Ann. Sc. Norm. Super. Pisa Cl. Sci. (5)
Vol. XVIII (2018), 831-844

Hyperbolic polygons of minimal perimeter in punctured discs

JOAN PORTI

Abstract. We prove that, among the polygons in a punctured disc with fixed angles, the perimeter is minimized by the polygon with an inscribed horocycle centered at the puncture. We generalize this to a disc with a cone point and to an annulus with a geodesic boundary component and a complete end. Then we apply this result to describe the minimum of the spine systole on the moduli space of punctured surfaces.

Mathematics Subject Classification (2010): 51M16 (primary); 57M20 (secondary).

1. Introduction

Consider a complete hyperbolic disc with a puncture, *i.e.* with a cusp, $\mathbb{X}_0 = \mathbb{H}^2/\langle\gamma_0\rangle$ where $\langle\gamma_0\rangle$ is the infinite cyclic group generated by a parabolic transformation $\gamma_0 \in \operatorname{Isom}^+ \mathbb{H}^2$. Fix $n \geq 1$ and $0 < \beta_1, \ldots, \beta_n < \pi$ a family of angles. Define $\mathcal{P}$ to be the space of polygons in $\mathbb{X}_0$ with those (counterclockwise ordered) angles, that separate both ends of $\mathbb{X}_0$, and so that the cusp lies in the convex side of each angle. In Lemma 4.4 below we show that $\mathcal{P} \neq 0$, even for $n = 1$. We prove:

Theorem 1.1. *The unique minimum of the perimeter in $\mathcal{P}$ is realized by the polygon with an inscribed horocycle centered at the cusp.*

The case of a disc without any puncture (*i.e.* the hyperbolic plane $\mathbb{H}^2$) was considered in [7]. The generalization in this paper is motivated by an application to spines of minimal length of hyperbolic surfaces. A spine of a surface with finite topological type is a graph so that the surface retracts to it (for a closed surface one removes a point). Martelli, Novaga, Pluda, and Riolo [4] have shown that for each *closed* hyperbolic surface there are finitely many spines of minimal length, and their proof applies to the non compact case. Those spines are graphs with geodesic edges and with trivalent vertices, forming angles $2\pi/3$.

Partially supported by grant MTM2015–66165–P (AEI/Feder).

Received September 28, 2016; accepted in revised form February 15, 2017.
Published online May 2018.

Let $\mathcal{M}_{g,p}$ denote the moduli space of a surface of genus g with $p \geq 1$ punctures, with $p \geq 3$ when $g = 0$. The minimal length of a spine is called the *spine systole* of a surface and defines a function

$$S\colon \mathcal{M}_{g,p} \to (0, +\infty).$$

We see in Corollary 5.2 that S is a proper function.

Corollary 1.2. *The minimum of $S\colon \mathcal{M}_{g,p} \to \mathbb{R}$ is realized precisely by subgroups of the modular group,* i.e. *by surfaces $\mathbb{H}^2/\,\Gamma$ with Γ a subgroup of the congruence group $\Gamma(2)$.*

Here $\Gamma(2)$ denotes the congruence subgroup mod 2 of $\mathrm{PSL}(2, \mathbb{Z})$. When $p = 1$ those surfaces are classically called *cycloidal* [3, 5, 6]. Surfaces $\mathbb{H}^2/\,\Gamma$ with $\Gamma < \Gamma(2)$ satisfy an extremal property: there is a family of punctured horodiscs (*i.e.* punctured discs in $\mathbb{X}_0$ bounded by a horocycle), one for each cusp, whose interiors are embedded and pairwise disjoint, and whose complements are regions bounded by three horocyclic segments with tangent endpoints. In the cycloidal case ($p = 1$) there is precisely a unique such disc, which is maximal. See [3] for extremality properties of embedded discs, punctured or not, as well as [1].

In Corollary 5.1 we prove that $\min S = 3(2g - 2 + p)\log(3)$.

We shall consider a slightly more general situation, by replacing the cusp by a *cone point* of angle $\alpha \in (0, 2\pi)$ or a *geodesic* of length $r > 0$. Denote by $\mathbb{X}$ this space, and denote by $\mathbf{c}$ the cone point, the cusp, or the boundary component, according to the case we are considering. Consider again $\mathcal{P}$ the space of polygons in $\mathbb{X}$ with fixed angles $0 < \beta_1, \ldots, \beta_n < \pi$ that separate $\mathbf{c}$ from the (infinite volume) end of $\mathbb{X}$ and so that $\mathbf{c}$ lies in the convex side of each angle. If $\mathbf{c}$ is a cone point of angle α, then we need to assume furthermore that

$$\alpha + \sum_{i=1}^{n} \beta_i < n\pi\,, \tag{1.1}$$

so that $\mathcal{P} \neq \emptyset$ (see Lemma 4.4).

Definition 1.3. An *equidistant* to $\mathbf{c}$ is the following curve in $\mathbb{X}$:

- A horocycle centered at $\mathbf{c}$ when it is a cusp;
- A circle centered at $\mathbf{c}$ when it is a cone point;
- An equidistant line to $\mathbf{c}$ when it is a geodesic.

An equidistant has constant geodesic curvature κ, where $\kappa = 1, \kappa > 1$ or $\kappa < 1$ in the respective cases of the definition. The following generalizes Theorem 1.1.

Theorem 1.4. *The unique minimum of the perimeter in $\mathcal{P}$ is realized by the polygon with an inscribed equidistant to $\mathbf{c}$.*

In $\mathbb{H}^2$ a polygon is determined by the angles and edge lengths. In Lemma 4.5 we prove that this is true also for polygons in $\mathcal{P}$, in particular the position with respect to $\mathbf{c}$ is also determined by the angles and edge lengths.

The proof of Theorem 1.1 uses techniques from [7], that rely on ideas introduced in [9], with some modifications. The proof requires the Lorentz model of hyperbolic space so that several aspects of the three cases are unified. For instance $\mathbf{c}$ is represented by a point $\mathbf{x}_0$ in Lorentz space, which is lightlike for a cusp, timelike for a cone point, and spacelike for a geodesic.

Section 2 is devoted to the tools of Lorentz spaces we need. In Section 3 we construct the space of polygons $\mathcal{P}$ and we prove that it is a $(n-1)$-dimensional manifold. The main theorem is proved in Section 4 and the corollary on spines is proved in Section 5.

ACKNOWLEDGEMENTS. I am indebted to Christophe Bavard for pointing me to the reference [2], and to the referee for useful suggestions.

2. Lorentz space

The Lorentz space $\mathbb{R}^2_1$ is $\mathbb{R}^3$ equipped with the symmetric bilinear product with matrix

$$J = \begin{pmatrix} -1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

so that for $x, y \in \mathbb{R}^2_1$, $x \cdot y = x^t J y = -x^0 y^0 + x^1 y^1 + x^2 y^2$. The Lorentz model of the hyperbolic plane is then

$$\mathbb{H}^2 = \left\{ x \in \mathbb{R}^2_1 \mid x \cdot x = -1, \ x^0 > 0 \right\} .$$

From the equation $x \cdot x = -1$, the tangent space at a point is its orthogonal

$$T_x \mathbb{H}^2 = x^\perp = \left\{ y \in \mathbb{R}^2_1 \mid x \cdot y = 0 \right\} .$$

The de Sitter sphere is

$$\mathbb{S}^2_1 = \left\{ x \in \mathbb{R}^2_1 \mid x \cdot x = 1 \right\} .$$

Every point $x \in \mathbb{S}^2_1$ can be identified with an *oriented line* in $\mathbb{H}^2$

$$\left\{ y \in \mathbb{H}^2 \mid x \cdot y = 0 \right\} .$$

The orientation is provided by a normal vector. Indeed, given $x \in \mathbb{S}^2_1$, for any point p in the line x, x can be viewed as a vector in $T_p\mathbb{H}^2$ (since $x \cdot p = 0$) and x is orthogonal to the line it represents. We can also associate to x a halfplane bounded by this line

$$\left\{ y \in \mathbb{H}^2 \mid x \cdot y \leq 0 \right\} .$$

Remark 2.1. The vector $x \in \mathbb{S}^2_1$ is the outwards normal field at the boundary of the halfplane $\{y \in \mathbb{H}^2 \mid x \cdot y \leq 0\}$.

To prove this remark, given a point $y \in \mathbb{H}^2$ such that $x \cdot y = 0$, we consider the path $t \mapsto \varsigma(t) = y + t\,x + O(t^2)$, then $\varsigma(t) \cdot x = t + O(t^2)$. Hence $\varsigma'(0) = x$ and the derivative of $\varsigma(t) \cdot x$ at $t = 0$ is positive.

The light half-cone is

$$\mathbb{L} = \left\{ x \in \mathbb{R}^2_1 \mid x \cdot x = 0, x^0 > 0 \right\} .$$

Every $x \in \mathbb{L}$ can be identified with the horocycle

$$\left\{ y \in \mathbb{H}^2 \mid y \cdot x = -1 \right\} .$$

This is the boundary of the horodisc

$$\left\{ y \in \mathbb{H}^2 \mid y \cdot x \geq -1 \right\} .$$

On the other hand, the projective space on $\mathbb{L}$ can be identified with the ideal boundary $\partial_\infty \mathbb{H}^2$.

With the previous conventions, the Lorentz product is related to the incidence, see [8, Section 3.2]:

Proposition 2.2 (Incidence and Lorentz product).

(a) *Given two points $x, y \in \mathbb{H}^2$ at distance $d \geq 0$, then $x \cdot y = -\cosh d$;*
(b) *Given a point $x \in \mathbb{H}^2$ and an* oriented line *$y \in \mathbb{S}^2_1$ at distance $d \geq 0$, then $x \cdot y = \pm \sinh d$, where the sign is negative if and only if y belongs to the halfplane associated to x;*
(c) *The horocycle $x \in \mathbb{L}$ is centered at an ideal endpoint of a line $y \in \mathbb{S}^2_1$ if and only if $x \cdot y = 0$;*
(d) *The horocycle $x \in \mathbb{L}$ is tangent to the line $y \in \mathbb{S}^2_1$ if and only if $x \cdot y = \pm 1$, with negative sign if the halfplane corresponding to y contains the horodisc corresponding to y;*
(e) *If the oriented lines $x, y \in \mathbb{S}^2_1$ are disjoint at distance $d \geq 0$ ($d = 0$ means that they are asymptotic), then $x \cdot y = \pm \cosh d$, where the sign is positive when the orientations are compatible (one of the halfplanes is contained in the other);*
(f) *If the oriented lines $x, y \in \mathbb{S}^2_1$ meet at one point with angle α (taking care of the orientations), then $x \cdot y = \cos \alpha$.*

Following again [8, Section 3.2] the *Lorentzian cross product* $\boxtimes$ in $\mathbb{R}^2_1$ is defined by the rule

$$(u \boxtimes v) \cdot w = \det(u, v, w), \qquad \forall u, v, w \in \mathbb{R}^2_1,$$

where $\det(u, v, w)$ denotes the determinant of the matrix with entries the components of u, v, w. Namely $u \boxtimes v = J(u \times v)$, where $\times$ denotes the usual cross product in $\mathbb{R}^3$. In particular $(\mathbb{R}^2_1, \boxtimes)$ is a Lie algebra.

Remark 2.3. There is a natural bijection $\mathbb{R}^2_1 \leftrightarrow \mathfrak{so}(2, 1)$ that is:

- An isomorphism of Lie algebras $(\mathbb{R}^2_1, \boxtimes) \cong (\mathfrak{so}(2, 1), [,])$;
- An isomorphism of $SO_0(2, 1)$-modules, where the action on $\mathbb{R}^2_1$ is linear and on $\mathfrak{so}(2, 1)$ is the adjoint;
- A Lorentz isometry, where $\mathfrak{so}(2, 1)$ is equipped with a multiple of the Killing form.

Now fix $\mathbf{x}_0 \in \mathbb{H}^2, \mathbb{S}^2_1$, or $\mathbb{L}$. Namely $\mathbf{x}_0$ represents either a point in hyperbolic plane, an oriented line, or an ideal point (viewed projectively). Let $e_1, \dots, e_n \in \mathbb{S}^2_1$ be a collection of oriented lines.

Lemma 2.4. *The oriented lines $e_1, \dots, e_n \in \mathbb{S}^2_1$ are tangent to an equidistant to $\mathbf{x}_0$ if and only if*

$$|e_1 \cdot \mathbf{x}_0| = \cdots = |e_n \cdot \mathbf{x}_0| = \text{costant}.$$

In addition, the absolute values can be removed by taking care of orientations.

3. The space of polygons

Let $\mathcal{P}$ denote the space of polygons in $\mathbb{X}$ as in the introduction. It can be embedded in $T^1\mathbb{X} \times \mathbb{R}^n$ by looking at the tangent vector to a given edge at one of its vertices, and the edge lengths $l_1, \dots, l_n > 0$.

By convexity, the closure $\overline{\mathcal{P}}$ is obtained by considering edges of length zero or, when $\mathbf{c}$ is a cone point, by allowing a vertex or the interior of an edge to meet the cone point. In this case, $\alpha > \pi$ when $\mathbf{c}$ meets the interior of an edge, or $\alpha + \beta_i > 2\pi$ when $\mathbf{c}$ meets the i-th vertex.

As before, $\mathbf{x}_0 \in \mathbb{H}^2$ when $\mathbf{c}$ is a cone point, $\mathbf{x}_0 \in \mathbb{S}^2_1$ when $\mathbf{c}$ is geodesic, and $\mathbf{x}_0 \in \mathbb{L}$ when $\mathbf{c}$ is a cusp.

Fix e_0 an oriented line so that $e_0 \cdot \mathbf{x}_0 = 0$ and fix a point $p_0 \in e_0$ in this line. Let $g\colon (-\infty, \infty) \to \mathbb{H}^2$ denote a parametrization of e_0 so that $g(0) = p_0$ and $\{e_0, \dot{g}(0)\}$ is a positive frame in $T_{p_0}\mathbb{H}^2$. In addition, assume that:

- When $\mathbf{x}_0 \in \mathbb{H}^2$, then $\mathbf{x}_0 = p_0$;
- When $\mathbf{x}_0 \in \mathbb{S}^2_1$, then $\mathbf{x}_0 \cap e_0 = \{p_0\}$ and $\{e_0, \mathbf{x}_0\}$ is a positive frame (*i.e.* $\dot{g}(0) = \mathbf{x}_0$);
- When $\mathbf{x}_0 \in \mathbb{L}$, then $\mathbf{x}_0 \cdot p_0 = 1$ and $g(-\infty)$ is the projective class of $\mathbf{x}_0$.

See Figure 3.1.

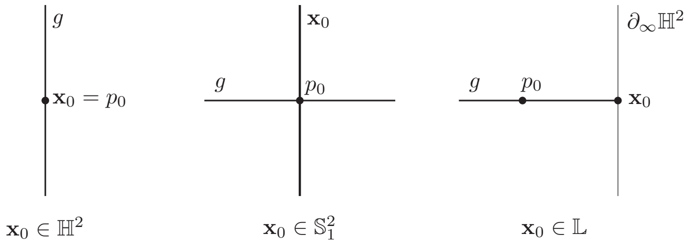

Figure 3.1. The point p_0 and the geodesic g, according to the different possibilities for $\mathbf{x}_0$.

Consider also an orientation preserving isometry $\gamma \in \mathrm{SO}_0(2,1)$ as follows:

- When $\mathbf{x}_0 \in \mathbb{H}^2$, γ is a (positively oriented) rotation of angle $\alpha \in (0, 2\pi)$ around $\mathbf{x}_0$;
- When $\mathbf{x}_0 \in \mathbb{S}^2_1$, γ is a loxodromic isometry with axis $\mathbf{x}_0$ of translation length r (in the direction $-e_0$);
- When $\mathbf{x}_0 \in \mathbb{L}$, γ is a parabolic transformation than fixes $\mathbf{x}_0$ (in the direction $-e_0$).

Chose $l_1, \ldots, l_n \in [0, +\infty)$ the lengths of the sides of the polygon. We shall also consider two parameters $l_0 \in \mathbb{R}$ and $\theta \in [0, 2\pi]/\{0 \sim 2\pi\} \cong S^1$.

We define maps v and w from the parameter spaces to the unit tangent bundle $T^1\mathbb{H}^2$ as follows. Start with the vector $\dot{g}(l_0) \in T_{g(l_0)}\mathbb{H}^2$ and rotate it by an angle θ, call this vector $v(l_0, \theta)$. This defines a map:

$$v \colon \mathbb{R} \times S^1 \to T^1\mathbb{H}^2 .$$

Then consider a polygonal path starting at $q_0 = g(l_0)$ in the direction of $v(l_0, \theta)$ that is the union of n segments of lengths $l_1, l_2, \ldots, l_n$ with ordered angles $\beta_1, \beta_2, \ldots, \beta_{n-1}$, so that at the end of the i-th edge turns left by the exterior angle $\pi - \beta_i$ and continue to the $(i+1)$-th edge. At the end of the n-th edge, consider the tangent unitary vector defining an angle β_n, *i.e.* turn left by the exterior angle $\pi - \beta_n$. This defines a map

$$w \colon \mathbb{R} \times S^1 \times \mathbb{R}^n \to T^1\mathbb{H}^2.$$

Then $\overline{\mathcal{P}}$ is contained in the set

$$\overline{\mathcal{P}} \subseteq \Big\{ (l_0, \theta, l_1, \ldots, l_n) \in I \times S^1 \times [0, +\infty)^n \mid w(l_0, \theta, l_1, \ldots, l_n) = \gamma v(l_0, \theta) \Big\}$$

where:

- $I = [0, +\infty)$ when $\mathbf{c}$ is a cone point;
- $I = (0, +\infty)$ when $\mathbf{c}$ is a geodesic;
- $I = \mathbb{R}$ when $\mathbf{c}$ is a cusp.

When $\mathbf{c}$ is a cone point of angle $\alpha < \pi$ or $\mathbf{c}$ is a geodesic, then $l_0 = 0$ is not possible by convexity.

Proposition 3.1. *The space $\mathcal{P}$ is a $(n-1)$-dimensional analytic manifold with tangent space at a point $p \in \mathcal{P}$:*

$$T_p\mathcal{P} = \left\{ (\dot{l}_0, \dot{\theta}, \dot{l}_1, \dots, \dot{l}_n) \in \mathbb{R}^{n+2} \mid \dot{l}_0(1-\gamma)e_0 + \dot{\theta}(1-\gamma)q_0 + \sum_{i=1}^{n} \dot{l}_i e_i = 0 \right\},$$

where $e_1, \dots, e_n \in \mathbb{S}^2_1$ denote the oriented lines that contain the oriented edges of p.

The unit tangent bundle $T^1\mathbb{H}^2$ is naturally identified to the isometry group $\mathrm{SO}_0(2,1)$, as the action is simply transitive. Thus the tangent space at a given point is naturally identified with $\mathfrak{so}(2,1) \cong \mathbb{R}^2_1$. In the next lemma the Lie algebras correspond to the tangent space at different points.

Lemma 3.2. *The tangent map $w_*\colon \mathbb{R}^{n+2} \to T_{w(l_0,\theta,l_1,\dots,l_n)}\mathbb{H}^2 \cong \mathfrak{so}(2,1)$ satisfies $w_*\left(\frac{\partial}{\partial l_i}\right) = e_i$ for $i = 0, \dots, n$ and $w_*\left(\frac{\partial}{\partial \theta}\right) = g(l_0) = q_0$.*

The tangent map $v_*\colon \mathbb{R}^2 \to T_{w(l_0,\theta)}\mathbb{H}^2 \cong \mathfrak{so}(2,1)$ satisfies $v_*\left(\frac{\partial}{\partial l_0}\right) = e_0$ and $v_*\left(\frac{\partial}{\partial \theta}\right) = g(l_0) = q_0$.

Proof. Increase one of the l_j by keeping the other l_k and θ constant means composing the map (either w or v) with an isometry with axis $e_i \in \mathbb{R}^2_1$, and its derivative corresponds to $e_i \in \mathfrak{so}(2,1)$ after the previous identifications of the tangent space to $T^1\mathbb{H}^2$ to $\mathfrak{so}(2,1) \cong \mathbb{R}^2_1$. The same argument applies to θ. ☐

Lemma 3.3. *Assuming that $l_0 > 0$ when $\mathbf{c}$ is a cone point, we have*

$$\langle (1-\gamma)e_0, (1-\gamma)q_0 \rangle = \mathbf{x}_0^{\perp},$$

where $q_0 = g(l_0)$.

Proof. We start checking that, for the different possibilities of $\mathbf{x}_0$,

$$\langle e_0, q_0, \mathbf{x}_0 \rangle = \mathbb{R}^2_1. \tag{3.1}$$

Namely, when $\mathbf{x}_0$ is a horocycle, we may assume up to isometry that

$$\mathbf{x}_0 = \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}, \quad e_0 = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}, \quad q_0 = \begin{pmatrix} \cosh(t) \\ \sinh(t) \\ 0 \end{pmatrix},$$

for some $t \in \mathbb{R}$. When $\mathbf{x}_0$ is a geodesic,

$$\mathbf{x}_0 = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \quad e_0 = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}, \quad q_0 = \begin{pmatrix} \cosh(t) \\ \sinh(t) \\ 0 \end{pmatrix},$$

for some $t > 0$. When $\mathbf{x}_0$ is a point in hyperbolic plane, since we assume $l_0 > 0$,

$$\mathbf{x}_0 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \quad e_0 = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}, \quad q_0 = \begin{pmatrix} \cosh(t) \\ \sinh(t) \\ 0 \end{pmatrix},$$

for some $t > 0$. This establishes (3.1). Then, since $\ker(1 - \gamma) = \langle \mathbf{x}_0 \rangle$ and γ is an isometry, the lemma follows. $\square$

Proof of Proposition 3.1. Consider M the matrix of size $3 \times (n + 2)$ with columns $M_1, \dots, M_{n+2}$, where

$$M_1 = (1 - \gamma)e_0, \quad M_2 = (1 - \gamma)q_0, \quad M_3 = e_1, \dots, \quad M_{n+2} = e_n\,. \tag{3.2}$$

We aim to show that $\operatorname{rank}(M) = 3$, so that the maps w and γv are transversal. Assume first that in the elliptic case $l_0 > 0$. By Lemma 3.3, it suffices to have that $e_i \cdot \mathbf{x}_0 \neq 0$ for some $i = 1, \dots, n$. By the incidence relations, Proposition 2.2, it is impossible that $e_i \cdot \mathbf{x}_0 = 0$ for all $i = 1, \dots, n$ (*e.g.*, when there is a cusp this would mean that all edges belong to a geodesic ending at the cusp, and similarly for the other cases).

When $l_0 = 0$ in the elliptic case, $q_0 = \mathbf{x}_0$ hence $(1 - \gamma)q_0 = 0$. In this case, since $\mathbf{x}_0 = \gamma \mathbf{x}_0$ is the starting and final point on the polygonal path, it is a closed polygon in $\mathbb{H}^2$. In particular the number of edges is ≥ 3 and they are generic enough so that $e_1, \dots, e_n$ are linearly independent in $\mathbb{R}^2_1$. $\square$

Remark 3.4. The proof of Proposition 3.1 yields that $\overline{\mathcal{P}}$ is contained in a smooth manifold of the same dimension as $\mathcal{P}$, with the tangent space described by Proposition 3.1. We shall use this to integrate tangent vectors into deformations of polygons.

4. Proof of the main theorem

The proof of Theorem 1.4 follows from the following 4 lemmas.

Lemma 4.1. *The perimeter* $\overline{\mathcal{P}} \to [0, +\infty)$ *is a proper function.*

Lemma 4.2. *A polygon in* $\mathcal{P}$ *is a critical point of the perimeter if and only if it has an inscribed equidistant.*

Lemma 4.3. *A polygon in* $\overline{\mathcal{P}} - \mathcal{P}$ *can be perturbed to* $\mathcal{P}$ *while decreasing the perimeter.*

Lemma 4.4. *There exists a unique polygon in* $\mathcal{P}$ *with an inscribed equidistant.*

Proof of Lemma 4.1. Seeking a contradiction, assume that we have a sequence of parameters in $\overline{\mathcal{P}}$ with $l_0 \to +\infty$ but $l_1, \dots, l_n \geq 0$ are bounded. This is not possible because the distance between $g(l_0)$ and $\gamma g(l_0)$ converges to infinity as $l_0 \to +\infty$, but this distance is bounded by the perimeter $l_1 + \cdots + l_n$. This establishes properness when **c** is a cone point or a geodesic. When **c** is a cusp, there could be a sequence of polygons with $l_0 \to -\infty$, while $l_1, \dots, l_n \geq 0$ are bounded. This implies that the sequence of polygons are contained in horodiscs with area going to zero, but this contradicts Gauss-Bonnet theorem: the area depends only on the angles $\beta_1, \dots, \beta_n$. $\square$

Proof of Lemma 4.2. Being a critical point means that whenever $\dot{l}_0, \dot{\theta}, \dot{l}_1, \dots, \dot{l}_n$ satisfy

$$\dot{l}_0(1-\gamma)e_0 + \dot{\theta}(1-\gamma)q_0 + \dot{l}_1 e_1 + \cdots + \dot{l}_n e_n = 0\,,$$

then $\dot{l}_1 + \cdots + \dot{l}_n = 0$. Let M be the matrix of size $3 \times (n+2)$ defined by columns as in Equation (3.2), in the proof of Proposition 3.1. By the proof of the same proposition, $\operatorname{rank}(M) = 3$. Let $\overline{M}$ be the matrix of size $4 \times (n+2)$ obtained by adding the row

$$\begin{pmatrix} 0 & 0 & 1 & 1 & \cdots & 1 \end{pmatrix}$$

to the bottom of M. Being a critical point means that $\ker M = \ker \overline{M}$, *i.e.* that $\operatorname{rank}(\overline{M}) = 3$. Set

$$\begin{pmatrix} z^0 \\ z^1 \\ z^2 \\ 1 \end{pmatrix} \in \ker\left(\overline{M}^t\right) \qquad \text{and} \qquad z = \begin{pmatrix} -z^0 \\ z^1 \\ z^2 \end{pmatrix} \neq 0\,.$$

By hypothesis

$$\big((1-\gamma)e_0\big)\cdot z = \big((1-\gamma)q_0\big)\cdot z = 0.$$

Thus, by Lemma 3.3 z is a multiple of $\mathbf{x}_0$: $z = \lambda \mathbf{x}_0$ for some $\lambda \in \mathbb{R} \setminus \{0\}$. Hence

$$l_1 \cdot \mathbf{x}_0 = \cdots = l_n \cdot \mathbf{x}_0 = -1/\lambda.$$

By Lemma 2.4, and discarding the values of λ that contradict convexity, the lemma is proved. $\square$

Proof of Lemma 4.3. We consider first the case where the cone point **c** meets a single vertex, say the first one. By convexity, $\beta_n + \alpha > 2\pi$. By the previous construction $l_0 = 0$, and we aim to deform the parameters so that l_0 increases but the perimeter decreases. When $l_0 = 0$, deforming θ does not change the resulting polygon. Thus we chose θ so that the line that bisects e_0 and γe_0 is the same that bisects e_1 and e_n but the corresponding half-lines are opposite, see Figure 4.1.

Since $\mathbf{x}_0 = p_0 = q_0$ belongs to the lines $e_0, \gamma e_0, e_1$ and e_n, we view them as tangent vectors to $\mathbf{x}_0$, *i.e.* they lie in the plane $T_{\mathbf{x}_0}\mathbb{H}^2$. Now $e_0 - \gamma e_0$ and $e_1 + e_n$ are

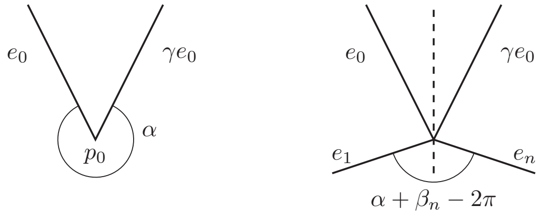

Figure 4.1. The relative position of the lines e_0 and γe_0 and the vertex of the polygon.

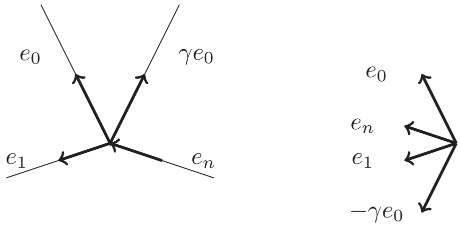

Figure 4.2. The oriented lines viewed as vectors in $T_{p_0}\mathbb{H}^2$.

both tangent vectors perpendicular to the bisector, and they both point in the same direction (see Figure 4.2):

$$(1-\gamma)e_0 = \lambda(e_1 + e_n) \qquad \text{for some } \lambda > 0.$$

Hence we may consider a deformation tangent to the vector $\dot{l}_0 = 1$, $\dot{\theta} = 0$, $\dot{l}_1 = \dot{l}_n = -\lambda$, $\dot{l}_2 = \dot{l}_3 = \cdots = \dot{l}_{n-1} = 0$. This is a vector tangent to the manifold in the equations defined in Proposition 3.1, and we have shown that this is a smooth point, see Remark 3.4. Hence the tangent vector corresponds to a deformation, and by construction it pushes the cone point to the interior of the polygon ($\dot{l}_0 > 0$) and the derivative of the perimeter is

$$\dot{l}_1 + \cdots + \dot{l}_n = -2\lambda < 0\,.$$

When **c** meets the interior of an edge, the proof is analogous by viewing an interior point as a vertex of angle π. When some of the l_i vanishes, this is precisely the content of Lemma 11 in [7]. In general, the tangent vectors to deformations can be added in order to combine the different deformations, namely pushing the cone point away from the polygon and increasing the length of edges of length 0 in the same deformation, using again Remark 3.4. □

Proof of Lemma 4.4. The existence and uniqueness is proved by gluing certain polygons.

Assume first that **c** is a cusp. For each vertex i, consider an ideal hyperbolic triangle with angles 0, $\pi/2$, and $\beta_i/2$. Double this triangle by a reflection on the edge opposite to the right angle, obtaining a quadrilateral with angles 0, $\pi/2$, β_i, and $\pi/2$, see Figure 4.3. The angles do not determine this quadrilateral, there are quadrilaterals that are non symmetric, but this is the only one whose finite edges are tangent to an horocycle centered at the ideal point. From those quadrilaterals one can construct the polygon, and it is unique by the tangency to the horocycle.

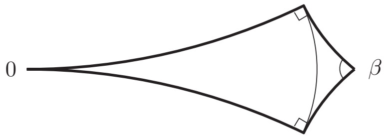

Figure 4.3. The quadrilateral in the proof of Lemma 4.4 when **c** is a cusp, with the inscribed horocycle.

When **c** is a cone point, given α_i the building block is a triangle with angles $\alpha_i/2$, $\pi/2$ and $\beta_i/2$. Double it along the long edge, to get a quadrilateral with angles α_i, $\pi/2$, β_i, and $\pi/2$, see Figure 4.4, so that the edges that meet at angle β_i are tangent to the circle centered at the vertex of angle α_i. Let $r(\alpha_i, \beta_i)$ denote the radius of this circle, which is the length of the two edges adjacent to the vertex with angle α_i. For fixed β_i the radius $r(\alpha_i, \beta_i)$ is strictly decreasing on α_i, with $r(\pi - \beta_i, \beta_i) = 0$ and $r(0, \beta_i) = +\infty$. Thus by gluing the blocs one can realize any cone angle $< n\pi - \beta_1 - \cdots - \beta_n$, in particular α by Assumption (1.1). Uniqueness also follows.

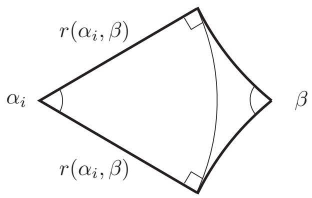

Figure 4.4. The quadrilateral in the proof of Lemma 4.4 when **c** is a cone point, with the inscribed circle.

When **c** is a geodesic, the building blocks are similar: symmetric pentagons with four right angles and one angle β_i (that is the double of a quadrilateral with three right angles and one angle $\beta_i/2$), Figure 4.5. The argument now is similar, as $r = r(d_i, \beta_i)$ is a strictly decreasing function on the length d_i of the segment opposite to β_i, $r(+\infty, \beta_i) = 0$ (approaching a triangle with two ideal vertices), and $r(0, \beta_i) = +\infty$ (approaching a triangle with one ideal vertex). □

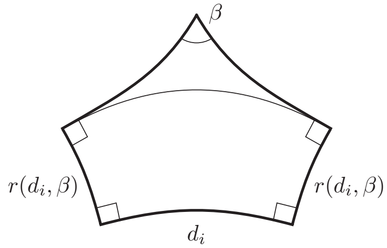

Figure 4.5. The pentagon in the proof of Lemma 4.4 when **c** is a geodesic, with the inscribed equidistant.

This concludes the proof of Theorem 1.1. Notice that Lemma 4.4 also establishes that $\mathcal{P}$ is non empty. The proof of Theorem 1.1 also shows that $\mathcal{P}$ has dimension $n-1$. One may still ask whether the edge lengths and angles determine a polygon in $\mathcal{P}$, as there are two further parameters that determine the position relative to $\mathbf{c}$.

Lemma 4.5. *A polygon in $\mathcal{P}$ is determined by its edge lengths $l_1, \dots, l_n > 0$ and angles $\beta_1, \dots, \beta_n \in (0, 2\pi)$. In particular its position relative to $\mathbf{c}$ is determined by the lengths and the angles.*

Proof. We unfold the polygon in $\mathbb{H}^2$: namely we consider a piecewise geodesic path consisting of n segments of lengths $l_1, \dots, l_n > 0$ and angles $\beta_1, \dots, \beta_{n-1} \in (0, 2\pi)$. When c is a cusp, the lemma follows because there exists a unique *oriented* parabolic isometry that joins the endpoints of this path. In fact, without taking into account the orientation there are two of them, but if we want the cusp to be in the convex side there is only one choice (two different points in $\mathbb{H}^2$ can be joined by precisely two curves of constant geodesic curvature 1). This establishes the lemma when $\mathbf{c}$ is a cusp. Notice that joining the endpoints by a parabolic isometry is a necessary condition, but not sufficient. The proof when $\mathbf{c}$ is a cone point or a geodesic is analogous, instead of parabolic isometries one must consider rotations of given angle, or loxodromic elements of given translation length, respectively. $\square$

5. Spines of minimal length

Let F be a non compact, complete, and orientable hyperbolic surface with finite topology. As said in the introduction, a spine is a graph in F so that F retracts to it, and the proof of Martelli, Novaga, Pluda, and Riolo [4] in the compact case yields the existence of spines of minimal length. Those are piecewise geodesic graphs with trivalent vertices, so that the angles are $2\pi/3$.

Proof of Corollary 1.2. The endpoints of surfaces in $\mathcal{M}_{g,p}$ are cusps, recall that $p \geq 1$. If we cut open a surface in $\mathcal{M}_{g,p}$ along a spine of minimal length, then we obtain polygons with angles $2\pi/3$ in punctured discs, one for each end of the surface. Since the perimeter is minimized by the polygon with an inscribed horocycle, this length is minimized precisely by surfaces obtained from these polygonal domains (that in particular are regular). Thus surfaces minimizing the spine systole are an orbifold covering of the 2-sphere with a puncture and two cone points of order 2 and 3 respectively, namely the modular orbifold $\mathbb{H}^2/\mathrm{PSL}(2,\mathbb{Z})$. Therefore the surfaces that minimize the spine systole are $\mathbb{H}^2/\Gamma$ for some $\Gamma < \mathrm{PSL}(2,\mathbb{Z})$. In fact $\Gamma < \Gamma(2)$, see for instance the proof of [2, Proposition A.4].

On the other hand, since the modular orbifold $\mathbb{H}^2/\mathrm{PSL}(2,\mathbb{Z})$ has a horodisc centered at the cusp whose interior is properly embedded and its closure has self-intersection precisely at the cone point of order 2, every modular surface is obtained from punctured polygonal domains with an inscribed horocycle as above. $\square$

The edge length of the polygon of angles β in a punctured disc with an inscribed horocycle is

$$2\log\frac{1+\cos\beta/2}{\sin(\beta/2)} = 2\sinh^{-1}\big(\cot(\beta/2)\big) \tag{5.1}$$

independently of the number of edges. For spines of minimal length, we are interested in $\beta = 2\pi/3$. This yields

$$2\log\frac{1+1/2}{\sqrt{3}/2} = \log(3). \tag{5.2}$$

Thus we have:

Corollary 5.1. *Let F be an orientable hyperbolic surface with finite topology, of genus g and with $p \geq 1$ ends. Then the length l of a spine in F satisfies*

$$l \geq 3(2g+p-2)\log(3),$$

with equality if and only if $F = \mathbb{H}^2/\Gamma$ for some $\Gamma < \Gamma(2) < \mathrm{PSL}(2,\mathbb{Z})$ and the spine has minimal length.

Proof. For a general surface, as its retraction to its convex core is distance decreasing, we may assume that F is a surface with boundary components and cusps. Using the constructions of Lemma 4.4, *e.g.*, Figures 4.3 and 4.5, the length of a regular polygon with angles $2\pi/3$ is bounded below by the cusped case, and the minimum is realized by surfaces $\mathbb{H}^2/\Gamma$ for some $\Gamma < \Gamma(2) < \mathrm{PSL}(2,\mathbb{Z})$. As a minimal spine is a trivalent graph, the number of edges is $-3\chi(F) = 3(2g+p-2)$. Hence the corollary follows from (5.2). □

Finally, a spine of minimal length may be nonunique, but from Lemma 4.5 we have that a shortest spine determines the surface, once we know how it embeeds. This topological information is encodded by a ribbon structure on the spine, namely a cyclic ordering on the half edges incident to each vertex.

Corollary 5.2. *A surface is uniquely determined by the spine of minimal length equipped with a ribbon structure. In particular $S\colon \mathcal{M}_{g,p} \to (0,+\infty)$ is proper.*

References

[1] C. BAVARD, *Disques extrémaux et surfaces modulaires*, Ann. Fac. Sci. Toulouse Math. (6) **5** (1996), 191–202.

[2] M. GENDULPHE, *Le lemme de Schwarz et la borne supérieure du rayon d'injectivité des surfaces*, Manuscripta Math. **148** (2015), 399–413.

[3] E. GIRONDO and G. GONZÁLEZ-DIEZ, *On extremal Riemann surfaces and their uniformizing Fuchsian groups*, Glasg. Math. J. **44** (2002), 149–157.

[4] B. MARTELLI, M. NOVAGA, A. PLUDA and S. RIOLO, *Spines of minimal length*, Ann. Sc. Norm. Super. Pisa Cl. Sci. (5) **17** (2017), 1067–1090.
[5] M. H. MILLINGTON, *On cycloidal subgroups of the modular group*, Proc. London Math. Soc. (3) **19** (1969), 164–176.
[6] H. PETERSSON, *Über einen einfachen Typus von Untergruppen der Modulgruppe*, Arch. Math. **4** (1953), 308–315.
[7] J. PORTI, *Hyperbolic polygons of minimal perimeter with given angles*, Geom. Dedicata **156** (2012), 165–170.
[8] J. G. RATCLIFFE, "Foundations of Hyperbolic Manifolds", Graduate Texts in Mathematics, Vol. 149, Springer-Verlag, New York, 1994.
[9] J.-M. SCHLENKER, *Small deformations of polygons and polyhedra*, Trans. Amer. Math. Soc. **359** (2007), 2155–2189.

Departament de Matemàtiques
Universitat Autònoma de Barcelona,
08193 Cerdanyola del Vallès, Spain
and
Barcelona Graduate School
of Mathematics (BGSMath)
porti@mat.uab.cat