

## Durham Research Online

---

**Deposited in DRO:**

31 October 2018

**Version of attached file:**

Accepted Version

**Peer-review status of attached file:**

Peer-reviewed

**Citation for published item:**

Parker, John R. and Peyerimhoff, Norbert and Siburg, Karl Friedrich (2018) 'Minimizing length of billiard trajectories in hyperbolic polygons.', *Conformal geometry and dynamics.*, 22 . pp. 315-332.

**Further information on publisher's website:**

<https://doi.org/10.1090/ecgd/328>

**Publisher's copyright statement:**

Accepted manuscript available under a CC-BY-NC-ND licence.

**Additional information:**

---

**Use policy**

---

The full-text may be used and/or reproduced, and given to third parties in any format or medium, without prior permission or charge, for personal research or study, educational, or not-for-profit purposes provided that:

- a full bibliographic reference is made to the original source
- a [link](#) is made to the metadata record in DRO
- the full-text is not changed in any way

The full-text must not be sold in any format or medium without the formal permission of the copyright holders.

Please consult the [full DRO policy](#) for further details.

---

## MINIMIZING LENGTH OF BILLIARD TRAJECTORIES IN HYPERBOLIC POLYGONS

JOHN R. PARKER, NORBERT PEYERIMHOFF, AND KARL FRIEDRICH SIBURG

ABSTRACT. Closed billiard trajectories in a polygon in the hyperbolic plane can be coded by the order in which they hit the sides of the polygon. In this paper, we consider the average length of cyclically related closed billiard trajectories in ideal hyperbolic polygons and prove the conjecture that this average length is minimized for regular hyperbolic polygons. The proof uses a strict convexity property of the geodesic length function in Teichmüller space with respect to the Weil-Petersson metric, a fundamental result established by Wolpert.

### 1. INTRODUCTION

To play billiards in a Euclidean polygon, the rules are as follows: An infinitesimal ball travels along a straight line (geodesic) at constant speed, and when it hits a side of the polygon then it changes its direction so the angle of incidence agrees with the angle of reflection. The path followed by such a ball is called a *billiard trajectory*.

It also makes sense to play billiards in a hyperbolic polygon, as here we also have well-defined meanings of geodesics and angles of incidence and reflection. To our knowledge, the first instance where such a dynamical system was considered is in an article by E. Artin [1] written in German (see [4] for an English translation). Using continued fractions, he constructs dense bi-infinite billiard trajectories in half of the fundamental polygon of the modular surface. In fact, there are many striking connections between geodesics on the modular surface, their symbolic coding via cutting sequences and number theory such as binary quadratic forms and continued fractions (see [8] for a well known classical reference and also, e.g., [2] for very recent developments).

A billiard trajectory is said to be *closed* if after a finite time it returns to the same point with the same direction. A natural setting is to consider closed billiard trajectories in ideal hyperbolic polygons where all vertices lie on the boundary at infinity. A key piece of information of a billiard trajectory in such an ideal hyperbolic polygon is its *billiard sequence* obtained by recording the order of the sides where the ball hits the boundary. Unlike the Euclidean case, where there may be uncountably many closed billiard trajectories, although they are homotopic, with the same billiard sequence, in the hyperbolic case there is at most one (which is a consequence of the fact that the curvature is strictly negative). In a polygon with  $k$ -sides, there are  $k$  different *anti-clockwise enumerations* of the polygon's sides with

---

Received by the editors 30 October 2018.

2010 *Mathematics Subject Classification*. Primary 37D40, Secondary 32G15, 53A35, 37F30.

©XXXX American Mathematical Society

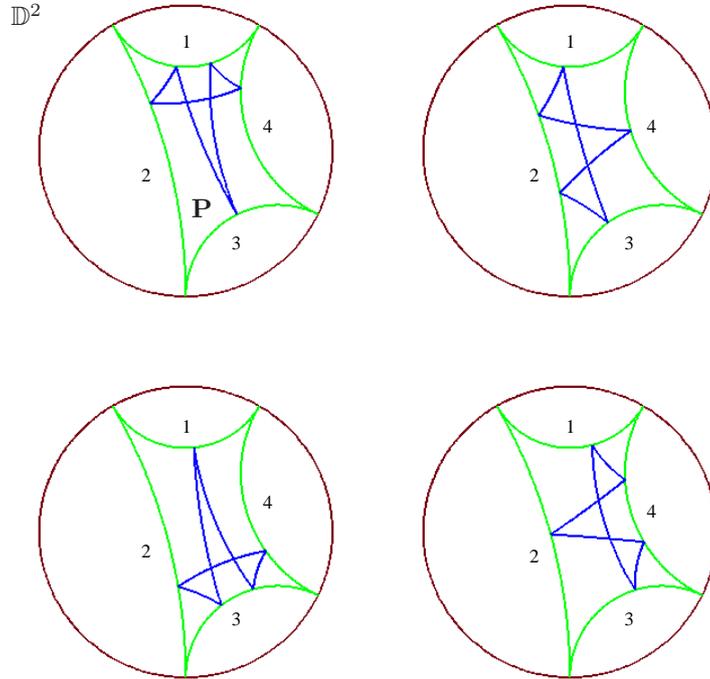


FIGURE 1. Illustration of all closed billiard trajectories in an ideal quadrilateral in the Poincaré disk  $\mathbb{D}^2$  which are cyclically related to  $(1, 2, 4, 1, 3)$ . Their billiard sequences are  $(2, 3, 1, 2, 4)$ ,  $(3, 4, 2, 3, 1)$  and  $(4, 1, 3, 4, 2)$ .

labels  $1, 2, \dots, k$ . For each such labelling of the polygon, a closed billiard trajectory  $\gamma$  gives rise to a finite sequence of these numbers, which we call the billiard sequence of  $\gamma$  with respect to this labelling. Two billiard trajectories in a given ideal polygon are said to be *cyclically related* if, under different anti-clockwise enumerations of the polygon's sides, they yield the same billiard sequences. Figure 1 illustrates this concept: Our hyperbolic polygon there is an ideal quadrilateral in the Poincaré unit disc  $\mathbb{D}^2$  and the original closed billiard trajectory has the finite billiard sequence  $(1, 2, 4, 1, 3)$ .

A closed billiard trajectory has a well-defined *hyperbolic length*. Given two different ideal hyperbolic  $k$ -gons with anti-clockwise labellings of their sides, we can compare billiard trajectories in both polygons with the same billiard sequences. For closed billiard trajectories with a given billiard sequence, it is interesting to ask for which polygons this length is minimised. A first conjecture may be that the minimising polygon is the regular polygon, i.e., the polygon whose symmetry group is the full dihedral group. However, it is easy to see that this is actually not

the case. Indeed, there are many billiard sequences for which we can find polygons whose corresponding trajectories have arbitrary small lengths. Therefore, we consider *families of all cyclically related* closed billiard trajectories in a polygon and their averaged lengths. A conjecture in [3] states for *ideal* hyperbolic polygons that the *average length function* of any family of cyclically related closed billiard trajectories is uniquely minimised by the regular polygon. Note that in a regular polygon, cyclically related billiard trajectories are just rotations of each other about the centre of the polygon and that all of them have the same length. The aim of this paper is to confirm this conjecture, that is, to prove the following result:

**Theorem 1.1.** *In anti-clockwise labelled ideal hyperbolic polygons with  $k \geq 3$  sides, the average length function of any family of cyclically related closed billiard trajectories corresponding to a given billiard sequence is uniquely minimised by the regular polygon.*

Let us roughly outline the strategy of proof: Firstly, we associate to every polygon a hyperbolic surface by gluing two oppositely oriented copies of the polygon pointwise along corresponding sides. Note that the surface is noncompact and has  $k$  cusps. We refer to this surface as a *billiard surface*. Then every even-sided closed billiard trajectory in the polygon lifts to a pair of closed geodesics of the same length in the corresponding surface and every odd-sided closed billiard trajectory lifts to one closed geodesic in the surface of twice the length of the original billiard trajectory. In short, billiard trajectories in the polygon correspond to geodesics in the billiard surface.

This allows us to rephrase the conjecture as a result on the length of closed geodesics in billiard surfaces. In order to apply powerful results of Teichmüller theory and the *Weil-Petersson metric*, we consider the billiard surfaces as points in *Teichmüller space* and we call the subspace of all billiard surfaces the *billiard space*. Specifically, we use the results of Wolpert that geodesic length functions are strictly convex with respect to the Weil-Petersson Metric of Teichmüller space, and of Kerckhoff that summed length functions of filling curves are proper.

Introducing a Weil-Petersson isometry of Teichmüller space (the so called *flip map*) which fixes the billiard space pointwise, we show that the billiard space is a geodesically convex subset of Teichmüller space with respect to the Weil-Petersson metric. The average length function of a family of cyclically related closed billiard trajectories corresponds to a geodesic length function of a filling family of closed curves in Teichmüller space. By the above mentioned results, the geodesic length function becomes minimal in a unique point in Teichmüller space and it only remains to show that this minimising point in Teichmüller space corresponds to the billiard surface associated to the regular polygon.

The following five sections of this paper follow essentially the arguments of proof described above. In Appendix A, we briefly discuss an analogous problem in the Euclidean setting: Here the polygons are rectangles of area one and the unique minimizing billiard table turns out to be the unit square.

**Acknowledgement:** The authors gratefully acknowledge the inspiring and helpful discussions with Andreas Knauf and Joan Porti concerning the strategy of proof. They also thank Andy Hayden for numerous general detailed discussions

concerning the topic and many aspects of the proof. NP greatly enjoyed the hospitality of the TU Dortmund and the Isaac Newton Institute, Cambridge, while he was working on certain parts of this article.

## 2. CYCLICALLY RELATED BILLIARD TRAJECTORIES ARE FILLING

In this section, we prove a particular property of the connected components of the complement of a union of all rotations of a closed billiard trajectory in a regular hyperbolic polygon. Let us first introduce this property for families of curves, which is called *filling*, in full detail. This definition for polygons is guided by the desire that the lift of a family of filling curves in the corresponding billiard surface (which will be introduced later) is also filling. Note that, for finite volume Riemann surfaces, a family of curves is called filling if each connected component of their complement is an open topological disc or a disc with one puncture corresponding to one of the cusps of the surface. Since the main result of this section (Proposition 2.4 below) holds both for compact and ideal regular polygons, we formulate it in this generality. Let us first introduce some basic notation.

**Definition 2.1.** *Let  $\mathbf{P}$  be a (closed) hyperbolic  $k$ -gon. Let  $x_1, \dots, x_k$  (which lie in  $\mathbb{D}^2$  or its boundary  $\partial\mathbb{D}^2$ ) denote the vertices of  $\mathbf{P}$  cyclically ordered anti-clockwise. Let  $[x_i, x_{i+1}]$  denote the (geodesic) side of  $\mathbf{P}$  with endpoints  $x_i$  and  $x_{i+1}$ , where indices are taken modulo  $k$ . We use the convention that each side contains both its endpoints.*

Henceforth, we only consider hyperbolic polygons  $\mathbf{P} \subset \mathbb{D}^2$  with interior angles equal to the fixed value  $\pi/l$  (compact case) or 0 (ideal case). Such a polygon gives rise to a tessellation of  $\mathbb{D}^2$  via repeated reflections and to a natural projection map from  $\mathbb{D}^2$  to  $\mathbf{P}$ . Then the projection of every oriented bi-infinite geodesic in  $\mathbb{D}^2$  can be viewed as a billiard trajectory in  $\mathbf{P}$ , as long as the geodesic in  $\mathbb{D}^2$  is not completely contained in the union of the boundaries of the tiles in this tessellation. Conversely, given a billiard trajectory in  $\mathbf{P}$  with a start point, it can be *unfolded* to a bi-infinite geodesic in  $\mathbb{D}^2$  by reflecting the billiard table along its sides instead of the trajectory, whenever it hits the boundary. Note that this viewpoint allows us to define billiard trajectories of  $\mathbf{P}$  even in the case when they hit the vertices  $x_i$  of  $\mathbf{P}$ . We also like to mention for the sake of simplicity that, if there is no danger of misinterpretation, we often do not distinguish between billiard trajectories and geodesics given as arc-length parametrized curves and their geometric representation as subsets of polygons or surfaces.

**Definition 2.2.** *An arc  $\alpha$  of a closed billiard trajectory in  $\mathbf{P}$  is a closed geodesic arc whose interior lies in  $\mathbf{P}$  and whose endpoints lie on  $\partial\mathbf{P}$ . Note that an endpoint of  $\alpha$  may be a vertex of  $\mathbf{P}$ , but such a vertex must lie in  $\mathbb{D}^2$ .*

Now we can introduce the concept of being *filling*. The different types of connected components in Definition 2.3 are illustrated in Figure 2.

**Definition 2.3.** *Let  $\mathbf{P}$  be a (closed) hyperbolic polygon and  $\gamma$  be a union of closed billiard trajectories in  $\mathbf{P}$ . We say that  $\gamma$  fills  $\mathbf{P}$  if  $\gamma$  is connected and each connected component of  $\mathbf{P} \setminus \gamma$  is a topological disc whose boundary is one of:*

- (a) a topological circle in  $\gamma$  made up of segments of geodesic arcs of  $\gamma$ ;
- (b) a topological arc in  $\gamma$  (made up of segments of geodesic arcs in  $\gamma$ ) and an arc of one side of  $\mathbf{P}$ , possibly including one or both vertices in this side; or

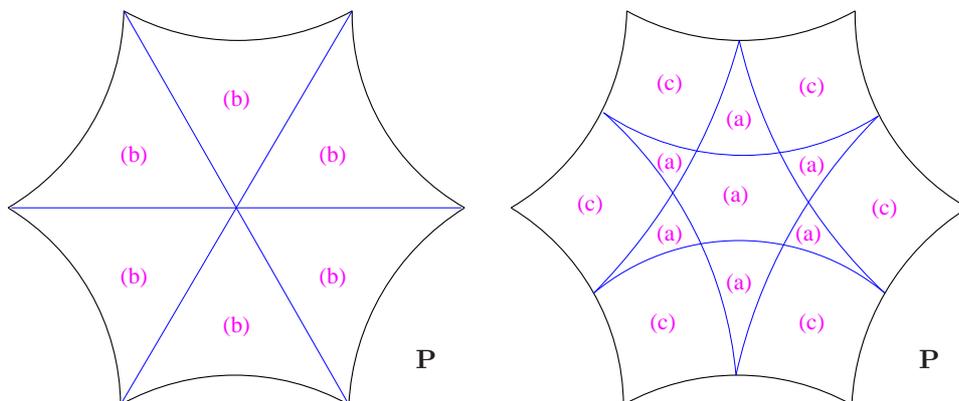


FIGURE 2. Examples of connected components of type (a), (b) and (c) in Definition 2.3 of  $\mathbf{P} \setminus \gamma$ .  $\gamma$  is the union of the blue arcs.

- (c) a topological arc in  $\gamma$ , exactly one vertex of  $\mathbf{P}$  and an arc in each of the two sides ending in this vertex, but not including either of the other vertices in these sides.

Now we state the main result of this section.

**Proposition 2.4.** *Let  $\mathbf{P}_0$  be a regular hyperbolic  $k$ -gon and let  $\rho$  denote the anti-clockwise rotation through angle  $2\pi/k$  about the centre of  $\mathbf{P}_0$ . Let  $\gamma_0$  be a closed billiard trajectory in  $\mathbf{P}_0$  and let  $\gamma_i = \rho^i(\gamma_0)$  for  $i = 1, \dots, k-1$ . Then  $\gamma = \bigcup_{i=0}^{k-1} \gamma_i$  fills  $\mathbf{P}_0$ .*

Note that the curves  $\gamma_i$  are the closed billiard trajectories cyclically related to  $\gamma_0$ . An important lemma for the proof is the following.

**Lemma 2.5.** *If  $\gamma_0$  is a closed billiard trajectory in a hyperbolic polygon  $\mathbf{P}$  then there are two non-adjacent sides of  $\mathbf{P}$  that intersect  $\gamma_0$  (not necessarily as endpoints of a single geodesic arc of the trajectory).*

*Proof.* We suppose the result is false. That is, suppose that  $\gamma_0$  is a closed billiard trajectory in  $\mathbf{P}$  and there are two sides  $[x_{i-1}, x_i]$  and  $[x_i, x_{i+1}]$  of  $\mathbf{P}$  so that that every arc of  $\gamma_0$  has one endpoint in  $[x_{i-1}, x_i]$  and the other endpoint in  $[x_i, x_{i+1}]$ . Note that  $\gamma_0$  cannot pass through  $x_i$  since a geodesic arc from  $x_i$  to a point in either of these two sides must be contained in this side. Moreover, if  $\gamma_0$  passes through  $x_{i-1}$  (or  $x_{i+1}$ ) then we could find an arc in  $\gamma_0$  connecting the non-adjacent sides  $[x_{i-2}, x_{i-1}]$  and  $[x_i, x_{i+1}]$  (or the non-adjacent sides  $[x_{i-1}, x_i]$  and  $[x_{i+1}, x_{i+2}]$  respectively). Let  $<$  denote the natural anticlockwise order on  $[x_{i-1}, x_i] \cup [x_i, x_{i+1}]$ .

There are finitely many intersection points of  $\gamma_0$  with  $\partial\mathbf{P}$ . Write them as  $y_j$  where  $-n \leq j \leq m$  and  $j \neq 0$  where

$$x_{i-1} < y_{-n} < y_{-n+1} < \dots < y_{-1} < x_i < y_1 < \dots < y_{m-1} < y_m < x_{i+1}.$$

Every geodesic arc in  $\gamma_0$  connects a point  $y_{-r}$  with negative index and a point  $y_s$  with positive index.

Consider  $y_{-n}$ . Suppose  $a$  is the largest index so that there is an arc of  $\gamma_0$  from  $y_{-n}$  to  $y_a$  (see Figure 3 for illustration). Then there is a point  $y_b$  with  $b \leq a$  so that the

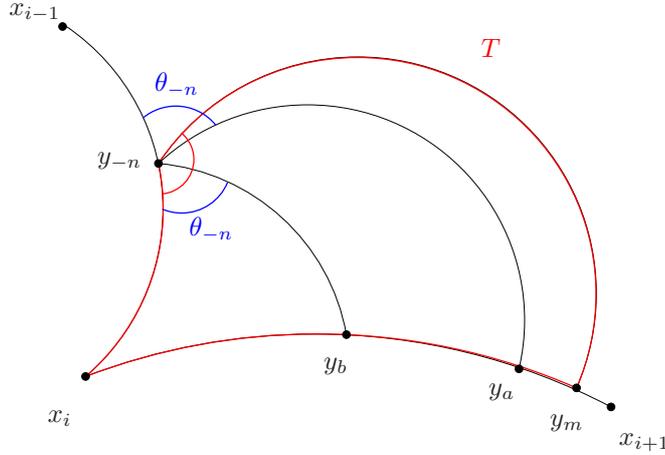


FIGURE 3. Note that  $\theta_{-n} \leq \pi/2$  and, therefore, the red internal angle of the triangle  $T$  at  $y_{-n}$  must be at least  $\pi/2$ .

arc  $[y_{-n}, y_b]$  is adjacent to  $[y_{-n}, y_a]$  in the billiard trajectory  $\gamma_0$ . (Note that there could be other arcs of  $\gamma_0$  with endpoint  $y_{-n}$ .) Since the angle of incidence equals the angle of reflection, the angle  $\theta_{-n}$  between the arcs  $[y_b, y_{-n}]$  and  $[y_{-n}, x_i]$  equals the angle between the arcs  $[y_a, y_{-n}]$  and  $[y_{-n}, x_{i-1}]$ . Since  $b \leq a$  then  $\theta_{-n} \leq \pi/2$ .

Similarly for  $y_m$ . Let  $-c$  be the smallest index so that there is an arc from  $y_m$  to  $y_{-c}$ . Then the angle  $\theta_m$  between the arcs  $[y_{-c}, y_m]$  and  $[y_m, x_{i+1}]$  is at most  $\pi/2$ .

Now consider the solid closed geodesic triangle  $T$  with vertices  $x_i$ ,  $y_m$  and  $y_{-n}$ . The entire billiard trajectory must be contained in  $T$ . The internal angle of  $T$  at  $y_{-n}$  is at least  $\pi - \theta_{-n}$  and the internal angle at  $y_m$  is at least  $\pi - \theta_m$ . But both these angles are at least  $\pi/2$ , which contradicts the fact that the sum of internal angles of a hyperbolic triangle are less than  $\pi$ .  $\square$

**Definition 2.6.** Let  $\alpha$  and  $\beta$  be two closed geodesic arcs in a hyperbolic polygon  $\mathbf{P}$  with distinct endpoints. We say that the endpoints of  $\alpha$  and  $\beta$  in  $\partial\mathbf{P}$  interlace if each interval of  $\partial\mathbf{P}$  between the endpoints of  $\alpha$  contains an endpoint of  $\beta$  and vice versa.

We leave the easy proof of the following fact to the reader.

**Lemma 2.7.** Let  $\alpha$  and  $\beta$  be two closed geodesic arcs in a hyperbolic polygon  $\mathbf{P}$  with distinct endpoints. If the endpoints of  $\alpha$  and  $\beta$  interlace then  $\alpha$  and  $\beta$  intersect in an interior point of  $\mathbf{P}$ .

Finally, we give a detailed proof of our main result of this section.

*Proof of Proposition 2.4.* We begin by proving  $\gamma$  is arcwise connected. We divide the proof into two cases.

First, suppose that there is an arc  $\alpha_0$  of  $\gamma_0$  so that there are two non-adjacent sides of  $\mathbf{P}_0$  containing the endpoints of  $\alpha_0$ . Let  $[x_{i-1}, x_i]$  and  $[x_{j-1}, x_j]$  denote these two sides. Since these edges are not adjacent  $x_{i-1}$ ,  $x_i$ ,  $x_{j-1}$  and  $x_j$  are all distinct. Now consider  $\alpha_1 = \rho(\alpha_0)$ . It intersects the boundary of  $\mathbf{P}_0$  in the sides  $[x_i, x_{i+1}]$  and  $[x_j, x_{j+1}]$ . The intervals  $[x_{i-1}, x_i]$ ,  $[x_i, x_{i+1}]$ ,  $[x_{j-1}, x_j]$  and  $[x_j, x_{j+1}]$

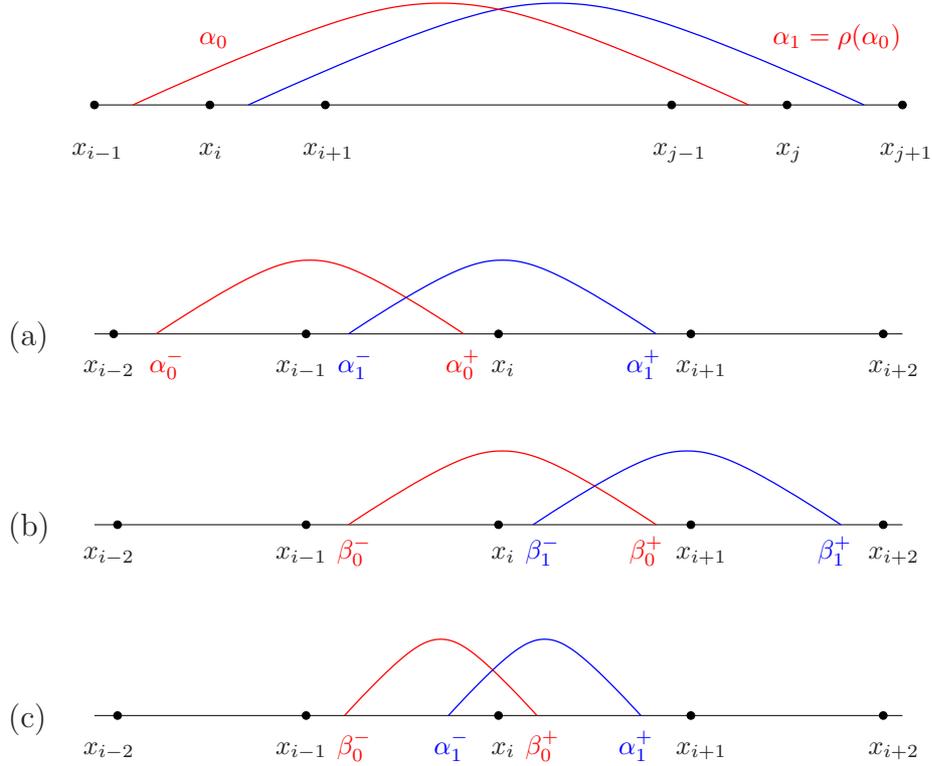


FIGURE 4. Configurations of interlacing in the proof of Proposition 2.4. The boundary  $\partial\mathbf{P}$  is straightened to simplify the illustration.

are distinct by construction and occur in this cyclic order. Therefore, as we move around the boundary of  $\mathbf{P}_0$  the endpoints of  $\alpha_0$  and  $\alpha_1$  interlace (see top of Figure 4). This means that  $\alpha_0$  and  $\alpha_1$  intersect in  $\mathbf{P}_0$  by Lemma 2.7. Hence  $\gamma_0$  and  $\gamma_1$  intersect. Applying powers of  $\rho$  we see that  $\gamma_i$  and  $\gamma_{i+1}$  intersect, thus proving that  $\gamma$  is arcwise connected.

Secondly, suppose that every arc of  $\gamma_0$  connects adjacent sides of  $\mathbf{P}_0$ . Every such arc has to connect interior points of the adjacent sides of  $\mathbf{P}_0$  for, otherwise, we would be in the first case. Using Lemma 2.5, we can find consecutive arcs  $\alpha_0$  and  $\beta_0$  of  $\gamma_0$  meeting  $\partial\mathbf{P}_0$  in three successive sides. To be precise, suppose one end  $\alpha_0^-$  of  $\alpha_0$  is a point in  $[x_{i-2}, x_{i-1}]$ , the common endpoint  $\alpha_0^+ = \beta_0^-$  of  $\alpha_0$  and  $\beta_0$  lies in  $[x_{i-1}, x_i]$  and the other endpoint  $\beta_0^+$  of  $\beta_0$  lies in  $[x_i, x_{i+1}]$ . Note that because  $\alpha_0$  and  $\beta_0$  are geodesic arcs, their only intersection point is their common endpoint. They therefore form an m-shaped curve. Now consider  $\alpha_1 \cup \beta_1 = \rho(\alpha_0 \cup \beta_0)$ , with endpoints  $\alpha_1^- \in [x_{i-1}, x_i]$ ,  $\alpha_1^+ = \beta_1^- \in [x_i, x_{i+1}]$  and  $\beta_1^+ \in [x_{i+1}, x_{i+2}]$ . If the sets  $\{\alpha_0^-, \alpha_0^+, \beta_0^+\}$  and  $\{\alpha_1^-, \alpha_1^+, \beta_1^+\}$  have a point in common, then  $\gamma_0$  and  $\gamma_1$  intersect, and so  $\gamma$  is connected as above. Thus, we may assume these two sets are disjoint. It suffices to show that certain endpoints of these arcs interlace and so, using Lemma 2.7, the corresponding arcs intersect. If  $<$  denotes the natural anti-clockwise order on  $[x_{i-1}, x_i] \cup [x_i, x_{i+1}]$ , then it is easy to show:

- (a) if  $\alpha_1^- < \alpha_0^+$  then  $\alpha_0^-, \alpha_1^-, \alpha_0^+, \alpha_1^+$  interlace, and so  $\alpha_0$  and  $\alpha_1$  intersect;
- (b) if  $\beta_1^- < \beta_0^+$  then  $\beta_0^-, \beta_1^-, \beta_0^+, \beta_1^+$  interlace, and so  $\beta_0$  and  $\beta_1$  intersect;
- (c) if  $\beta_0^- < \alpha_1^-$  and  $\beta_0^+ < \alpha_1^+$  then  $\beta_0^-, \alpha_1^-, \beta_0^+, \alpha_1^+$  interlace, and so  $\beta_0$  and  $\alpha_1$  intersect.

The cases (a)-(c) are illustrated in Figure 4. We observe that, since  $\alpha_0^+ = \beta_0^-$  and  $\alpha_1^+ = \beta_1^-$  then condition (c) is precisely the condition that both (a) and (b) fail. Therefore these three cases exhaust all possibilities. The argument then follows as in the first case.

This shows that  $\gamma$  is arcwise connected. Since  $\mathbf{P}_0$  is topologically a disc every connected component  $U$  of  $\mathbf{P}_0 \setminus \gamma$  is a topological disc. If every point of  $\partial U$  lies in  $\gamma$ , then we have case (a) of Definition 2.3. So suppose that  $\partial U$  contains a point of  $\partial \mathbf{P}_0$  that is not contained in  $\gamma$ . Then  $\partial U$  contains a non-empty topological arc of  $\partial \mathbf{P}_0$  both of whose endpoints lie in  $\gamma$  (note this arc is not necessarily contained in just one side of  $\mathbf{P}_0$ ). We claim that  $\partial U$  can contain at most one such topological arc in  $\partial \mathbf{P}_0$ . Suppose this is false. Then we can find four points  $b_1, b_2, c_1, c_2$  in  $\partial U$  so that (a) the points  $c_1, c_2$  lie in  $\gamma$ , (b) the points  $b_1, b_2$  lie in the interior of arcs of  $\partial \mathbf{P}_0$  not intersecting  $\gamma$  and (c) these four points  $b_1, c_1, b_2, c_2$  are interlaced. Therefore we can find a Jordan arc  $\delta$  from  $b_1$  to  $b_2$  (that is from  $\partial \mathbf{P}_0$  to itself) in  $U$  (except for its end points) so that the two connected components of  $\mathbf{P}_0 \setminus \delta$  each contains a point of  $\gamma$ , namely  $c_1$  and  $c_2$ . This contradicts the connectedness of  $\gamma$ . Hence  $U$  can contain at most one topological arc of  $\partial \mathbf{P}_0$  in its boundary. Recall, we are assuming such an arc exists, or else we are in case (a) of Definition 2.3. Call this arc  $\varepsilon$ . If the interior of  $\varepsilon$  is contained in only one side of  $\mathbf{P}_0$ , then we are in case (b) of Definition 2.3. If the interior of  $\varepsilon$  contains points in precisely two different sides of  $\mathbf{P}_0$ , then these two sides must be adjacent, say  $[x_{i-1}, x_i]$  and  $[x_i, x_{i+1}]$ , and their common vertex  $x_i$  must also be contained in the interior of  $\varepsilon$ . In particular,  $\gamma$  does not pass through the vertex  $x_i$ . Since  $\gamma$  is preserved by the symmetry map  $\rho$ , we see that  $\gamma$  does not pass through any vertex of  $\mathbf{P}_0$ . Since  $\gamma$  intersects the sides  $[x_{i-1}, x_i]$  and  $[x_i, x_{i+1}]$  and does not contain their endpoints, it must contain points of both their interiors. In particular,  $\varepsilon$  does not contain  $x_{i-1}$  or  $x_{i+1}$ . Hence we are in case (c) of Definition 2.3. Finally, suppose that the interior of  $\varepsilon$  contains points from at least three sides of  $\mathbf{P}_0$ . As  $\varepsilon$  is connected, this means it contains a whole side of  $\mathbf{P}_0$ , which contradicts the fact that, by symmetry, each side of  $\mathbf{P}_0$  intersects  $\gamma$ . Thus, the only possibilities for  $\partial U$  are (a), (b) and (c) from Definition 2.3 as required.  $\square$

### 3. TEICHMÜLLER SPACE AND FENCHEL-NIELSEN COORDINATES

From now on we fix  $k \geq 3$  and we only consider *ideal*  $k$ -gons. In contrast to the convention in the previous section, our ideal  $k$ -gons  $\mathbf{P}$  do not contain the vertices at infinity, but they contain the sides and are therefore closed subsets of  $\mathbb{D}^2$ . A key observation in the proof of Theorem 1.1 is that every ideal  $k$ -gon  $\mathbf{P}$  gives rise to a Riemann surface  $S_{\mathbf{P}}$  (its billiard surface) via a gluing process of two copies of  $\mathbf{P}$ , denoted by  $\mathbf{P}^+$  and  $\mathbf{P}^-$ , along corresponding sides, and that every closed billiard trajectory in  $\mathbf{P}$  gives rise to one or two closed geodesics in  $S_{\mathbf{P}}$ . This allows us to apply powerful results from Teichmüller theory.

Let us first set up the Teichmüller space framework and introduce the relevant objects. A Riemann surface (of finite type)  $S$  is a 2-dimensional oriented differentiable manifold with finitely many ends, carrying a Riemannian metric of constant

curvature minus one. We suppose that  $S$  has finite area with respect to this metric. In particular, the ends are realised as cusps. The universal covering of  $S$  agrees with  $\mathbb{D}^2$  and the canonical complex structure of  $\mathbb{D}^2$  induces a complex structure on  $S$ . Thus it makes sense to consider holomorphic and anti-holomorphic isometries of  $S$ .

Let  $\mathbf{P} \subset \mathbb{D}^2$  be an ideal  $k$ -gon (with anti-clockwise enumerated vertices  $x_1, \dots, x_k$  as in Definition 2.1) and  $S_{\mathbf{P}}$  be the corresponding billiard surface. Note that  $S_{\mathbf{P}}$  is, topologically, homeomorphic to a  $k$ -punctured sphere. The ends of  $S_{\mathbf{P}}$  correspond to the vertices  $x_j$ . Note that  $S_{\mathbf{P}}$  is a *labelled billiard surface* since its ends carry labels in  $\{1, 2, \dots, k\}$ . Similarly,  $\mathbf{P}$  is a *labelled polygon* where the bi-infinite geodesic side  $(x_i, x_{i+1})$  of  $\mathbf{P}$  without its end points is endowed with the label  $i \pmod{k}$ .  $S_{\mathbf{P}}$  has a natural anti-holomorphic isometry  $J_{\mathbf{P}} : S_{\mathbf{P}} \rightarrow S_{\mathbf{P}}$  interchanging  $\mathbf{P}^+$  and  $\mathbf{P}^-$ , and with fixed point set  $\cup_{i=1}^k (x_i, x_{i+1})$ . Let  $\mathbf{P}_0 \subset \mathbb{D}^2$  be an ideal regular  $k$ -gon with anti-clockwise labelling,  $R_0 = S_{\mathbf{P}_0}$  be the corresponding labelled billiard surface, and  $J_0 = J_{\mathbf{P}_0} : R_0 \rightarrow R_0$  be the corresponding anti-holomorphic isometry.

**Definition 3.1.** *The Teichmüller space  $T(R_0)$  is the set of all equivalence classes of pairs  $(S, f)$  where  $S$  is an oriented Riemann surface and  $f : R_0 \rightarrow S$  is a quasiconformal mapping. Two such pairs  $(S, f)$  and  $(S', f')$  are equivalent if the map  $f' \circ f^{-1} : S \rightarrow S'$  is homotopic to an orientation preserving isometry. We denote the equivalence class associated to the pair  $(S, f)$  by  $[S, f]$ . A point  $[S, f]$  in Teichmüller space  $T(R_0)$  is also called a marked Riemann surface.*

The Teichmüller space  $T(R_0)$  carries a natural complex manifold structure and the anti-holomorphic isometry  $J_0 : R_0 \rightarrow R_0$  gives rise to an anti-holomorphic automorphism  $\mathcal{F}$  on Teichmüller space (see [6, pp. 229]), which we call the flip map.

**Definition 3.2.** *Let  $\varphi : R_0 \rightarrow R_0$  be an orientation preserving quasiconformal mapping. Then we define the induced map  $\varphi_* : T(R_0) \rightarrow T(R_0)$  as*

$$\varphi_*([S, f : R_0 \rightarrow S]) = [S, f \circ \varphi : R_0 \rightarrow S].$$

The flip map  $\mathcal{F} : T(R_0) \rightarrow T(R_0)$  is defined as

$$(3.1) \quad \mathcal{F}([S, f : R_0 \rightarrow S]) = [S^*, j_S \circ f \circ J_0 : R_0 \rightarrow S^*].$$

Here,  $S^*$  is the same surface as  $S$  but with the opposite orientation and  $j_S : S \rightarrow S^*$  is, as a map, the pointwise identity.

Let  $\rho : \mathbf{P}_0 \rightarrow \mathbf{P}_0$  be the anti-clockwise rotation through angle  $2\pi/k$  about the centre of  $\mathbf{P}_0$ . By abuse of notation, we denote the associated rotation in the corresponding billiard surface, again, by  $\rho : R_0 \rightarrow R_0$ . The induced map  $\rho_* : T(R_0) \rightarrow T(R_0)$  has order  $k$ . Note that the special point  $x_0 = [R_0, \text{id} : R_0 \rightarrow R_0] \in T(R_0)$  is a common fixed point of both  $\rho_*$  and the flip map  $\mathcal{F}$ .

Our next aim is to introduce suitable Fenchel-Nielsen coordinates  $(l, \tau)$ , which yield a diffeomorphism between  $T(R_0)$  and  $(\mathbb{R}^+)^{k-3} \times \mathbb{R}^{k-3}$ . We first decompose  $\mathbf{P}_0$  into right angled compact hexagons, right angled pentagons with one ideal vertex and right angled quadrilaterals with two ideal vertices. Such a decomposition induces a decomposition of  $R_0$  into  $k-2$  pairs of pants  $Y_1, \dots, Y_{k-2}$  with three/two/one geodesic boundary cycles, respectively. Each of these pairs of pants  $Y_j$  is invariant (as a set) under the reflection  $J_0$ , and they have their own reflections  $J_{Y_j}$  which agree with the restrictions of  $J_0$  to  $Y_j$ . For illustration, we now use

the following colour convention: The  $k - 3$  boundary cycles  $C_1, \dots, C_{k-3} \subset R_0$  of the pants decomposition  $R_0 = \bigcup_{j=1}^{k-2} Y_j$  are *green lines*. The bi-infinite geodesics  $(x_i, x_{i+1}) \subset R_0$  are *red lines*. Cutting  $R_0$  along all red lines splits the surface into the two polygons  $\mathbf{P}_0^+$  and  $\mathbf{P}_0^-$ .

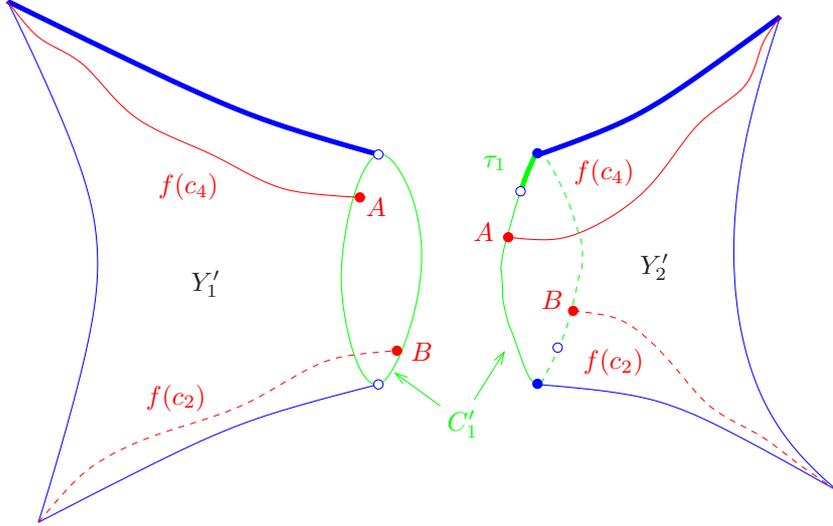


FIGURE 5. The green boundary cycles of  $Y'_1$  and  $Y'_2$  are identified such that the  $A$ 's and  $B$ 's fit together. The red curve  $f(c_4)$  is freely homotopic (fixing the end points) to the union of the two thick blue arcs and the thick green arc of length  $\tau_1$  in between.

The Fenchel-Nielsen coordinates of a point  $[S, f] \in T(R_0)$  are now given as follows: Let  $C'_j \subset S$  be the unique closed geodesic corresponding to  $f(C_j)$  modulo free homotopy in  $S$ . Again, we think of the curves  $C'_j$  as *green lines*. They give rise to a pants decomposition  $S = \bigcup_{j=1}^{k-2} Y'_j$  agreeing, combinatorially, with the pants decomposition of  $R_0$ . The length parameters of  $[S, f]$  are then given by the lengths  $l_j \in \mathbb{R}^+$  of the boundary cycles  $C'_j \subset S$ .

For every geodesic  $c_i = (x_i, x_{i+1}) \subset R_0$  of  $R_0$ , let its image  $f(c_i) \subset S$  again carry the *colour red*. Note that the bi-infinite curves  $f(c_i) \subset S$  are generally no longer geodesics. Each pair of pants  $Y'_j$  in the decomposition of  $S$  comes equipped with a triplet of *blue geodesic arcs*, namely the fixed point set of the intrinsic reflection  $J_{Y'_j}$  of this pair of pants. Now, for every bi-infinite red curve  $f(c_i) \subset S$  there exists a unique regular freely homotopic curve connecting the same ends, which is made up of alternating blue and green arcs (regular means here that we do not allow to go back and forth in certain parts of the curve). This means that the curve  $f(c_i)$  defines an arc in each green boundary cycle  $C'_j$  along its path, and the length of this arc provides a unique twist parameter  $\tau_j \in \mathbb{R}$ . Note that the sign of the twist parameter is uniquely determined by the orientations of the pairs of pants and their boundary cycles. Note also, that every boundary cycle  $C'_j$  defines an  $X$ -piece (two pairs of pants glued along  $C'_j$ ) and there are two curves  $f(c_{i_1})$  and  $f(c_{i_2})$  intersecting it. The twist parameter  $\tau_j \in \mathbb{R}$  is independent of the choice of  $f(c_{i_1})$

or  $f(c_{i_2})$  (see Figure 5 for illustration of the twist parameter  $\tau_1$  in  $S = Y'_1 \cup Y'_2$ ). For further details we refer to, e.g., [5, Section 7.6].

**Definition 3.3.** *We denote by  $B(R_0)$  the subset of  $T(R_0)$  with vanishing twist parameters. The points  $[S, f] \in B(R_0)$  are called marked billiard surfaces. We refer to  $B(R_0)$  as the billiard space associated to  $\mathbf{P}_0$ .*

#### 4. PROPERTIES OF THE BILLIARD SPACE

Now we explain that each point of  $B(R_0)$  can be realised by a labelled billiard surface  $S$  together with an (almost canonical) quasiconformal mapping  $f : R_0 \rightarrow S$ , respecting the labelling (i.e., mapping the  $i$ -th end of  $R_0$  to the  $i$ -th end of  $S$ , for  $i = 1, \dots, k$ ). Given the length coordinates  $(l_1, \dots, l_{k-3})$ , we can construct an ideal hyperbolic  $k$ -gon  $\mathbf{P}$  with these parameters in its decomposition into hexagons, pentagons and quadrilaterals consistent with the decomposition of  $\mathbf{P}_0$ . Next, we choose quasiconformal maps from each building block (hexagon/pentagon/quadrilateral) of  $\mathbf{P}_0$  to the corresponding building block of  $\mathbf{P}$  mapping corresponding boundary components onto each other such that they can be combined to a global quasiconformal map  $f_{\mathbf{P}} : R_0 \rightarrow S_{\mathbf{P}}$ , equivariant under the global reflections  $J_0$  and  $J_{S_{\mathbf{P}}}$ :

$$(4.1) \quad f_{\mathbf{P}} \circ J_0 = J_{S_{\mathbf{P}}} \circ f_{\mathbf{P}}.$$

By construction, the union  $\bigcup_{i=1}^k c_i$  of the bi-infinite *red lines* of  $R_0$  are mapped under  $f_{\mathbf{P}}$  onto the union of the *blue geodesic arcs* of the pairs of pants  $Y'_j$  of  $S_{\mathbf{P}}$ , and the *green* boundary cycles of the pants decomposition of  $R_0$  are mapped under  $f_{\mathbf{P}}$  onto the corresponding *green* boundary cycles of the pants decomposition of  $S_{\mathbf{P}}$ . This fact guarantees that all twist parameters of  $[S_{\mathbf{P}}, f_{\mathbf{P}}]$  are zero. In this context, we can think of  $B(R_0)$  as the “subset of labelled billiard surfaces” in  $T(R_0)$ .

**Remark 4.1.** *As seen above, a general point  $x \in B(R_0)$  is an equivalence class  $x = [S_{\mathbf{P}}, f_{\mathbf{P}} : R_0 \rightarrow S_{\mathbf{P}}]$  with an almost canonical quasiconformal mapping  $f_{\mathbf{P}}$ . Note that  $x = [S_{\mathbf{P}}, f_{\mathbf{P}}]$  agrees with  $x_0 = [R_0, \text{id} : R_0 \rightarrow R_0] \in B(R_0)$  if and only if the polygon  $\mathbf{P}$  is regular.*

The Teichmüller space  $T(R_0)$  carries a complex manifold structure with a natural symplectic form  $\omega_{WP}$ , the *Weil-Petersson symplectic form*. By Wolpert’s Theorem (see [10]),  $\omega_{WP}$  can be written in terms of the Fenchel-Nielsen coordinates  $(l, \tau)$  of  $T(R_0)$  as follows:

$$\omega_{WP} = - \sum_{j=1}^{k-3} d\tau_j \wedge dl_j.$$

The symplectic form  $\omega_{WP}$  and the almost complex structure on  $T(R_0)$  induce a Kähler metric  $g_{WP}$  on  $T(R_0)$ , the *Weil-Petersson metric*. While the Riemannian metric  $g_{WP}$  is generally not complete (see [9]), it is still true that any pair of points  $x_1, x_2 \in T(R_0)$  can be joined by a unique Weil-Petersson geodesic (see [11]). The billiard space  $B(R_0)$  has the following useful properties.

**Proposition 4.2.** *The billiard space  $B(R_0)$  is a Lagrangian submanifold of the symplectic manifold  $(T(R_0), \omega_{WP})$ . Moreover,  $B(R_0)$  is a geodesically convex subset of  $(T(R_0), g_{WP})$ , i.e., for given  $x_1, x_2 \in B(R_0)$  the unique Weil-Petersson geodesic connecting  $x_1$  and  $x_2$  lies entirely in  $B(R_0)$ .*

*Proof.* The flip map  $\mathcal{F}$ , defined in (3.1), is an isometry with respect to  $g_{WP}$  (see [6, p. 230]). Written in our Fenchel-Nielsen coordinates we have

$$(4.2) \quad \mathcal{F}(l, \tau) = (l, -\tau).$$

This follows from [6, p. 230, bottom formula] and the fact that  $x_0 = [R_0, \text{id}]$  is a fixed point of  $\mathcal{F}$ . Therefore, the fixed point set of  $\mathcal{F}$  is the space  $B(R_0)$  of marked billiard surfaces. By the above considerations,  $B(R_0)$  is a Lagrangian submanifold and, as the fixed point set of an isometry,  $B(R_0)$  is geodesically convex.  $\square$

Finally, we give an important characterisation of the point  $x_0 \in B(R_0)$ .

**Proposition 4.3.** *The only simultaneous fixed point of  $\rho_*$  and  $\mathcal{F}$  in  $T(R_0)$  is  $x_0 = [R_0, \text{id}] \in B(R_0)$ .*

*Proof.* Let  $x \in T(R_0)$  be a simultaneous fixed point of  $\mathcal{F}$  and  $\rho_*$ .

The fixed point property  $\mathcal{F}(x) = x$  and (4.2) imply that  $x \in B(R_0)$ . Therefore,  $x$  has a representation  $x = [S_{\mathbf{P}}, f_{\mathbf{P}} : R_0 \rightarrow S_{\mathbf{P}}]$  with  $\mathbf{P} \subset \mathbb{D}^2$  a labelled ideal hyperbolic  $k$ -gon and  $f_{\mathbf{P}} \circ J_0 = J_{S_{\mathbf{P}}} \circ f_{\mathbf{P}}$ . By Remark 4.1, we only have to show that  $\mathbf{P}$  is regular.

The fixed point property  $\rho_*(x) = x$  means that  $g_0 := f_{\mathbf{P}} \circ \rho \circ f_{\mathbf{P}}^{-1} : S_{\mathbf{P}} \rightarrow S_{\mathbf{P}}$  is homotopic to an isometry  $g_1 : S_{\mathbf{P}} \rightarrow S_{\mathbf{P}}$ . Since a homotopy between two maps on  $S_{\mathbf{P}}$  preserves the ends,  $g_1$  maps the end  $j$  of  $S_{\mathbf{P}}$  to the end  $j+1$  (modulo  $k$ ). Let  $c_j$  be the unique geodesic in  $S_{\mathbf{P}}$  connecting the ends  $j$  and  $j+1$  modulo  $k$ . Then the set  $C = \bigcup_{j=1}^k c_j$  splits  $S_{\mathbf{P}}$  into the ideal polygons  $\mathbf{P}^+$  and  $\mathbf{P}^-$ , both isometric to  $\mathbf{P}$ , and we have  $g_1(c_j) = c_{j+1}$  for all  $j$ , modulo  $k$ . This means that  $g_1(C) = C$  and the isometry  $g_1$  either interchanges  $\mathbf{P}^+$  and  $\mathbf{P}^-$  or preserves them as sets. Since  $g_1$  is orientation preserving, we cannot have  $g_1(\mathbf{P}^+) = \mathbf{P}^-$ . This shows that we have  $g_1 : \mathbf{P}^+ \rightarrow \mathbf{P}^+$ .

Now we embed  $\mathbf{P}^+$  into  $\mathbb{D}^2$  and compactify  $\mathbf{P}^+$  by adding the ideal vertices  $x_1, \dots, x_k \in \partial\mathbb{D}^2$  corresponding to the ends  $1, \dots, k$ , respectively. Then the isometry  $g_1$  extends to a continuous map, denoted again by  $g_1$ , of the compactification  $\overline{\mathbf{P}^+}$ . By Brouwer's Fixed Point Theorem, there exists  $z_0 \in \overline{\mathbf{P}^+}$  such that  $g_1(z_0) = z_0$ . This point must be an interior point of  $\overline{\mathbf{P}^+}$  since the boundary of  $\overline{\mathbf{P}^+}$ , consisting of the points  $x_1, \dots, x_k$  and the geodesics  $c_j$ , cannot have a fixed point (recall that  $g_1$  maps  $x_j$  to  $x_{j+1}$  modulo  $k$ ). Let  $r_j$  be the geodesic ray connecting  $z_0$  with the ideal point  $x_j$ . Then  $g_1$  maps the triangle with vertices  $z_0, x_j, x_{j+1}$  to the triangle with vertices  $z_0, x_{j+1}, x_{j+2}$  modulo  $k$ . Therefore, all the triangles with vertices  $z_0, x_j, x_{j+1}$  for  $j = 1, \dots, k$  are isometric to one another. Since isometries preserve angles, the angle between the rays  $r_j$  and  $r_{j+1}$  at  $z_0$  must therefore be  $2\pi/k$ . This shows that  $\mathbf{P}^+ \subset \mathbb{D}^2$  is a regular  $k$ -gon, finishing the proof.  $\square$

## 5. GEODESIC LENGTHS FUNCTIONS AND CYCLICALLY RELATED BILLIARD TRAJECTORIES

Recall that  $\mathbf{P}_0 \subset \mathbb{D}^2$  denotes a labelled regular ideal  $k$ -gon and that  $R_0$  is its associated billiard surface with rotational symmetry  $\rho : R_0 \rightarrow R_0$ . Let us now introduce geodesic length functions on Teichmüller space.

**Definition 5.1.** *A closed curve in  $R_0$  is called essential if it is neither null-homotopic nor spirals around one of the ends of  $R_0$ . Let  $\tilde{\gamma} = \{\tilde{\gamma}_1, \dots, \tilde{\gamma}_N\}$  be a finite family of essential closed curves  $\tilde{\gamma}_i : S^1 \rightarrow R_0$ . For  $x = [S, f] \in T(R_0)$ , let  $\tilde{\gamma}'_i$*

be the unique closed geodesic which is freely homotopic to  $f(\tilde{\gamma}_i)$ . Then the geodesic length function associated to  $\tilde{\gamma}$  is a map

$$L = L_{\tilde{\gamma}} : T(R_0) \rightarrow [0, \infty),$$

defined by

$$L(x) = \sum_{i=1}^N \text{length}(\tilde{\gamma}'_i).$$

Note that we can continuously deform a curve spiralling around one of the ends of a hyperbolic surface into an arbitrarily short curve by moving it up into the end. This is the reason that such curves are not considered to be essential. The following fundamental convexity result of Wolpert will be key for the proof of Theorem 1.1.

**Theorem 5.2** ([11, Cor. 4.7]). *Let  $\tilde{\gamma} \subset R_0$  be a finite family of essential closed curves and  $L = L_{\tilde{\gamma}} : T(R_0) \rightarrow (0, \infty)$  be the associated geodesic length function. Then the function  $L$  is continuous and strictly convex along every Weil-Petersson geodesic.*

Let us now link this concept with cyclically related closed billiard trajectories in different ideal hyperbolic  $k$ -gons. This requires further notation.

Let  $\mathbf{P} \subset \mathbb{D}^2$  be a labelled ideal  $k$ -gon. It was shown in [3, Thm 2.1] that a finite sequence  $\mathbf{a} = (a_0, a_1, \dots, a_{n-1})$  is a billiard sequence (i.e., a coding of a closed billiard trajectory) if and only if (a) consecutive values  $a_j$  and  $a_{j+1}$  with indices taken modulo  $n$  do not coincide and (b) if  $\mathbf{a}$  contains only two different labels, then they must not be neighbours (i.e., must not differ by  $\pm 1$  modulo  $k$ ). Let  $\gamma_{\mathbf{a}, \mathbf{P}}$  be the family consisting of the unique closed billiard trajectory associated to  $\mathbf{a}$  and all its cyclically related billiard trajectories in  $\mathbf{P}$ . Then  $\gamma_{\mathbf{a}, \mathbf{P}}$  consists of  $k$  piecewise geodesic closed curves  $\gamma_i$ . Let  $L_{av}(\mathbf{P}, \mathbf{a})$  be the average length of these curves, i.e.,

$$L_{av}(\mathbf{P}, \mathbf{a}) = \frac{1}{k} \sum_{\gamma_i \in \gamma_{\mathbf{a}, \mathbf{P}}} \text{length}(\gamma_i).$$

Let  $\pi_{\mathbf{P}} : S_{\mathbf{P}} \rightarrow \mathbf{P}$  be the canonical projection and  $\tilde{\gamma}_{\mathbf{a}, \mathbf{P}} = \pi_{\mathbf{P}}^{-1}(\gamma_{\mathbf{a}, \mathbf{P}})$  be the lift of these billiard trajectories in the corresponding billiard surface. Note that  $\tilde{\gamma}_{\mathbf{a}, \mathbf{P}}$  consists of  $2k$  or  $k$  closed geodesics in  $S_{\mathbf{P}}$ , depending on whether  $n$  is even or odd: Let  $\tilde{\gamma}_i$  be one of the closed geodesics in  $\tilde{\gamma}_{\mathbf{a}, \mathbf{P}}$ . Then there exists a fixed integer  $t$ , such that every label  $s = a_j$  corresponds to a transversal crossing between  $\tilde{\gamma}_i$  and a bi-infinite geodesic  $(x_{s+t}, x_{s+t+1})$  (where indices are taken modulo  $k$ ), i.e.,  $\tilde{\gamma}_i$  changes from  $\mathbf{P}^{\pm}$  to  $\mathbf{P}^{\mp}$ . After  $n$  such changes  $\tilde{\gamma}_i$  will not close up if  $n$  is odd. This is the reason why, in this case,  $\tilde{\gamma}_{\mathbf{a}, \mathbf{P}}$  consists of  $k$  geodesics corresponding to cutting sequences<sup>1</sup> cyclically related to the doubling  $\mathbf{a}\mathbf{a} = (a_0, a_1, \dots, a_{n-1}, a_0, \dots, a_{n-1})$ . But it is obvious that we have in both cases

$$\sum_{\tilde{\gamma}_i \in \tilde{\gamma}_{\mathbf{a}, \mathbf{P}}} \text{length}(\tilde{\gamma}_i) = 2kL_{av}(\mathbf{P}, \mathbf{a}).$$

Moreover, the left hand side can be rewritten as the geodesic length function associated to  $\tilde{\gamma}_{\mathbf{a}} = \tilde{\gamma}_{\mathbf{a}, \mathbf{P}_0}$ , i.e.,

$$(5.1) \quad L_{\tilde{\gamma}_{\mathbf{a}}}([S_{\mathbf{P}}, f_{\mathbf{P}}]) = 2kL_{av}(\mathbf{P}, \mathbf{a}),$$

<sup>1</sup>As in the case of a labelled polygon  $\mathbf{P}$ , we can associate a symbolic coding to a closed curve in a labelled billiard surface  $S_{\mathbf{P}}$  reflecting its crossings with the bi-infinite geodesics connecting subsequent ends. We refer to it as the *cutting sequence* associated to the curve.

with  $f_{\mathbf{P}} : R_0 \rightarrow S_{\mathbf{P}}$  introduced at the beginning of Section 4. Note here that each closed curve in  $f_{\mathbf{P}}(\tilde{\gamma}_{\mathbf{a}})$  is freely homotopic to a corresponding curve in the family  $\tilde{\gamma}_{\mathbf{a}, \mathbf{P}}$  since both closed curves in  $S_{\mathbf{P}}$  have the same cutting sequences.

We finish this section with the following useful observation.

**Lemma 5.3.** *Let  $\mathbf{a} = (a_0, \dots, a_{n-1})$  be a billiard sequence. Then we have, for all  $x \in T(R_0)$ ,*

$$L_{\tilde{\gamma}_{\mathbf{a}}}(x) = L_{\tilde{\gamma}_{\mathbf{a}}}(\rho_*(x)) = L_{\tilde{\gamma}_{\mathbf{a}}}(\mathcal{F}(x)).$$

*Proof.* Note that  $\tilde{\gamma}_{\mathbf{a}} = \{\tilde{\gamma}_1, \dots, \tilde{\gamma}_N\}$  with  $N = k$  or  $N = 2k$  is a family of closed geodesics in  $R_0$  which, as a set, is invariant under  $\rho$  and  $J_0$  by its very construction. If  $x = [S, f : R_0 \rightarrow S]$  then  $\rho_*(x) = [S, f \circ \rho : R_0 \rightarrow S]$  and we have

$$L_{\tilde{\gamma}_{\mathbf{a}}}(\rho_*(x)) = \sum_{i=1}^N \text{length}(\tilde{\gamma}'_i),$$

where  $\tilde{\gamma}'_i$  is freely homotopic to  $f \circ \rho(\tilde{\gamma}_i)$ . The result for  $\rho$  follows now from the fact that  $\rho$  only permutes the closed curves  $\tilde{\gamma}_i$ . The result for the flip map  $\mathcal{F}$  follows analogously from the fact that also  $J_P$  only permutes the closed curves  $\tilde{\gamma}_i$ .  $\square$

## 6. PROOF OF THEOREM 1.1

As before, let  $\mathbf{P}_0 \subset \mathbb{D}^2$  be a labelled regular ideal  $k$ -gon and  $\mathbf{a}$  be a finite billiard sequence. Theorem 1.1 in the Introduction states that

$$(6.1) \quad L_{av}(\mathbf{P}, \mathbf{a}) \geq L_{av}(\mathbf{P}_0, \mathbf{a})$$

for all ideal  $k$ -gons  $\mathbf{P} \subset \mathbb{D}^2$  with equality if and only if  $P$  is regular. Recall that  $\tilde{\gamma}_{\mathbf{a}}$  is a family of closed geodesics in  $R_0$  associated to the billiard sequence  $\mathbf{a}$  and that  $x_0 = [R_0, \text{id}] \in B(R_0) \subset T(R_0)$ . Then (6.1) is a consequence of the following, by identity (5.1): For any finite billiard sequence  $\mathbf{a}$ , the geodesic length function associated to  $\tilde{\gamma}_{\mathbf{a}}$  satisfies

$$(6.2) \quad L_{\tilde{\gamma}_{\mathbf{a}}}(x) \geq L_{\tilde{\gamma}_{\mathbf{a}}}(x_0),$$

with equality iff  $x = x_0$ . So our goal is to prove (6.2).

Let us return to the property of closed curves to be filling, but now in the setting of the Riemann surface  $R_0$ .

**Definition 6.1.** *A family of closed curves  $\{\tilde{\gamma}_1, \dots, \tilde{\gamma}_N\}$  in  $R_0$  is called filling if each connected component of  $R_0 \setminus \bigcup_{i=1}^N \tilde{\gamma}_i$  is topologically an open disc or a once-punctured open disc.*

The importance of being filling becomes clear in the following result by Kerckhoff.

**Proposition 6.2** ([7, Lemma 3.1]). *Let  $\{\tilde{\gamma}_1, \dots, \tilde{\gamma}_N\}$  be a finite family of closed curves and  $L : T(R_0) \rightarrow (0, \infty)$  be the associated geodesic length function, introduced in Definition 5.1. If this family is filling, then  $L$  is a proper function.*

Proposition 6.2 and Theorem 5.2 together imply the following corollary. The proof of this corollary is well known (see, e.g., last paragraph of [11] or also [7, Thm 3]) but we include it here for the reader's convenience.

**Corollary 6.3.** *Let  $\{\tilde{\gamma}_1, \dots, \tilde{\gamma}_N\}$  be a finite family of closed essential curves which is filling and  $L : T(R_0) \rightarrow (0, \infty)$  be the associated geodesic length function. Then there is a unique point  $x_{min} \in T(R_0)$  where  $L$  assumes its global minimum.*

*Proof.* Let  $L_0 = \inf\{L(x) \mid x \in T(R_0)\} \geq 0$  and  $x_m \in T(R_0)$  be a sequence satisfying  $\lim_{m \rightarrow \infty} L(x_m) \rightarrow L_0$ . Since  $L$  is proper, by Proposition 6.2,  $L^{-1}([0, L_0 + 1])$  is compact and there exists a convergent subsequence  $x_{m_j} \rightarrow x_{min} \in T(R_0)$  with

$$0 < L(x_{min}) = \lim_{j \rightarrow \infty} L(x_{m_j}) = L_0.$$

Assume we have another point  $x' \in T(R_0)$  with  $L(x') = L_0$ . Then there exists a unique geodesic connecting  $x_{min}$  and  $x'$ , along which  $L$  is strict convex, by Theorem 5.2. This would lead to a point  $x'' \in T(R_0)$  between  $x_{min}$  and  $x'$  with  $L(x'') < L(x_{min}) = L_0$ , which is a contradiction.  $\square$

Let us, finally, present the proof of (6.2): We first explain why  $\tilde{\gamma}_{\mathbf{a}} = \{\tilde{\gamma}_1, \dots, \tilde{\gamma}_N\}$  with  $N = k$  or  $N = 2k$  is filling in  $R_0$ . We know from Proposition 2.4 that if  $\gamma_0$  denotes the closed billiard trajectory corresponding to  $\mathbf{a}$  in  $\mathbf{P}_0$  and  $\gamma_i = \rho^i(\gamma_0)$ , then  $\gamma = \bigcup_{i=0}^{k-1} \gamma_i$  fills  $\mathbf{P}_0$ . Now,  $R_0$  consists of two copies  $\mathbf{P}_0^\pm$  of  $\mathbf{P}_0$ , glued along their boundaries. Under the identification  $\mathbf{P}_0 = \mathbf{P}_0^+$ , we have

$$R_0 \setminus \bigcup_{i=1}^N \tilde{\gamma}_i = \left( \mathbf{P}_0 \setminus \bigcup_{i=0}^{k-1} \gamma_i \right) \cup J_0 \left( \mathbf{P}_0 \setminus \bigcup_{i=0}^{k-1} \gamma_i \right),$$

and from the domains with properties (a), (b), (c) in Definition 2.3 it is easy to see that the connected components of  $R_0 \setminus \bigcup_{i=1}^N \tilde{\gamma}_i$  are either topologically an open disc or a once-punctured open disc. This shows that  $\tilde{\gamma}_{\mathbf{a}}$  is filling. Moreover, the geodesics  $\tilde{\gamma}_i$  are essential and we conclude from Corollary 6.3 that there exists a *unique* point  $x_{min} \in T(R_0)$  with

$$L(x) > L(x_{min}) \quad \text{for all } x \in T(R_0), x \neq x_{min},$$

where  $L$  denotes the geodesic length function associated to  $\tilde{\gamma}_{\mathbf{a}}$ . It only remains to identify this global minimum. We know from Lemma 5.3 that  $L(x) = L(\rho_*(x)) = L(\mathcal{F}(x))$ , and the uniqueness of the minimum implies that we have

$$x_{min} = \rho_*(x_{min}) = \mathcal{F}(x_{min}).$$

It then follows from Proposition 4.3 that we must have  $x_0 = x_{min}$ .  $\square$

#### APPENDIX A. BILLIARD IN EUCLIDEAN RECTANGLES

In this appendix, we discuss an Euclidean analogue of the conjecture, namely we consider lengths of cyclically related closed billiard trajectories in Euclidean rectangles of area one. For every  $c > 0$ , we introduce the rectangular billiard table  $\mathbf{P}_c = [0, c] \times [0, 1/c] \subset \mathbb{R}^2$ . Every closed billiard trajectory in  $\mathbf{P}_c$  is, up to free homotopy, in one-one correspondence with a vector  $(nc, m/c)$  with  $(n, m) \in \mathbb{Z}^2 \setminus \{(0, 0)\}$ . The closed billiard trajectories cyclically related to  $(nc, m/c)$  are  $(-mc, n/c)$ ,  $(-nc, -m/c)$  and  $(mc, -n/c)$ . The lengths of these four cyclically related billiard trajectories add up to

$$L_{n,m}(c) = 2\sqrt{n^2c^2 + \frac{m^2}{c^2}} + 2\sqrt{m^2c^2 + \frac{n^2}{c^2}}.$$

The Euclidean analogue of the conjecture in this "baby case" then reads as

$$(A.1) \quad L_{n,m}(c) \geq L_{n,m}(1),$$

with equality if and only if  $c = 1$ . (A.1) is equivalent to

$$\sqrt{n^2c^2 + \frac{m^2}{c^2}} + \sqrt{m^2c^2 + \frac{n^2}{c^2}} \geq 2\sqrt{n^2 + m^2}.$$

Squaring both sides leads to

$$2\sqrt{n^2c^2 + \frac{m^2}{c^2}}\sqrt{m^2c^2 + \frac{n^2}{c^2}} \geq 2(n^2 + m^2) - \left(c^2 - \frac{1}{c^2}\right)(n^2 + m^2).$$

This shows that we have (A.1) if

$$\sqrt{n^2c^2 + \frac{m^2}{c^2}}\sqrt{m^2c^2 + \frac{n^2}{c^2}} \geq (n^2 + m^2).$$

Squaring again yields

$$\left(c^4 + \frac{1}{c^4}\right)n^2m^2 \geq 2n^2m^2,$$

which holds obviously for all  $c > 0$ . It is easy to see that the equality case leads to  $c = 1$ , completing the elementary proof.

Recall that in the case of *hyperbolic polygons* we associated to every billiard table a billiard surface. Let us briefly explain what this means in our case: Reflections of the billiard table  $\mathbf{P}_c$  along its sides leads to the rectangle  $[0, 2c] \times [0, 2/c]$  which, after identification of its opposite sides, becomes a torus, denoted by  $S_c$ , the billiard surface associated to the billiard table  $\mathbf{P}_c$ . Then every closed billiard trajectory, traversed twice, can be viewed as a closed geodesic in  $S_c$ . Such a closed geodesic is then again, up to free homotopy, in one-one correspondence with a vector  $(2nc, 2m/c)$ . Our inequality above about cyclically related closed billiard trajectories then naturally translates to a corresponding statement about closed geodesics in the associated billiard surfaces. The relevant Teichmüller space is then the space of all closed flat oriented surfaces (of genus 1), which we can identify with the hyperbolic upper half plane  $\mathbb{H}^2 = \{z \in \mathbb{C} \mid \text{Im}(z) > 0\}$ . More concretely, we associate to every point  $\tau \in \mathbb{H}^2$  the lattice  $\Gamma_\tau$  generated by 1 and  $\tau$ , and we multiply this lattice by a suitable multiplicative factor, then denoted by  $\tilde{\Gamma}_\tau$ , to have covolume 4. Then the point  $\tau \in \mathbb{H}^2$  corresponds to the marked flat surface  $\mathbb{R}^2/\tilde{\Gamma}_\tau$ . In particular, the marked billiard surface  $S_c$  corresponds to the point  $i/c^2 \in \mathbb{H}^2$ , and the Weil-Petersson metric  $g_{WP}$  at  $z = x + iy \in \mathbb{H}^2$  agrees, up to a multiplicative factor, with the hyperbolic metric  $\frac{dx^2 + dy^2}{y^2}$  (see [6, Section 7.3.5]). The positive vertical imaginary axis in  $\mathbb{H}^2$  is therefore a Weyl-Petersson geodesic. Since this axis represents the set of all marked billiard surfaces, we can confirm in this case that the space of all marked billiard surfaces is a geodesically convex set in the Teichmüller space  $\mathbb{H}^2$ .

We finish this appendix by the remark that the restriction to *Euclidean rectangles* of area one is essential: Let us consider the bigger class of Euclidean quadrilaterals of area one (dropping the requirement that all angles are equal to  $\pi/2$ ). Then Figure 6 illustrates that the square is no longer necessarily the billiard table which minimises the total length of cyclically related closed billiard trajectories: the total length of all billiard trajectories cyclically related to the finite billiard sequence (1, 3) is obviously smaller in the parallelogram. Note also that reflections of the parallelogram along its sides does no longer lead to a tessellation of the Euclidean plane and, therefore, we cannot construct a billiard surface (flat torus) from this billiard table by the above mentioned method.

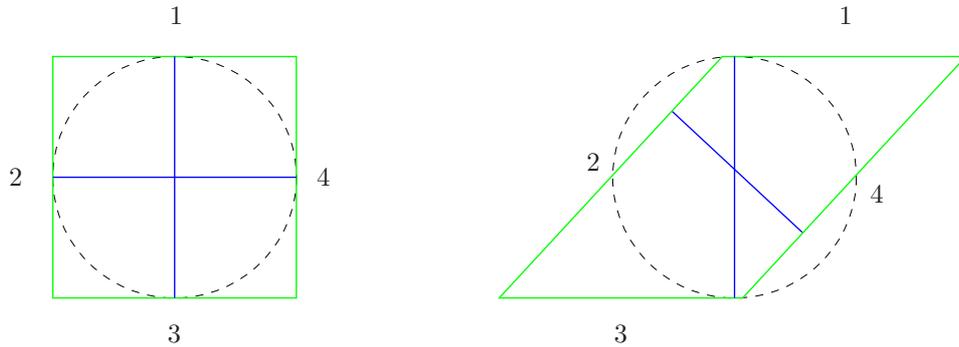


FIGURE 6. Closed cyclically related billiard trajectories to the billiard sequences  $(1, 3)$ ,  $(2, 4)$ ,  $(3, 1)$ , and  $(4, 2)$ . The billiard tables are the square and a parallelogram of area one.

REFERENCES

1. E. Artin, *Ein mechanisches System mit quasiergodischen Bahnen*, Abh. Math. Sem. Univ. Hamburg **3** (1) (1924), 170–175.
2. J. Bourgain and A. Kontorovich, *Beyond expansion II: Low lying fundamental geodesics*, J. Eur. Math. Soc. **19** (5) (2017), 1331–1359.
3. S. Castle, N. Peyerimhoff and K.F. Siburg, *Billiards in ideal hyperbolic polygons*, Discrete Contin. Dyn. Syst. **29** (3) (2011), 893–908.
4. F. Douma, English translation of E. Artin’s article [1] *Ein mechanisches System mit quasiergodischen Bahnen* at <http://www.maths.dur.ac.uk/~dma0np/>
5. J. H. Hubbard, *Teichmüller Theory and applications to geometry, topology, and dynamics*, Vol. 1, Matrix Editions, Ithaca, NY, 2006.
6. Y. Iwayoshi and M. Taniguchi, *An Introduction to Teichmüller Spaces*, Springer-Verlag, Tokyo, 1992.
7. S. P. Kerckhoff, *The Nielsen realization problem*, Ann. of Math. (2) **117** (2) (1983), 235–265.
8. C. Series, *The modular surface and continued fractions*, J. London Math. Soc. (2) **31** (1) (1985), 69–80.
9. S. A. Wolpert, *Noncompleteness of the Weil-Petersson metric for Teichmüller space*, Pacific J. Math. **61** (2) (1975), 573–577.
10. S. A. Wolpert, *On the Weil-Petersson geometry of the moduli space of curves*, Amer. J. Math. **107** (4) (1985), 969–997.
11. S. A. Wolpert, *Geodesic length functions and the Nielsen problem*, J. Differential Geom. **25** (2) (1987), 275–296.

DEPARTMENT OF MATHEMATICAL SCIENCES, DURHAM UNIVERSITY, SCIENCE LABORATORIES,  
SOUTH ROAD, DURHAM, DH1 3LE, UNITED KINGDOM  
*E-mail address:* [j.r.parker@durham.ac.uk](mailto:j.r.parker@durham.ac.uk)

DEPARTMENT OF MATHEMATICAL SCIENCES, DURHAM UNIVERSITY, SCIENCE LABORATORIES,  
SOUTH ROAD, DURHAM, DH1 3LE, UNITED KINGDOM  
*E-mail address:* [norbert.peyerimhoff@durham.ac.uk](mailto:norbert.peyerimhoff@durham.ac.uk)

TECHNISCHE UNIVERSITÄT DORTMUND, FAKULTÄT FÜR MATHEMATIK, LEHRSTUHL LS IX, VO-  
GELPOTHSWEG 87, 44 227 DORTMUND, GERMANY  
*E-mail address:* [karlfriedrich.siburg@uni-dortmund.de](mailto:karlfriedrich.siburg@uni-dortmund.de)