(t) + 2r(t))\ u(t). \tag{4.2}$$

The monodromy condition (4.1) holds if and only if

$$\int_0^{2\pi} \big((p(t) - 2q(t) + 2r(t))\ v(t) - (q(t) - 2p(t) + 2r(t))\ u(t)\big)\ dt = 0 \tag{4.3}$$

or, equivalently,

$$\int_0^{2\pi} [u'(t), v(t)](v(t) - u(t))\ dt = 0. \tag{4.4}$$

Proof. One has:

$$w_1(t) = u(t) + c(t),\ w_2(t) = v(t) + c(t),\ w_3(t) = -u(t) - v(t) + c(t). \tag{4.5}$$

The curve is horizontal if and only if

$$[w_2(t) - w_1(t), w_3'(t)] = [w_3(t) - w_2(t), w_1'(t)] = [w_1(t) - w_3(t), w_2'(t)] = 0$$

or, in view of (4.5),

$$[u(t),c'(t)] + 2[v(t),c'(t)] = 2r(t) - p(t), \quad 2[u(t),c'(t)] + [v(t),c'(t)] = 2r(t) - q(t)$$

(there are only two equations because the third is their consequence). It follows that

$$3[u(t),c'(t)] = 2r(t) - 2q(t) + p(t),\ 3[v(t),c'(t)] = 2r(t) - 2p(t) + q(t),$$

and since $[u(t), v(t)] = 1$, this implies (4.2).

Abusing notation, let T be the shift $t \mapsto t + 2\pi/3$ of the argument of the functions $u(t), v(t), p(t), q(t), r(t)$; this defines an action of $\mathbb{Z}_3$ on the space generated by these functions. One has:

$$T\colon\ u \mapsto v \mapsto -u - v,\ p \mapsto q \mapsto p + q - 2r,\ r \mapsto -r + q \mapsto p - r. \tag{4.6}$$

The monodromy condition (4.1) is satisfied if and only if $c(t)$ is $2\pi/3$-periodic. Applying (4.6) to (4.2), it follows that $c'(t)$ is $2\pi/3$-periodic, and hence $c(t)$ is $2\pi/3$-periodic if and only if $c(t)$ is 2π-periodic, that is, if $\int_0^{2\pi} c'(t)dt = 0$. In view of (4.2), this is equation (4.3). Finally, for every function $f(t)$, one has

$$\int_0^{2\pi} T(f)(t)\,dt = \int_0^{2\pi} f(t)\,dt.$$

Applying this to (4.6), we notice that the function $p+q-r$ is T-invariant, and hence

$$\int_0^{2\pi} (p(t)+q(t)-r(t))u(t)\,dt = \int_0^{2\pi} (p(t)+q(t)-r(t))v(t)\,dt = 0.$$

Therefore

$$\int_0^{2\pi} q(t)u(t)\,dt = \int_0^{2\pi} (p(t)+q(t)-2r(t))v(t)\,dt = -\int_0^{2\pi} r(t)v(t)\,dt.$$

Likewise one obtains other identities for the integrals of the vector-functions pu, pv, qu, qv, ru, rv:

$$\int_0^{2\pi} p(t)v(t)\,dt = -\int_0^{2\pi} r(t)u(t)\,dt$$

and

$$\int_0^{2\pi} q(t)v(t)\,dt = \int_0^{2\pi} p(t)u(t)\,dt = \int_0^{2\pi} r(t)u(t)\,dt + \int_0^{2\pi} r(t)v(t)\,dt.$$

These identities imply that (4.3) and (4.4) are equivalent. □

To summarize, one constructs an outer billiard table having a 1-parameter family of 3-periodic trajectories from a closed curve in $SL(2,\mathbb{R})$ satisfying (4.4), that is, satisfying two scalar equations.

It is interesting to consider the following discretization of the problem in which a smooth curve is replaced by a polygon. Let P be a convex polygon and ABC an inscribed triangle whose vertices are some vertices of P. One is allowed to slide one vertex of the triangle along an edge of P to the adjacent vertex of P, provided this edge is parallel to the opposite side of the triangle. Using a sequence of such moves, one wants to turn the triangle around inside P. For which polygons is this possible?

Consider an example. Let P be a regular $k = 3n+1$-gon with vertices $z_1, \dots z_k$, and let $A = z_1, B = z_{n+2}, C = z_{2n+2}$. Then the segment $z_1 z_2$ is parallel to the segment BC, and one can move point A to z_2. After that one can move point C to z_{2n+3}, then point B to z_{n+3}, etc. – see Figure 4 for $k = 7$.

One can deform a regular $k = 3n+1$-gon preserving the desired property: the condition to satisfy is that the segments $z_i z_{i+1}$ and $z_{i+n+1} z_{i+2n+1}$ be parallel for all $i = 1, \dots, k$. This gives k conditions, whereas a k-gon has $2k$ degrees of freedom, so it is not surprising that nontrivial deformations exist. We do not dwell on details here.

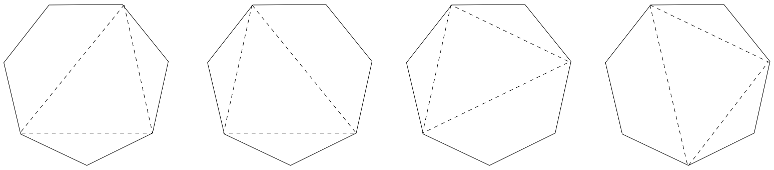

FIGURE 4. Rotating a triangle inside a regular septagon

Let us finish this section with a problem. *Suppose that a convex domain has the property that an inscribed triangle of extremal area can be turned around inside the domain. Find the lower and upper bounds for the ratio of the areas of the triangle and the domain.* We conjecture that a parallelogram and an ellipse are two extreme cases.

5. THE SET OF 3-PERIODIC ORBITS HAS EMPTY INTERIOR

It is an old conjecture, with motivation in spectral geometry, that the set of n-periodic billiard trajectories in a billiard with smooth boundary (not necessarily convex) has zero measure. For $n = 2$, this is easy to show, and for $n = 3$, this is a theorem of Rychlik [13] with a number of different proofs available [2, 14, 18, 19]. For $n \geq 4$, the conjecture is open.

This section concerns the outer counterpart of this conjecture. Consider n outer billiard tables $D_1, \dots, D_n$ with smooth boundaries and let $T_1, \dots, T_n$ be the respective outer billiard maps. We allow both clockwise and counterclockwise orientations of the boundaries, that is, “right" and “left" outer billiard maps. We are interested in the set F of fixed points of the composition $T_n \circ T_{n-1} \circ \cdots \circ T_1$.

CONJECTURE 5.1. *For every $n \geq 3$, the set F has zero measure.*

REMARK 5.2. For strictly convex outer billiards with corners, Conjecture 5.1 fails. An example, shown in Figure 5, is a rounded square: the set of 4-periodic orbits has nonempty interior. One can generalize: if a plane polygonal outer billiard has an n-periodic orbit then a whole neighborhood of this point consists of periodic points as well (with period $2n$, if n is odd). According to a theorem of Culter [3], every polygonal outer billiard has periodic orbits,[2] thus one can use any (rounded) polygon to construct an open set of periodic orbits.

We prove the following particular case of Conjecture 5.1.

THEOREM 5.3. *Let D_1, D_2, D_3 be three smooth strictly convex outer billiard tables and T_1, T_2, T_3 the respective outer billiard maps. Then the set F of fixed points of the map $T_3 \circ T_2 \circ T_1$ has empty interior.*

As in [13, 14, 18, 19], this theorem implies that F has zero measure; we do not dwell on this refinement.

---

[2] For inner polygonal billiards, this is an outstanding open conjecture.

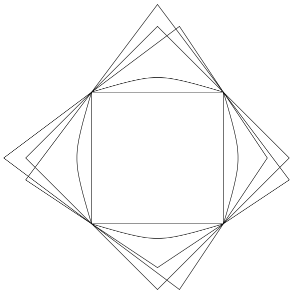

FIGURE 5. An open set of 4-periodic orbits for a piece-wise smooth outer billiard table

As a preparation to the proof of Theorem 5.3, let us recall two results on outer billiards. The first is an outer analog of the "mirror equation" of geometrical optics that plays an important role in the theory of inner billiards.

Let D be an outer billiard table with smooth boundary and T the outer billiard map. Let $T(x) = y$, see Figure 6.

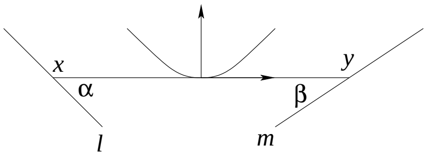

FIGURE 6. The differential of the outer billiard map and the mirror equation

PROPOSITION 5.4. *In the Cartesian coordinate system in Figure 6, the differential of T at point x is given by the matrix*

$$dT = \begin{pmatrix} -1 & -\frac{2\rho}{r} \\ 0 & -1 \end{pmatrix}$$

where $r = |xy|/2$ and ρ is the radius of curvature of the outer billiard curve at the reflection point. If dT takes a line l to a line m, see Figure 6, then one has the following "mirror equation":

$$\cot\alpha + \cot\beta = \frac{2\rho}{r}. \tag{5.1}$$

If we change the orientation of the outer billiard curve then the sign of ρ changes to the opposite. For a proof of Proposition 5.4, see [7].

The second result is the following "area construction", an outer counterpart of the string construction in the theory of inner billiards, that recovers a billiard table from a caustic.

PROPOSITION 5.5. *Let D be an outer billiard table and C an invariant curve of the outer billiard map T. Let $x \in C$ and $T(x) = y$. Then the area cut by the segment*

xy from C is the same for all $x \in C$. In other words, D can be recovered from C as an envelope of segments of constant areas.

For a proof of Proposition 5.5, see [4, 15, 16] or [7].

In particular, let C consist of two nonparallel lines, l and m, as in Figure 6. Then the area construction yields a hyperbola for which the lines l and m are the asymptotes. If the outer billiard map T takes x to y and $dT(l) = m$, as in Figure 6, then the hyperbola for which the lines l and m are the asymptotes is second-order tangent to the outer billiard curve at the midpoint of the segment xy. Note also that of 4 possible orientations of the pair of lines l and m in Figure 6, only two are consistent with the action of dT: one line is oriented toward their intersection point while the other is oriented from it.

Proof of Theorem 5.3. We give two proofs. The first is reminiscent of the one given in [19].

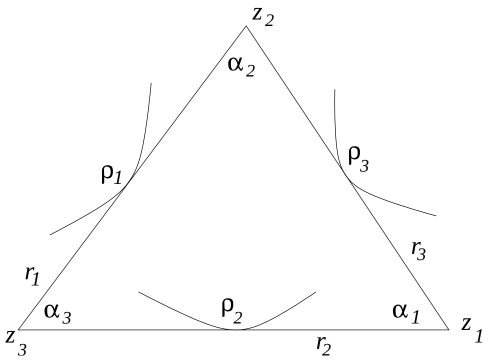

FIGURE 7. Proving Theorem 5.3

Consider Figure 7. One has three outer billiard maps, T_1, T_2, T_3 that cyclically permute points z_1, z_2, z_3:

$$T_1(z_2) = z_3,\ T_2(z_3) = z_1,\ T_3(z_1) = z_2.$$

Let α_i, $i = 1,2,3$, be the angles of the triangle, $2r_i$ its side lengths and ρ_i the curvature radii of the outer billiard curves at the reflection points. Let A_i be the matrix of dT_i in the coordinate system as in Figure 6, and let B_i be matrix of the clockwise rotation through angle $\pi - \alpha_i$.

Assume that z_2 belongs to an open set U of fixed points of the map $T_3 \circ T_2 \circ T_1$. Then $d(T_3 \circ T_2 \circ T_1)$ is the identity in U. Using Proposition 5.4, this implies that

$$B_2 A_3 B_1 A_2 B_3 A_1 = \mathrm{Id} \quad \text{or} \quad A_1 B_2 A_3 = B_3^{-1} A_2^{-1} B_1^{-1}. \tag{5.2}$$

The matrices involved are as follows:

$$A_i = \begin{pmatrix} -1 & \frac{-2\rho_i}{r_i} \\ 0 & -1 \end{pmatrix}, \quad B_i = \begin{pmatrix} -\cos\alpha_i & -\sin\alpha_i \\ \sin\alpha_i & -\cos\alpha_i \end{pmatrix}.$$

It is straightforward to compute the triple products of the matrices on both sides of the second equation (5.2). Apply both products to the vector $(1,0)$ and equate the second components:

$$\sin\alpha_2 = -\sin(\alpha_1+\alpha_3) + 2\sin\alpha_1\sin\alpha_3\frac{\rho_2}{r_2},$$

and since $\alpha_2 = \pi - \alpha_1 - \alpha_3$,

$$\frac{\rho_2}{r_2} = \frac{\sin(\alpha_1+\alpha_3)}{\sin\alpha_1\sin\alpha_3} = \cot\alpha_1 + \cot\alpha_3. \tag{5.3}$$

Let A be the area of the triangle z_1, z_2, z_3 and h_2 its altitude from the vertex z_2. Then by elementary geometry,

$$2r_1\sin\alpha_3 = 2r_3\sin\alpha_1 = h_2,\ \ r_1\cos\alpha_3 + r_3\cos\alpha_1 = r_2,\ \ A = h_2 r_2,$$

which makes it possible to rewrite (5.3) as

$$A\rho_2 = 2r_2^3. \tag{5.4}$$

For orbits z_1, z_2, z_3 with $z_2 \in U$, the area A remains constant. Move points z_1 and z_3 slightly along the line z_1z_3 in such a way that the midpoint of the segment z_1z_3 remains the same. We obtain a family of orbits in U for which ρ_2 does not change but r_2 does, in contradiction with (5.4). This completes our first proof.

The second argument is a variation on the proof given in [2]. Let M^5 be the manifold of nondegenerate triangles whose oriented areas equal 1 and let $\mathscr{F}^3$ be the dual Birkhoff distribution on M. If there is an open set of fixed points of the map $T_3 \circ T_2 \circ T_1$ then one has a 2-dimensional disc $U \subset M$ tangent to $\mathscr{F}$. We will show that this is impossible.

If such a disc U exists, let x and y be local coordinates in it. Then $u = \partial/\partial x$ and $v = \partial/\partial y$ are commuting vector fields on U.

As a basis of vector fields in M generating $\mathscr{F}$ we can choose the fields

$$W_1 = (z_2 - z_1)\partial z_2 + (z_1 - z_3)\partial z_3,\ \ W_2 = (z_3 - z_2)\partial z_3 + (z_2 - z_1)\partial z_1$$

and

$$W_3 = (z_1 - z_3)\partial z_1 + (z_3 - z_2)\partial z_2,$$

slight modifications of the ones introduced in Section 2. Then

$$u = f_1W_1 + f_2W_2 + f_3W_3,\ \ v = g_1W_1 + g_2W_2 + g_3W_3,$$

where f_i, g_i are smooth functions on U. Without loss of generality, $f_1 \neq 0$ and $g_2 \neq 0$. Then we take a linear combination of u and v, with coefficients that are smooth functions, to obtain new fields

$$\bar{u} = W_1 + fW_3,\ \ \bar{v} = W_2 + gW_3,$$

so $[\bar{u}, \bar{v}] \in \operatorname{Span}(\bar{u}, \bar{v})$. In particular, $[\bar{u}, \bar{v}]$ is tangent to $\mathscr{F}$.

It follows that $d\alpha_i(\bar{u}, \bar{v}) = 0$ for $i = 1, 2, 3$ where α_i are the 1-forms introduced in Section 2. It is straightforward to compute that

$$d\alpha_2(\bar{u}, \bar{v}) = -2(1+g),\ \ d\alpha_3(\bar{u}, \bar{v}) = 2(1+f),$$

and hence $f = g = -1$. Thus $\bar{u} = W_1 - W_3$, $\bar{v} = W_2 - W_3$.

Now one computes the commutator:

$$[W_1 - W_3, W_2 - W_3] = [W_1, W_2] + [W_2, W_3] + [W_3, W_1] = 4(W_1 + W_2 + W_3)$$

(we leave this straightforward but a little tedious computation to the reader). Since the fields W_1, W_2, W_3 are everywhere linearly independent, the vector $W_1 + W_2 + W_3$ is not a linear combination of $W_1 - W_3$ and $W_2 - W_3$. This is a contradiction proving that such a disc U does not exist. ☐

REMARK 5.6. The reason we consider n outer billiards in Conjecture 5.1, instead of just one, is that this makes it possible to relax the convexity assumptions and include outer billiard tables that are convex outward (similarly to the case of inner billiards where one does not assume the billiard table to be convex).

In fact, the proof of Theorem 5.3 considerably simplifies if one considers a single strictly convex outer billiard table. Namely, let z_1 be a 3-periodic point of an outer billiard map T. Then the differential $dT^3(z_1)$ cannot be equal to the identity.

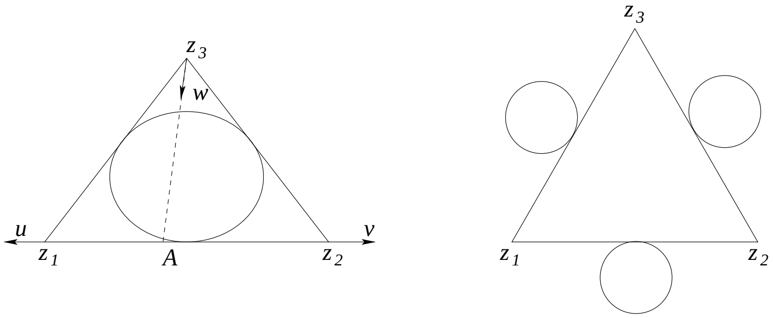

FIGURE 8. Can $dT^3 = \mathrm{Id}$?

Indeed, let $T(z_1) = z_2, T(z_2) = z_3, T(z_3) = z_1$, see Figure 8. Assume $dT^3(z_1) = \mathrm{Id}$. Let u and v be the opposite unit tangent vectors at points z_1 and z_2 having the direction of the side z_1z_2. Then $dT(u) = v$. Let $w = dT(v) = dT^{-1}(u)$. Since T is an orientation-preserving transformation, the vector w lies on the same sides of the lines z_2z_3 and z_1z_3 as the outer billiard table D. Thus w lies inside the wedge $z_1z_3z_2$. Let A be the intersection point of the line through point z_3 in the direction of vector w with the line z_1z_2. It follows from the area construction (see the paragraph after Proposition 5.5) that the outer billiard curve is second-order tangent to the hyperbola inscribed in the angle z_1Az_3 and tangent to the segment z_1z_3 at its midpoint. This hyperbola is convex outward, contradicting the convexity of D.

On the other hand, $d(T_2 \circ T_1 \circ T_3)$ may be equal to the identity at a fixed point of the map $T_2 \circ T_1 \circ T_3$. An example is shown in Figure 8 on the right: $z_1z_2z_3$ is a unit equilateral triangle, and the three outer billiard tables are circles of radii $1/\sqrt{3}$. Then equation (5.4) holds, and $d(T_2 \circ T_1 \circ T_3)(z_1) = \mathrm{Id}$.

Acknowledgments. We are grateful to Yu. Baryshnikov, S. Lvovski, R. Montgomery, K. F. Siburg, V. Zharnitsky for stimulating discussions and to the referees for their numerous suggestions that improved the exposition. The second author is indebted to Max-Planck-Institut in Bonn for its invariable hospitality.

REFERENCES

[1] Yu. Baryshnikov, V. Zharnitsky, *Billiards and nonholonomic distributions*, Zap. Nauchn. Sem. POMI **300** (2003), 56–64; translation in J. Math. Sci., **128** (2005), 2706–2710.

[2] Yu. Baryshnikov, V. Zharnitsky, *Sub-Riemannian geometry and periodic orbits in classical billiards*, Math. Res. Lett., **13** (2006), 587–598.

[3] C. Culter, *Asymptotics of polygonal dual billiards*, in preparation.

[4] F. Dogru, S. Tabachnikov, *Dual billiards*, Math. Intelligencer, **27** (2005), no. 4, 18–25.

[5] H. Geiges, *Contact geometry*, Handbook of Differential Geometry, v. 2, 315–382, Elsevier, Amsterdam, 2006.

[6] D. Genin, *Hyperbolic outer billiards: a first example*, Nonlinearity, **19** (2006), 1403–1413.

[7] E. Gutkin, A. Katok, *Caustics for inner and outer billiards*, Comm. Math. Phys., **173** (1995), 101–133.

[8] L. Hsu, *Calculus of variations via the Griffiths formalism*, J. Diff. Geom., **36** (1992), 551–589.

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

[10] O. Knill, *On nonconvex caustics of convex billiards*, Elem. Math., **53** (1998), 89–106.

[11] J. Landsberg, *Exterior differential systems and billiards*, Geometry, integrability and quantization, 35–54, Softex, Sofia, 2006.

[12] R. Montgomery, *A tour of subriemannian geometries, their geodesics and applications*, Amer. Math. Soc., Providence, RI, 2002.

[13] M. Rychlik, *Periodic points of the billiard ball map in a convex domain*, J. Diff. Geom., **30** (1989), 191–205.

[14] L. Stojanov, *Note on the periodic points of the billiard*, J. Diff. Geom., **34** (1991), 835–837.

[15] S. Tabachnikov, *Billiards*, Panor. Synth. **1**, Soc. Math. France, 1995.

[16] S. Tabachnikov, *Geometry and billiards*, Amer. Math. Soc., Providence, RI, 2005.

[17] A. Tumanov, V. Zharnitsky, *Periodic orbits in outer billiard*, preprint math.DS/0608529.

[18] Ya. Vorobets, *On the measure of the set of periodic points of a billiard*, Math. Notes, **55** (1994), 455–460.

[19] M. Wojtkowski, *Two applications of Jacobi fields to the billiard ball problem*, J. Diff. Geom., **40** (1994), 155–164.

DANIEL GENIN `<genin@math.psu.edu>`: Department of Mathematics, Penn State University, University Park, PA 16802, USA

SERGE TABACHNIKOV `<tabachni@math.psu.edu>`: Department of Mathematics, Penn State University, University Park, PA 16802, USA