PROCEEDINGS OF THE
AMERICAN MATHEMATICAL SOCIETY
Volume 127, Number 5, Pages 1347–1356
S 0002-9939(99)04778-4
Article electronically published on January 28, 1999

ISOPERIMETRIC CURVES ON HYPERBOLIC SURFACES

COLIN ADAMS AND FRANK MORGAN

(Communicated by Christopher B. Croke)

Abstract. Least-perimeter enclosures of prescribed area on hyperbolic surfaces are characterized.

1. Introduction

The isoperimetric problem of identifying the least-perimeter enclosure of given area A (henceforth called a minimizer) is solved only for a few Riemannian surfaces: the Euclidean plane, a round sphere, a round projective plane, the hyperbolic plane, a circular cone, a circular cylinder, a flat torus or Klein bottle, and recently a general surface of revolution with decreasing Gauss curvature (see the survey [HHM]). For most Riemannian surfaces the isoperimetric problem remains open.

In this paper, we will investigate the isoperimetric problem on hyperbolic surfaces. Theorem 2.2 shows that least-perimeter enclosures of prescribed area A are of four types (corresponding to the types of constant-curvature curves in hyperbolic space):

(1) circles,
(2) horocycles around cusps,
(3) boundaries of annular neighborhoods of geodesics, and
(4) collections of "neighboring curves" of geodesics.

If the surface has at least one puncture (cusp), then types (1) and (3) do not occur and the least perimeter $L \leq A$; if moreover $A < \pi$, then a minimizer is of type (2). The proof turns out to use little more than existence and regularity theory, the Gauss-Bonnet Theorem, and simple area formulas (see, e.g. Lemma 2.3). Elementary estimates then show for example that if one component of an isoperimetric region is an annulus, then it is the only component. One needs to check that certain types of candidates remain embedded.

Complete solutions to the isoperimetric problem on the once-punctured torus and the thrice-punctured sphere follow immediately. Proposition 2.4 deduces a complete solution for the four-punctured sphere.

Section 3 considers genus two surfaces without punctures and shows that there are choices of metrics for which pairs of pants and tori with boundary can and cannot occur as minimizers. In particular, the "maximal" genus two surface has only discs, annuli, and their complements as minimizers.

Received by the editors August 6, 1997.
1991 *Mathematics Subject Classification.* Primary 30F99, 53C99, 53A99, 49Q99.
Key words and phrases. Isoperimetric problem, hyperbolic surface.

Section 4 provides an algorithm to find a minimizer of prescribed area on a hyperbolic surface.

Section 5 considers higher dimensions and shows that cusp neighborhoods remain minimizing for small prescribed volumes.

We assume that all of our hyperbolic surfaces are connected and geometrically finite. Such surfaces may be compact or have finitely many ends: cusps (with exponentially shrinking thickness and finite area) or flared ends (asymptotic to the hyperbolic plane).

We would like to thank Hugh Howards for helpful conversations. This work was partially supported by National Science Foundation grants.

2. Least-perimeter enclosures in hyperbolic surfaces

The existence and regularity of least-perimeter enclosures (Lemma 2.1) follows from [HM] and plays an important role in our characterization of minimizers in Theorem 2.2.

2.1. Lemma (Existence and regularity)**.** *In a* (*geometrically finite*) *hyperbolic surface, there exists a perimeter minimizer among regions of prescribed area bounded by embedded rectifiable curves. It consists of curves of equal constant curvature.*

Proof. Suppose for some minimizing sequence, all of k components, some component goes off to infinity. If it goes out a cusp, the enclosed area goes to 0, and it may be discarded. It it goes out a flared end, it can be translated back inside a fixed compact region. Hence the results of [HM, §3], stated for compact surfaces, apply and provide the asserted existence and regularity.

The following characterization of minimizers is the main result of this paper.

2.2. Theorem. *Let S be a hyperbolic surface. For given area $0 < A < \mathrm{area}(S)$, a perimeter-minimizing system of embedded rectifiable curves bounding a region R of area A consists of a set of curves of one of the following four types* (*and all curves in the set have the same constant curvature*):

(1) *a circle,*

(2) *horocycles around cusps,*

(3) *two "neighboring curves" at constant distance from a geodesic, bounding an annulus or complement,*

(4) *geodesics or single "neighboring curves."*

The total perimeter L of a minimizer of area A satisfies

$$(*) \qquad L \leq \sqrt{A^2 + 4\pi A},$$

where equality holds precisely for a circle bounding a disc. If S has at least one cusp, then cases (1) *and* (3) *do not occur and $L \leq A$, with equality precisely for horocycles*; *if moreover $A < \pi$, then a minimizer consists of neighborhoods of an arbitrary collection of cusps bounded by horocycles, of total length A.*

Proof. By Lemma 2.1, a minimizer exists and consists of curves of constant curvature κ. Lifting to the Poincaré disc model for $\mathbf{H}^2$, a geodesic ($\kappa = 0$) is a diameter or a Euclidean circular arc normal to the boundary. Any other constant-curvature curve is determined by a choice of base point, a unit normal direction, and a positive curvature $|\kappa|$. Up to an isometry of $\mathbf{H}^2$, the base point can be chosen at the origin and the unit normal direction along the positive x-axis. Any positive-constant-curvature curve through this base point with this normal vector comes as the portion inside the unit disc of a Euclidean circle C centered on the positive x-axis and passing through the origin. If C lies within the open disc, it has $|\kappa| > 1$ and projects to a circle bounding a disc. If C is tangent to the unit circle, it is a horocycle with $|\kappa| = 1$. Otherwise C has $|\kappa| < 1$ and it is one of the two boundaries of a neighborhood of fixed hyperbolic radius about the geodesic with endpoints at the two points where C intersects the unit circle.

A minimizer cannot have more than one circle, since sliding one until it hits another (or itself) would contradict regularity. For a circle, by Gauss-Bonnet, $L\kappa = 2\pi + A$. Therefore $dL/dA = \kappa = (2\pi + A)/L$. Integrating from 0 to A yields $L^2 = A^2 + 4\pi A$. For other types of boundary curves, $dL/dA = \kappa$ is less than it is for a circle. Therefore $L \leq \sqrt{A^2 + 4\pi A}$ always holds, with equality precisely for a circle. Moreover, there is an $A_0 \geq 0$ such that if $A < A_0$, the minimizer is a circle, while if $A > A_0$, it is not a circle and (for $\Delta A > 0$)

$$\Delta L/\Delta A < 1. \qquad (**)$$

Suppose a minimizer properly includes both neighboring curves of some geodesic and the annulus between them. Let L_1 denote their total length and let A_1 denote the area of the annulus. Let L_2 denote the rest of the perimeter and A_2 the rest of the area. If $A_1 < A_0$ and $A_2 < A_0$, then

$$\begin{aligned} L(A_1 + A_2) = L_1 + L_2 &\geq \sqrt{A_1^2 + 4\pi A_1} + \sqrt{A_2^2 + 4\pi A_2} \\ &> \sqrt{(A_1 + A_2)^2 + 4\pi(A_1 + A_2)}, \end{aligned}$$

a contradiction of $(*)$. If $A_2 \geq A_0$, then by $(**)$

$$L(A_1 + A_2) \leq L(A_2) + A_1 < L_2 + L_1$$

by Lemma 2.3 below, a contradiction. Similarly if $A_2 < A_0$ but $A_1 \geq A_0$, then

$$L(A_1 + A_2) \leq L_1 + A_2 < L_1 + L_2,$$

the same contradiction.

Similarly suppose a minimizer properly includes both neighboring curves of some geodesic but not the annulus between them. If the hyperbolic surface has finite area, taking complements reproduces the previous contradiction. Otherwise a minimizer is never the complement of a circle, and for $\Delta A > 0$, $\Delta L/\Delta A > -1$. Therefore

$$L(A) \leq L(A_1 + A_2) + A_1 < L_2 + L_1 = L,$$

a contradiction. Therefore the minimizer must be of one of the four asserted types.

Henceforth assume S has a cusp. Type (1) curves cannot occur, because sliding the circle out the cusp until it hits itself would contradict regularity. Hence the minimizer always has $|\kappa| \leq 1$, and always satisfies $L(A) \leq A$. In particular, by Lemma 2.3, annuli cannot occur. Therefore complements of annuli cannot occur.

Finally assume $A < \pi$. We claim *there is no minimizer with* $-1 \leq \kappa < 1$ *and length* $L \leq A$, so $-A + \kappa L < 0$. Otherwise, applying Gauss-Bonnet to the enclosed region yields

$$2\pi\chi = -A + \kappa L < 0,$$

$\chi \leq -1$, $-A + \kappa L \leq -2\pi, \kappa L \leq -\pi, \kappa < 0, L \geq \pi > A$, a contradiction.

The remaining possibilities, systems of curves with $\kappa = 1$, consist of horocycles bounding cusp neighborhoods.

Since $\kappa = 1$, as we slide a horocycle out a cusp, $dL/dA = 1$, and therefore its length equals the area of the cusp neighborhood. By the previous claim, such systems remain minimizing as long as they exist, either for all $A < \pi$ or until they bump up against themselves at some $A_1 < \pi$. But if one bumps, by regularity the minimizer has perimeter less than A_1, contradicting the claim.

Remark. The final hypothesis $A < \pi$ is sharp. A once-punctured torus or a thrice-punctured sphere (with three cusps) has area 2π, and a least-perimeter region of area π consists of cusp neighborhoods or their complement, by the theorem.

Given $A > \pi$, there is a hyperbolic sphere with four punctures (four cusps) for which the two minimizers are of type (4), as can be seen from consideration of the limiting case of two thrice-punctured spheres with one shared puncture point.

Embedded cusp neighborhoods. Theorem 2.2 implies that any cusp has an embedded neighborhood of area greater than π bounded by a horocycle. Adams [A] has shown any cusp must have an embedded neighborhood of area at least 4 bounded by a horocycle.

The following useful lemma provides simple formulas for candidate least perimeters, given the length of the underlying geodesics.

2.3. Lemma. *Consider a neighboring curve of length L and curvature κ at distance s from a geodesic of length ℓ, enclosing area A. Then*

$$L^2 = A^2 + \ell^2,$$

$A = \ell \sinh s$, $L = \ell \cosh s$, and $\kappa = \tanh s$. For an annular region, the perimeter L satisfies $L^2 = A^2 + 4\ell^2$. For a region bounded by single neighboring curves (of type 2.2 (4)) about geodesics of total length ℓ, $L^2 = \Delta A^2 + \ell^2$, where ΔA is the difference from the area of the region bounded by the geodesics.

Proof. By Gauss-Bonnet, the curvature κ satisfies $L\kappa - A = 0$. Therefore

$$dL/dA = \kappa = A/L.$$

Integrating from 0 to A yields

$$L^2 = A^2 + \ell^2.$$

Hence for some parameter s, $A = \ell \sinh s$, $L = \ell \cosh s$, and $\kappa = \tanh s$. Since $dA/ds = L$, s is distance from the geodesic. The remaining formulas follow immediately.

We give another example.

2.4. Proposition. *On a hyperbolic 4-punctured sphere S (of area 4π) with systole ℓ (shortest simple closed geodesic length), the least-perimeter enclosure of* area A *is*

(1) *horocycles around cusps for $0 < A \le \pi + \ell^2/4\pi$ or $3\pi - \ell^2/4\pi \le A <$ area S,*
(2) *a neighboring curve for $\pi + \ell^2/4\pi \le A \le 3\pi - \ell^2/4\pi$.*

Proof. By Gauss-Bonnet, the area of S is 4π and every closed simple geodesic must enclose half the cusps and half the area. The division of cases follows from Lemma 2.3. We just need to show that the neighboring curve of the geodesic remains embedded from $\Delta A = 0$ to $\Delta A = \pi - \ell^2/4\pi$. If not, when it first bumped up against itself, by regularity it could not be minimizing, but there are no other possibilities in this range.

Remark. The preceding argument may not work for every hyperbolic surface because nonembedded curves bounding overlapping regions can be shorter than the shortest embedded curve. For example, on a high-genus, multiply-cusped hyperbolic surface with two cusps bounded by a short geodesic g, a double cover of g bounding the cusps with multiplicity two could be shorter than any embedded curve bounding a region of the same area.

3. The compact genus two surface

We will examine in more detail the possible types of minimizers one can obtain for the genus two surface. A typical metric might have minimizers of various types, as suggested by Figure 3.0. Of course small minimizers are circles, but not all minimizers are circles, because an annulus about the shortest simple closed geodesic of length ℓ, of width ℓ, beats a circle of the same area (no matter what ℓ is, by Lemma 2.3, more generally on any nontrivial unpunctured surface). Theorem 3.1 shows that for one "maximal" metric with long separating curves, the minimizers are all circles or boundaries of annuli. Perhaps other metrics, with a short, minimizing separating curve, have no annular minimizers (a separating systole greater than $2\pi\sqrt{7}$, if possible, would suffice to make circles and separating curves beat annular candidates). For some metrics pairs of pants occur, as can be seen by considering the limiting case of two thrice-punctured spheres, glued together at each of the three punctures. In any case, given the lengths of the underlying geodesics, Lemma 2.3 shows that all candidate perimeters as functions of area are hyperbolas asymptotic to lines of slope ± 1 (or for horocycles such lines themselves); see Figure 3.0.

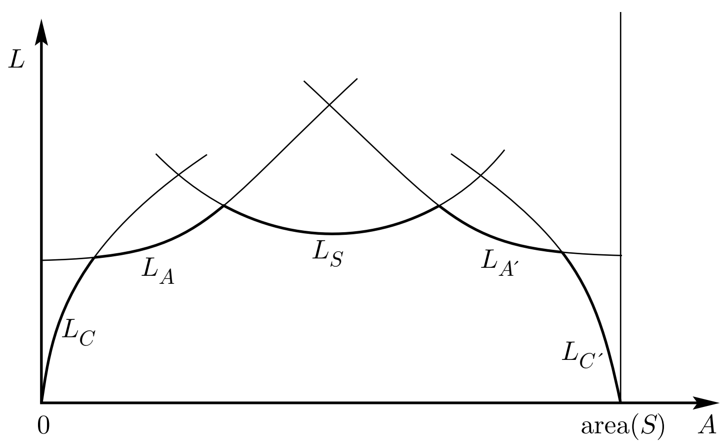

Figure 3.0. A speculative graph of least perimeter L as a function of prescribed area A for a compact genus two surface. By Lemma 2.3, the candidate perimeters, here circles of length L_C, boundaries of annuli of length L_A, and separating curves of length L_S, are hyperbolas asymptotic to lines of slope ± 1.

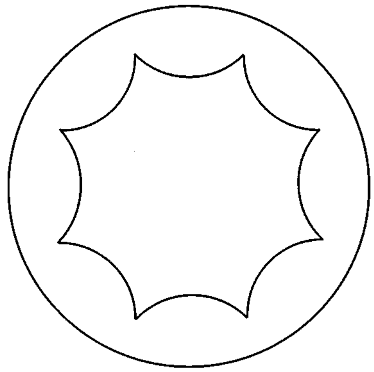

FIGURE 3.1. The fundamental domain of the maximal surface of genus two is an octagon with interior angles $\pi/4$ and opposite sides, all of length $3.057\ldots$, identified.

3.1. Theorem. *There exists a "maximal" hyperbolic metric on the genus two surface such that all minimizers are a disc, an annulus, or the complement of a disc or annulus.*

Proof. The *systole* of a hyperbolic metric on a surface is defined to be a shortest closed simple geodesic. P. Schmutz [S, Thms. 5.2, 5.3] gives a proof of the previously known fact that the length of a systole for any choice of hyperbolic metric on the compact genus two surface is bounded above by $\ell = 2\text{arccosh}(1+\sqrt{2}) = 3.057\ldots$ and then shows there is a unique surface realizing this systole length, called the *maximal surface*. This surface contains 12 systoles which together cut the surface into 16 equilateral triangles with angles $\pi/4$. The standard fundamental domain is the regular octagon of Figure 3.1, with all angles $\pi/4$, and opposite sides, all of length $3.057\ldots$, identified.

Since any annulus minimizer has core geodesic of length at least ℓ, 2.2($*$) and Lemma 2.3 imply that the circle beats the annulus for $A < \ell^2/\pi$. For $\ell^2/\pi < A < 2\pi$, the length of the boundary of the annulus increases to $2\sqrt{\pi^2+\ell^2} = 8.767\ldots$. Over this range of areas, the only possible competitors are a pair of pants or a nontrivial separating curve. A pair of pants has boundary length at least as large as the sum of the lengths of three geodesics, which is at least $3(3.057\ldots)$. By Lemma 3.2 below, a separating geodesic must have length at least $9.027\ldots$. Since the annulus beats both of these, by regularity, it must remain embedded over this entire range of areas. For $2\pi < A < 4\pi$, the complement of the annulus and of the disc become the minimizers.

3.2. Lemma. *A separating geodesic on the maximal genus two surface has length at least* $9.027\ldots$.

Proof. In the standard fundamental domain of Figure 3.1, the geodesic made up of four arcs, each connecting a distinct pair of edges that are separated by one edge on the boundary of the domain and each intersecting the boundary of the domain perpendicularly, has length $9.027\ldots$. Suppose that there exists a shorter separating geodesic. Realize it as a set of arcs properly embedded in the standard fundamental domain. We can assume that the arcs miss the vertices of the fundamental domain by choosing, if necessary, a different one of the six points where the systoles of the

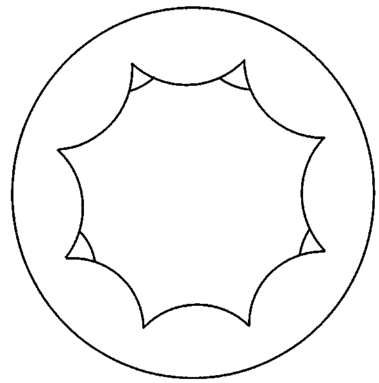

FIGURE 3.2. This pattern cannot occur for a geodesic.

surface intersect in fours to correspond to the vertices of the domain. (A separating geodesic of length less than 9.027... cannot intersect all six of the intersection points.)

If an arc cuts a triangle off the fundamental domain, call it a corner arc, and otherwise call it a long arc. First note that a long arc cuts off a component containing two proper subarcs of boundary edges and one, two, or three entire boundary edges. The shortest long arc cutting off one entire boundary edge is perpendicular to the boundary arcs at its endpoints and cuts a quadrilateral from the octagon. Bisecting the quadrilateral with an arc perpendicular to the long arc cuts the quadrilateral into two Lambert quadrilaterals (see §3.3), from which one can compute that the long arc has length at least 2.2567.... Four copies of a long arc realizing this lower bound give the separating geodesic of length 9.027.... The shortest possible long arc cutting off two entire boundary edges cuts a pentagon from the octagon with three adjacent $\pi/4$ angles and two right angles. Adding a geodesic arc connecting the two nonadjacent vertices with $\pi/4$ angles cuts an isosceles triangle off the pentagon. Hyperbolic trigonometry determines its edge lengths and angles. Bisecting the remaining quadrilateral by an arc that is perpendicular to the long arc and the new arc yields Lambert quadrilaterals from which the length of the long arc can be seen to be 2.881 .... The shortest possible long arc cutting off three entire boundary edges must pass through the center of the octagon and is a systole of length 3.057 ....

Thus, our geodesic separator can contain at most three long arcs. Moreover, it can pass through at most three consecutive corner arcs, as can be seen by noting that more than three corner arcs would span an angle of more than π around their common vertex in the universal cover, making it impossible for a geodesic to pass through them all. In the fundamental domain this rules out a separating geodesic which includes the pattern of corner arcs of Figure 3.2.

The cases of one long arc and three long arcs can be eliminated using the fact that each boundary edge must be separated from its opposite by an even number of arcs. Listing all of the possibilities for combinatorial patterns with two long arcs that satisfy the restrictions yields just two cases, as appear in Figure 3.3.

For the first such pattern, if the fundamental domain is decomposed into sixteen equilateral triangles all with angles $\pi/4$, by chopping the domain open along arcs

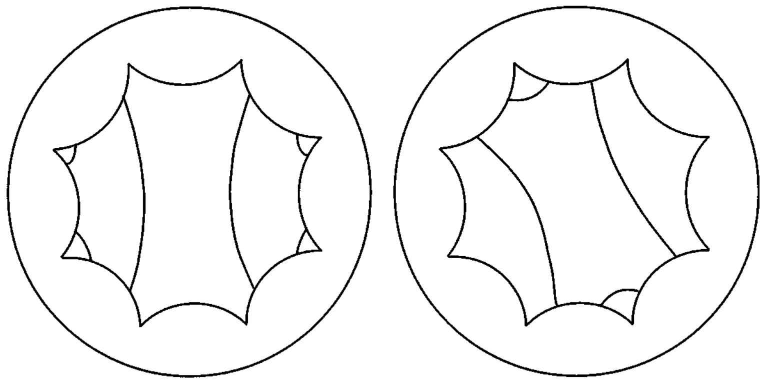

FIGURE 3.3. There are two possible patterns for shortest separating geodesics with two long arcs.

connecting the midpoints of opposite edges and arcs connecting the midpoints of adjacent edges, this geodesic is realized by a set of isometric geodesic segments, each of which cuts across one of sixteen of the triangles. Each segment has endpoints on two of the three edges of a given triangle, intersecting one edge perpendicularly and intersecting the other edge at its midpoint. Each such segment has length $1.5537\ldots$, and the geodesic realizing this pattern has a length of $24.86\ldots$.

In the decomposition of the fundamental domain into equilateral triangles, the geodesic corresponding to the second combinatorial pattern is realized by a set of isometric geodesic segments, each of which cuts across one of twelve of the triangles. Each segment has endpoints on two of the three edges of one of the triangles and its endpoints bisect each edge. This forces the length of each segment to be $1+\sqrt{2}$, and the corresponding geodesic has a length of $12(1+\sqrt{2})$, again longer than $9.027\ldots$.

3.3. Lambert quadrilaterals. A *Lambert* (geodesic) *quadrilateral* has three right angles and one other angle φ. If the two edges adjacent to the angle φ have lengths b_1, b_2 and the opposite edges have lengths a_1 and a_2 respectively, then the following two formulas hold:

$$\sinh(a_1)\sinh(a_2) = \cos\varphi,$$
$$\cosh(a_1) = \cosh(b_1)\sin\varphi.$$

4. An algorithm for finding minimizers

Theorem 4.1 provides an algorithm for finding minimizers on a hyperbolic surface.

4.1. Theorem. *Given a hyperbolic metric on a surface S and an area $0 < A < \text{area}(S) \leq \infty$, the following algorithm finds all least-perimeter regions of area A.*

(1) *Determine all embedded geodesics of length less than $\sqrt{A^2 + 4\pi A}$ (see Remark below). By Theorem 2.2, the only possibilities for a least-perimeter region are discs, neighborhoods of cusps (all of total length A), annular neighborhoods of such geodesics, a region bounded by neighboring curves of three or more such geodesics, or complements. (In the cusped case, you may use A for the upper bound on the length of the geodesics and exclude discs and annuli.)*

(2) *Compute the perimeter for each possibility and order them from smallest perimeter to largest.* (*Given the lengths of underlying geodesics, such computations are easy, as shown by Lemma 2.3; cf. Figure 3.0.*) *The solution will be the first embedded possibility.* (*We do not know whether the shortest possibility is embedded in general.*)

(3) *If the least-perimeter region is a disc or cusp neighborhood, it is automatically embedded and a solution. If not, first use Lemma 2.3 to compute the common distance s from each geodesic to its neighboring curve. Second determine the shortest perpendicular length between any pair of geodesics or between any geodesic and itself. If this length is greater than $2s$, the region is embedded and a solution. Otherwise move on to the next region on the list.*

Remark. Step (1) in the algorithm can be realized computationally as follows. Starting with a fundamental domain for the surface in the hyperbolic plane, tile the hyperbolic plane with copies of the domain to cover a neighborhood of the original domain of radius $\sqrt{A^2 + 4\pi A}$. Any closed geodesic in the surface of length less than $\sqrt{A^2 + 4\pi A}$ must lift to a geodesic segment in H^2 which intersects the initial tile and which is therefore contained in this neighborhood. Hence all such geodesics must lift to axes of hyperbolic isometries that are products of generating isometries that glue edges to edges in this set of tiles. If the surface has punctures or flared ends, one cannot tile such a neighborhood with a finite number of tiles. However, if one tiles to cover a neighborhood of radius $\sqrt{A^2 + 4\pi A}$ about the original domain minus the lifts of a disjoint set of flared ends and cusps, all closed geodesics of length less than $\sqrt{A^2 + 4\pi A}$ will lift to geodesic segments in this neighborhood.

Proof of Theorem 4.1. The theorem follows immediately from the characterization of minimizers of Theorem 2.2. In step (3), a disc, if the shortest possibility, is the minimizer by 2.2($*$) and hence embedded by regularity; similarly for a cusp neighborhood in a cusped surface.

5. Higher dimensions

In n-dimensional hyperbolic manifolds M, least-perimeter enclosures of prescribed volume exist by the methods of geometric measure theory (at least if the volume is finite so that no volume disappears to infinity in the limit) and are smooth constant-mean-curvature hypersurfaces except for a singular set of Hausdorff dimension at most $n - 8$ [M, pp. 66, 90, 87; cf. pp. 128–131].

The following theorem shows that in cusped manifolds, cusp neighborhoods remain minimizing for small volumes.

5.1. Theorem. *Let M be a cusped n-dimensional hyperbolic manifold of finite volume. For some $\varepsilon > 0$, the least-perimeter enclosure of a region of volume $V \leq \varepsilon$ is an arbitrary collection of horosurfaces around cusps, of total area $A = (n-1)V$.*

Proof. Since horosurfaces around cusps have mean curvature $H = n-1$, as you slide a horosurface out a cusp $dA/dV = n - 1$, so $A = (n - 1)V$. For a constant-mean-curvature hypersurface with $H \leq n - 1$, through any point where the embedding radius of M is bounded below, the area is bounded below (by "monotonicity" [M, pp. 90–91]). Hence we can choose $\varepsilon > 0$ so that any minimizer with $V \leq \varepsilon$ and $|H| \leq n - 1$ must lie deep in the cusps.

Suppose for some $0 < V \leq \varepsilon$ there is a minimizer S other than cusp neighborhoods. Choose such to maximize $(n - 1)V - A$. Then $H = n - 1$, or varying V

slightly would increase $(n-1)V - A$. Therefore S lies deep in the cusps. Now moving horosurfaces down the cusps until first contact with S contradicts the maximum principle. (At a point on S of first contact the tangent cone to S must be a plane and S must be regular.)

References

[A] Colin Adams, *Maximal cusps, collars and systoles for hyperbolic surfaces*, Indiana Univ. Math. J., to appear.

[HM] Joel Hass and Frank Morgan, *Geodesics and soap bubbles in surfaces*, Math. Z. **223** (1996), 185–196. MR **97j:**53009

[HHM] Hugh Howards, Michael Hutchings, and Frank Morgan, *The isoperimetric problem on surfaces*, Amer. Math. Monthly, to appear.

[M] Frank Morgan, *Geometric Measure Theory: a Beginner's Guide*, Academic Press, second edition, 1995. MR **96c:**49001

[S] P. Schmutz, *Riemann surfaces with shortest geodesic of maximal length*, Geom. and Func. Anal. **3** (1993), 564–631. MR **95f:**30060

Department of Mathematics, Williams College, Williamstown, Massachusetts 01267

E-mail address: colin.adams@williams.edu

E-mail address: frank.morgan@williams.edu