Periodic billiard paths in right triangles are unstable

W. Patrick Hooper

January 24, 2007

A *billiard path* $\widehat{\gamma}$ in a triangle T is a bi-infinite path consisting of line segments in the interior of T joined by points on the triangle's edges so that the angle of incidence is equal to the angle of reflection. Billiard paths must never intersect the vertices of T. A *periodic billiard path* is a billiard path which closes up. The *orbit type* of a periodic billiard path is the sequence of edges it hits.

We parameterize the space, Δ, of triangles up to similarity by their angles. A periodic billiard path $\widehat{\gamma}$ in a triangle T is called *stable* if there is a neighborhood of T, $U \subset \Delta$, consisting of triangles that have periodic billiard paths with the same orbit type as $\widehat{\gamma}$. If a billiard path is not stable, it is called *unstable*.

The purpose of this paper is to prove the following theorem:

Theorem 1 *Every periodic billiard path in a right triangle is unstable.*

Previously, this theorem was known for a countable discrete set of right triangles, see [VGS91]. Also, it was known that certain types of periodic billiard paths in right triangles were unstable, see [GZ03]. Troubetzkoy proved that periodic billiard paths in right triangles are numerous. He showed, for instance, that in any right triangle T and for all but countably many points $P \in T$, there is a dense set of directions θ for which a billiard path starting at P and traveling in in the direction θ is periodic (see [Tro05a] and [Tro05b]). Some of the ideas in [Tro05a] are also reflected in this paper.

The study of which triangles admit stable periodic billiard paths is motivated by the open question, "Does every triangle have a periodic billiard path?" For triangles T with no non-trivial rational linear relations in the angles, all periodic billiard paths are stable (see [Tab95], Cor. 3.3.2).

Experimental evidence suggested this theorem should be true. Rich Schwartz and the author have designed a computer program, *McBilliards*[1], to investigate periodic billiard paths in triangles.

The author would like to thank Yair Minsky and Rich Schwartz for helpful conversations involving this work. The author is grateful to the anonymous referee, who suggested many improvements.

1 Group theory and stable paths

Given any triangle T we construct a locally Euclidean structure, $\mathcal{D}(T)$, on the thrice punctured sphere. This is the double of T across its boundary with the vertices removed. Geometrically, these vertices are cone singularities. See figure 1.

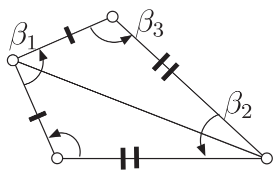

Figure 1: $\mathcal{D}(T)$ is constructed from two copies of the triangle T. The β_i are curves which travel around the punctures.

There is a canonical folding map $\mathcal{D}(T) \to T$ which is $2-1$ except on the edges of T. The billiard flow on the unit tangent bundle of T pulls back to the Euclidean geodesic flow on the the unit tangent bundle of $\mathcal{D}(T)$. Thus, a periodic billiard path $\widehat{\gamma}$ in T has a pull back to a closed Euclidean geodesic γ in $\mathcal{D}(T)$.

The locally Euclidean structure $\mathcal{D}(T)$ gives rise to the *holonomy representation* from the fundamental group to the orientation preserving isometry group of the plane,

$$hol : \pi_1(\mathcal{D}(T)) \to Isom_+(\mathbb{R}^2) \tag{1}$$

Let us give a brief description of the holonomy representation. Note that $\mathcal{D}(T)$ is locally isometric to $\mathbb{R}^2$. Thus in particular, any simply connected

[1] McBilliards is available from `http://mcbilliards.sourceforge.net/`.

open set U immersed in $\mathcal{D}(T)$ can be immersed isometrically into $\mathbb{R}^2$. In case U is in the universal cover, $\widetilde{\mathcal{D}(T)}$, this immersion into $\mathbb{R}^2$ is called the *developing map*, $dev : \widetilde{\mathcal{D}(T)} \to \mathbb{R}^2$. Covering space theory tells us that there is a canonical isomorphism ψ between the fundamental group, $\pi_1(\mathcal{D}(T))$, and the automorphisms of the universal cover. Thus for every $a \in \pi_1(\mathcal{D}(T))$ we get an automorphism of the universal cover $\psi(a) : \widetilde{\mathcal{D}(T)} \to \widetilde{\mathcal{D}(T)}$. Furthermore, there is a unique isometry of $\mathbb{R}^2$, $hol(a)$, defined so that the following diagram commutes:

$$\begin{array}{ccc} \widetilde{\mathcal{D}(T)} & \xrightarrow{\psi(a)} & \widetilde{\mathcal{D}(T)} \\ \Big\downarrow{\scriptstyle dev} & & \Big\downarrow{\scriptstyle dev} \\ \mathbb{R}^2 & \xrightarrow{hol(a)} & \mathbb{R}^2 \end{array} \tag{2}$$

For a more detailed description of the holonomy representation see section 3.4 of [Thu97].

The definition of the holonomy representation, depends on the choice of a base point P for universal cover, and a choice of an isometry from a small neighborhood of P to $\mathbb{R}^2$. Different choices will result in holonomy representations that differ by conjugation in $Isom_+(\mathbb{R}^2)$. Fortunately, we will be only interested in properties of $\pi_1(\mathcal{D}(T))$ which are invariant under conjugation, so we can ignore issues which arise from making such choices.

Now suppose that γ is a closed geodesic on $\mathcal{D}(T)$. A lift of γ to the universal cover is a geodesic $\widetilde{\gamma}$, so that $dev(\widetilde{\gamma})$ is a line. The holonomy around γ must translate along this line. Thus, we see that $hol(\gamma)$ must be a translation. We use this idea to generate a necessary algebraic condition for stability of a periodic billiard path.

Lemma 2 *If periodic billiard path $\widehat{\gamma}$ in a triangle T is stable then γ is null-homologous in $\mathcal{D}(T)$.*

Proof: We have the following commutative diagram:

$$\begin{array}{ccc} \pi_1(\mathcal{D}(T)) & \xrightarrow{hol} & Isom_+(\mathbb{R}^2) \\ \Big\downarrow & & \Big\downarrow{\scriptstyle \phi} \\ H_1(\mathcal{D}(T), \mathbb{Z}) & \xrightarrow{hol_{\mathrm{ab}}} & S^1 \end{array} \tag{3}$$

The down arrows correspond to abelianization. To see the abelianization of $Isom_+(\mathbb{R}^2)$ is the circle S^1, realize that the commutator subgroup of

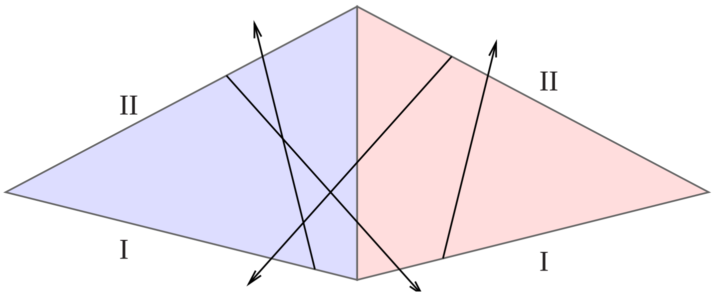

Figure 2: There are periodic billiard paths with orbit type $123\overline{123}$ in every acute triangle. This figure depicts such a billiard path with period 6 lifted to $\mathcal{D}(T)$. Note that by lemma 2, this curve is null-homologous in $\mathcal{D}(T)$. This path is closely related to the famous Fagnano curve, which is a billiard path of period 3 with the same orbit type.

$Isom_+(\mathbb{R}^2)$ is the group of translations. For any periodic billiard path $\widehat{\gamma}$ the isometry $hol(\gamma)$ is a translation, so $\gamma \in ker(\phi \circ hol)$.

If $\widehat{\gamma}$ is stable, then by definition we can find an open set U of triangles that have periodic billiard paths with the same orbit type as $\widehat{\gamma}$. In particular, therefore, we can choose $T' \in U$ so that there are no non-trivial rational linear relations between the angles of T'. That is, if α_i are the angles of T' then for all $q_i \in \mathbb{Q}$,

$$q_1\alpha_1 + q_2\alpha_2 + q_3\alpha_3 = 0 \text{ implies } q_i = 0 \text{ for all } i \tag{4}$$

The homology group of $\mathcal{D}(T')$ is generated by two loops β_1 and β_2 shown in figure 1. $hol_{\text{ab}}(\beta_1)$ and $hol_{\text{ab}}(\beta_2)$ correspond to rotations by angles $2\alpha_1$ and $2\alpha_2$ respectively. There is no non-trivial rational linear relation between α_1, α_2, and π. Thus $hol_{\text{ab}} : H_1(\mathcal{D}(T'), \mathbb{Z}) \to S^1$ is injective. By the commutativity of diagram 3 and because $\gamma' \in ker(\phi \circ hol)$, we know that $\widehat{\gamma}'$ must be null-homologous. Finally, because $\widehat{\gamma}$ and $\widehat{\gamma}'$ have the same orbit type, γ is also null-homologous. $\diamondsuit$

It should be noted that the converse of the lemma is also true, but we will not need it. If $\widehat{\gamma}$ is a periodic billiard path and γ is null-homologous then $\widehat{\gamma}$ is stable.

2 The minimal translation surface

The *universal abelian cover* of $\mathcal{D}(T)$ is defined to be

$$AC_\Delta = \widetilde{\mathcal{D}(T)}/[\pi_1(\mathcal{D}(T)), \pi_1(\mathcal{D}(T))] \tag{5}$$

Here $\widetilde{\mathcal{D}(T)}$ is the universal cover, and $[\pi_1(\mathcal{D}(T)), \pi_1(\mathcal{D}(T))] \subset \pi_1(\mathcal{D}(T))$ is the commutator subgroup acting by automorphisms of the universal cover. The automorphism group of the covering $AC_\Delta \to \mathcal{D}(T)$ is $H_1(\mathcal{D}(T), \mathbb{Z}) \cong \mathbb{Z}^2$.

Definition 1 *The minimal translation surface corresponding to T is*

$$MT(T) = AC_\Delta / ker(hol_{ab})$$

where $hol_{ab} : H_1(\mathcal{D}(T), \mathbb{Z}) \to S^1$ as in diagram 3.

MT(T) inherits a Euclidean structure from its covering of $\mathcal{D}(T)$. By construction, it is clear that the image of the holonomy representation

$$hol : \pi_1(\mathrm{MT}(T)) \to Isom_+(\mathbb{R}^2)$$

consists entirely of translations. $\mathbb{R}^2$ has a natural translation invariant trivialization of its unit tangent bundle. (We identify the direction a vector points by it's angle in $\mathbb{R}/2\pi$. This identification is invariant under translations.) Because $hol(\pi_1(\mathrm{MT}(T)))$ is a group of translations, the trivialization pulls back to a trivialization of the unit tangent bundle of MT(T). Thus, closed Euclidean geodesics in MT(T) are simple. In particular, the closed geodesic $\gamma \subset \mathcal{D}(T)$ lifts to a simple closed geodesic $\widetilde{\gamma}$ in MT(T).

We can also construct MT(T) more geometrically. Let $G = \mathbb{Z}_2 * \mathbb{Z}_2 * \mathbb{Z}_2$. Consider the representation $\rho : G \to Isom(\mathbb{R}^2)$ generated by reflections in the sides of T. The orbit of T under G is

$$G(T) = \{\rho(g)(T) | g \in G\} \tag{6}$$

Definition 2 *The minimal translation surface of T is the quotient of edge and triangle identifications.*

$$MT(T) = G(T)/\sim_1/\sim_2$$

Identify to triangles in $G(T)$ along an edge if they differ by reflection in that edge. Then identify any two triangles in $G(T)$ which differ by a translation. The result is the minimal translation surface $MT(T)$.

The minimal translation surface comes up in other natural ways as well. The geodesic flow on the unit tangent bundle of $\mathcal{D}(T)$ has invariant subsurfaces isometric to $\mathrm{MT}(T)$. Thus, this surface is often called the *invariant surface*. See [ZK75] and [Tab95] for more on this viewpoint. This surface is also used in [Tro05a] and [Tro05b].

It is enough to prove theorem 1 in the case of a right triangle T whose non-right angles are not rational multiples of π. In this case we can explicitly describe $\mathrm{MT}(T)$. $\mathrm{MT}(T)$ is a union of rhombi $\{R_i | i \in \mathbb{Z}\}$ where each rhombus R_i is glued to R_{i-1} and R_{i+1}. See figure 3. In particular, notice that opposite edges of R_i are glued to the same neighboring rhombus.

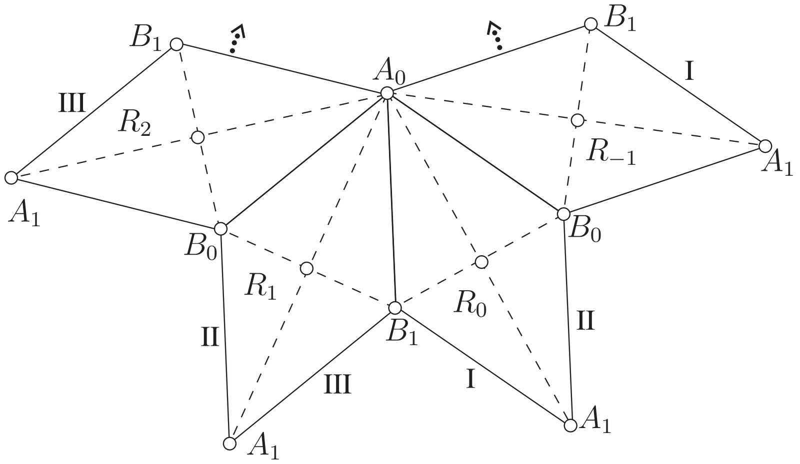

Figure 3: $\mathrm{MT}(T)$ for a generic right triangle T. Roman numerals indicate edge identifications.

Formally, the rhombi R_i have punctures in their centers. The centers of each rhombus are lifts of the vertex of $\mathcal{D}(T)$ with cone angle π. Since, the cone points of $\mathcal{D}(T)$ were removed, the center of each rhombus is punctured. Quite clearly, the punctures at the centers of the rhombi are removable singularities, but we will not remove them. Their presence allows us to compute topological invariants which will allow us to derive a contradiction from lemma 2. In addition there are 4 points of the surface where infinite branching occurs. These points are marked A_0, A_1, B_0, and B_1 in figure 3.

The automorphism group of the cover $\mathrm{MT}(T) \to \mathcal{D}(T)$ is isomorphic to $\mathbb{Z} \times \mathbb{Z}_2$. The unique involution i in the automorphism group acts by a simultaneous rotation by π about the puncture in the center of each rhombus R_i. This involution swaps A_0 with A_1 and B_0 with B_1.

Suppose $\widehat{\gamma}$ is a periodic billiard path in T, and γ is the corresponding closed geodesic on $\mathcal{D}(T)$. Since $hol(\gamma)$ is a translation, γ lifts to a simple closed geodesic $\widetilde{\gamma}$ in $\mathrm{MT}(T)$. The following result is the key to theorem 1.

Theorem 3 *The two closed curves $\widetilde{\gamma}$ and $i(\widetilde{\gamma})$ cut the surface $MT(T)$ into two pieces. One component is a twice punctured standard Euclidean cylinder whose boundary consists of these two curves.*

The punctures in the cylinder must come from centers of the rhombi in $\mathrm{MT}(T)$. We break this result into a series of shorter arguments. First:

Proposition 4 *$\widetilde{\gamma}$ and $i(\widetilde{\gamma})$ are disjoint.*

Proof: First of all, i acts by a rotation by π on directions in $\mathrm{MT}(T)$. Therefore $\widetilde{\gamma}$ and $i(\widetilde{\gamma})$ travel in opposite directions. We can reverse the orientation of $i(\widetilde{\gamma})$ to obtain a curve $-i(\widetilde{\gamma})$ which travels in the same direction as $\widetilde{\gamma}$. Thus, if they intersect, $\widetilde{\gamma} = -i(\widetilde{\gamma})$. Supposing they are equal, the action of i restricted to $\widetilde{\gamma}$ is an orientation reversing isometry of the circle. Such a map has two fixed points. But i is a non-trivial automorphism of $\mathrm{MT}(T) \to \mathcal{D}(T)$, so it has no fixed points (the centers of the rhombi are not in $\mathrm{MT}(T)$.) $\diamondsuit$

Proposition 5 *For each $n \in \mathbb{Z}$, $\widetilde{\gamma}$ intersects opposite edges of R_n an equal number of times.*

Proof: Fix an orientation for $\widetilde{\gamma}$. This gives us a direction $\vec{v}$ in $\mathrm{MT}(T)$, which dictates the direction $\widetilde{\gamma}$ travels. Fix a rhombus R_n. Note that, the edges of R_n are glued alternately to R_{n-1} and R_{n+1}. Furthermore, the only way to get from R_{n-1} to R_{n+1} is by traveling through R_n. A line l traveling through R_n in the direction of $\vec{v}$ which does not intersect a vertex must do one of three things.

1. l passes through opposite edges of the rhombus. In fact, there is at most one pair of opposite edges for which this is possible. In this case l travels from R_{n+u} through R_n and back into R_{n+u}, where u is either 1 or -1.

2. l passes from R_{n-1} through R_n to R_{n+1}. There is only one pair of edges for which this is possible.

3. l passes from R_{n+1} through R_n to R_{n-1}. In this case the two edges l passes through are opposite those discussed in item 2.

See figure 4. Our proposition boils down to the fact that for each R_n, intersections of types 2 and 3 between $\widetilde{\gamma}$ and R_n occur an equal number of times. $\widetilde{\gamma}$ gives rise to a closed walk on the integers by looking at the sequence of indices of rhombi $\widetilde{\gamma}$ hits. In order for a walk on the integers to close up, each time you walk upward from $n-1$ through n and up to $n+1$, you must later walk downward traveling from $n+1$ through n to $n-1$. Thus intersections of types 2 and 3 must occur equally often. $\diamondsuit$

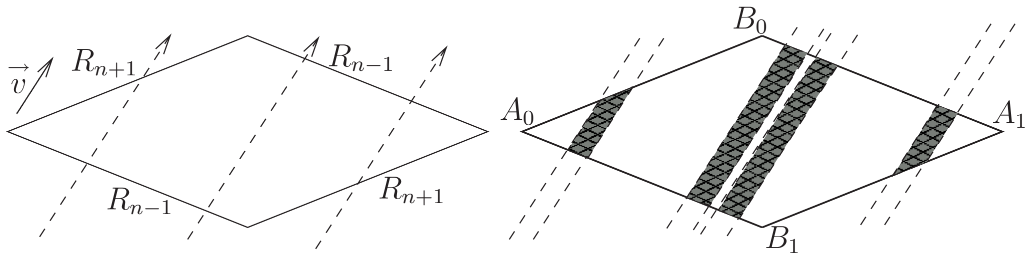

Figure 4: Left: Possible intersections types of lines traveling through R_n in direction $\vec{v}$. Right: The coloring used in the proof of theorem 3.

Proof of Theorem 3: Proposition 5 tells us that $\widetilde{\gamma} \cup i(\widetilde{\gamma})$ intersects each edge of R_n an even number of times, since i acts by a rotation by π.

On each rhombus R_n, we color the pieces of $R_n \backslash (\widetilde{\gamma} \cup i(\widetilde{\gamma}))$ black and white. Color the piece containing the vertex A_0 white. Then, we follow the coloring rule that every piece adjacent to a white piece is black and vice versa. The fact that $\widetilde{\gamma} \cup i(\widetilde{\gamma})$ intersects edge $\overline{A_0 B_0}$ an even number of times tells us that the piece containing B_0 must also be colored white. Similarly, it follows that the pieces containing B_1 and A_1 must also be colored white. See figure 4 for an illustration of the coloring. Also, the coloring is compatible between each adjacent pair of rhombi. That is, an edge of a rhombus is shared by two rhombi. The colorings of the edge induced by the first rhombus and the second rhombus are identical.

This coloring shows that $\mathrm{MT}(T) \setminus (\widetilde{\gamma} \cup i(\widetilde{\gamma}))$ consists of at least two components. Neighborhoods of the four infinite branch points A_0, A_1, B_0 and B_1 are colored white. $\mathrm{MT}(T)$ has no other branch points. Thus the black region is an oriented Euclidean surface with geodesic boundary (and possibly punctures coming from the centers of rhombi). The only surface satisfying this requirement is a cylinder.

Call the cylinder $\mathcal{C}$. By construction, the involution i preserves $\mathcal{C}$ and acts as an orientation preserving isometry. i swaps the boundaries of the cylinder, $\widetilde{\gamma}$ and $i(\widetilde{\gamma})$. A Euclidean isometry of a cylinder which swaps the boundary components, must have exactly two fixed points. Since i is a deck transformation of $\mathrm{MT}(T)$, it has no fixed points. However i does preserve the punctures of $\mathrm{MT}(T)$. Thus $\mathcal{C}$ must have exactly two punctures. $\diamondsuit$

Proof of Theorem 1: Let $\mathcal{C}$ be the cylindrical component of $\mathrm{MT}(T) \setminus (\widetilde{\gamma} \cup i(\widetilde{\gamma}))$ as above. Given appropriately oriented curves p_1 and p_2 traveling around the punctures of $\mathcal{C}$, in $H_1(\mathrm{MT}(T), \mathbb{Z})$,

$$[\![\widetilde{\gamma}]\!] + [\![i(\widetilde{\gamma})]\!] + [\![p_1]\!] + [\![p_2]\!] = 0 \tag{7}$$

because these four curves bound a subsurface of $\mathrm{MT}(T)$. Using the covering map $\mathrm{MT} \to \mathcal{D}$ we can push these four curves into $\mathcal{D}$ to obtain a relation in $H_1(\mathcal{D}, \mathbb{Z})$. Both $\widetilde{\gamma}$ and $i(\widetilde{\gamma})$ are sent to γ because i is a deck transformation of the covering. Similarly, there is a deck transformation taking p_1 to p_2. Both p_1 and p_2 are sent to the curve p which travel twice around the π-cone singularity of $\mathcal{D}$. Thus in $H_1(\mathcal{D}, \mathbb{Z})$,

$$[\![\gamma]\!] = -[\![p]\!] \neq 0 \tag{8}$$

Thus lemma 2 entails the theorem. $\diamondsuit$

Billiards in an isosceles triangle T' are closely related to billiards in right triangles. In particular, T' can be folded in half to obtain a right triangle T. Thus a periodic billiard path in a right triangle lifts to a periodic billiard path in an isosceles triangle.

Using definition 2 of the minimal translation surface, we see $\mathrm{MT}(T')$ contains $\mathrm{MT}(T)$ and can be obtained by filling in the punctures in the centers of the rhombi making up $\mathrm{MT}(T)$. The deck transformation i of $\mathrm{MT}(T)$ induces a self-isometry i' of $\mathrm{MT}(T')$ which fixes the centers of each rhombus.

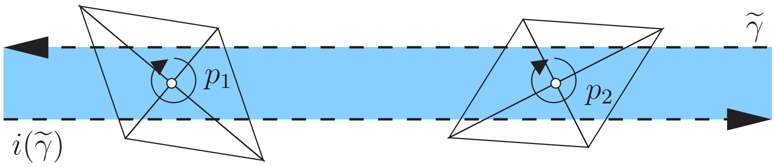

Figure 5: A characterization of the cylinder $\mathcal{C} \subset \mathrm{MT}(T)$ together with its punctures.

Let M be the midpoint of the edge preserved by the reflective symmetry of the isosceles triangle T'.

Corollary 6 *Assume T' is an isosceles triangle with no non-trivial rational linear relations among its two distinct angles. Let $\widehat{\gamma}$ be a periodic billiard path in T'. There exists a periodic billiard path $\widehat{\gamma}'$ in T' with the same orbit type as $\widehat{\gamma}$ and which passes through M twice in its period.*

Proof: We can lift $\widehat{\gamma}$ to a simple closed curve $\widetilde{\gamma} \subset \mathrm{MT}(T')$. If i' fixes a point of $\widetilde{\gamma}$, it must fix exactly two points, because it is an orientation reversing isometry of $\widetilde{\gamma}$. The fixed point set of i' in $\mathrm{MT}(T')$ is the set of lifts of M. Thus $\widehat{\gamma}$ hits M twice in its period.

Otherwise the folding map $T' \to T$ projects $\widehat{\gamma}$ to a periodic billiard path in the right triangle T. The core curve of the cylinder constructed in theorem 3 passes through two punctures of $\mathrm{MT}(T)$. Inside $\mathrm{MT}(T')$ these punctures have been replaced by lifts of M. The image of this core curve under the folding map $\mathrm{MT}(T') \to T'$ is a periodic billiard path hitting M twice. $\diamondsuit$

Galperin and Zvonkine called trajectories of the type that appear in Corollary 6 *mirror trajectories*. They showed that mirror trajectories are all unstable ([GZ03], theorem 3A). In fact our main result, theorem 1, follows from this theorem of Galperin and Zvonkine together with our Corollary 6.

References

[GZ03] G. Galperin and D. Zvonkine, *Periodic billiard trajectories in right triangles and right-angled tetrahedra*, Regul. Chaotic Dyn. **8** (2003), no. 1, 29–44. MR 1963966 (2004b:37081)

[Tab95] Serge Tabachnikov, *Billiards*, Panor. Synth. (1995), no. 1, vi+142. MR 1328336 (96c:58134)

[Thu97] William P. Thurston, *Three-dimensional geometry and topology. Vol. 1*, Princeton Mathematical Series, vol. 35, Princeton University Press, Princeton, NJ, 1997, Edited by Silvio Levy. MR 1435975 (97m:57016)

[Tro05a] Serge Troubetzkoy, *Periodic billiard orbits in right triangles*, Ann. Inst. Fourier (Grenoble) **55** (2005), no. 1, 29–46. MR 2141287

[Tro05b] ———, *Periodic billiard orbits in right triangles II*, arXiv:math.DS/0505637, 2005.

[VGS91] Ya. B. Vorobets, G. A. Galperin, and A. M. Stëpin, *Periodic billiard trajectories in polygons*, Uspekhi Mat. Nauk **46** (1991), no. 5(281), 165–166. MR 1160340 (93b:58119)

[ZK75] A. N. Zemljakov and A. B. Katok, *Topological transitivity of billiards in polygons*, Mat. Zametki **18** (1975), no. 2, 291–300. MR 0399423 (53 #3267)