TRANSACTIONS OF THE
AMERICAN MATHEMATICAL SOCIETY
Volume 352, Number 11, Pages 4889–4909
S 0002-9947(00)02482-X
Article electronically published on July 12, 2000

THE ISOPERIMETRIC PROBLEM ON SURFACES OF REVOLUTION OF DECREASING GAUSS CURVATURE

FRANK MORGAN, MICHAEL HUTCHINGS, AND HUGH HOWARDS

ABSTRACT. We prove that the least-perimeter way to enclose prescribed area in the plane with smooth, rotationally symmetric, complete metric of nonincreasing Gauss curvature consists of one or two circles, bounding a disc, the complement of a disc, or an annulus. We also provide a new isoperimetric inequality in general surfaces with boundary.

1. INTRODUCTION

Although it is well known that in the plane the least-perimeter way to enclose given area is a circle, the isoperimetric problem has remained open even in very simple surfaces, and the following result was just proved in 1996 by Benjamini and Cao [BeC, Thms. 5, 8].

1.1. Theorem (Benjamini and Cao). *The least-perimeter way to enclose given area in the paraboloid of revolution*

$$\{z = x^2 + y^2\} \subset \mathbf{R}^3 \tag{1}$$

is a circle of revolution.

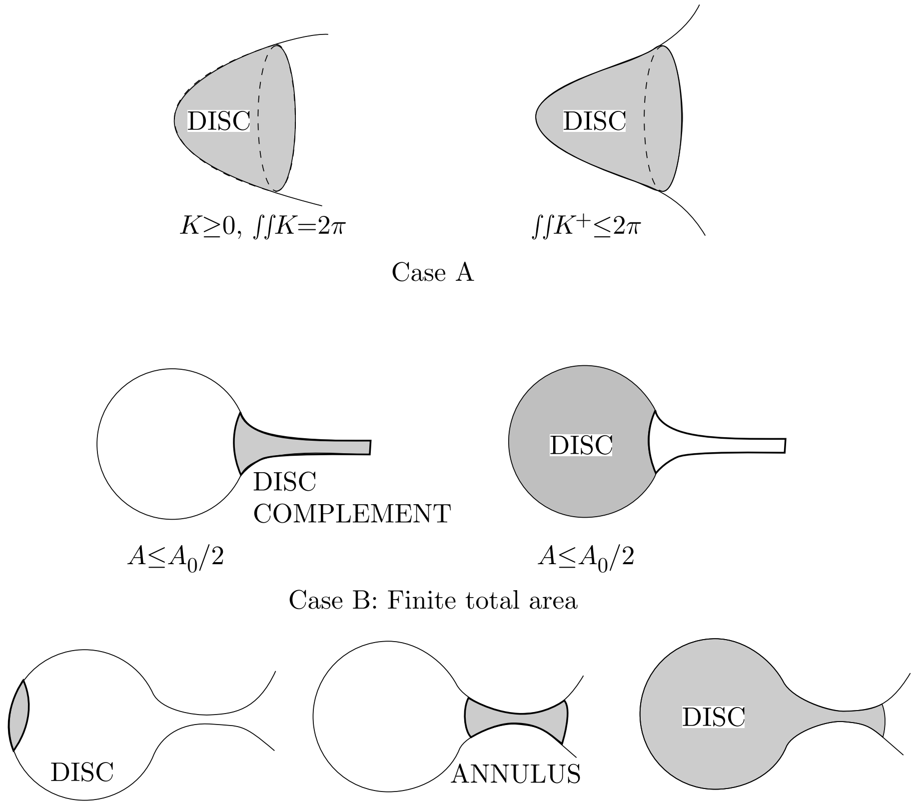
It was known that a solution should want to stay near the origin, where the Gauss curvature K is big. In fact, if a curve lies in a region where $K \leq K_0$, then its length L and area A satisfy the Bol-Fiala inequality [Os, 4.25]

$$L^2 \geq 4\pi A - K_0 A^2. \tag{2}$$

Unfortunately, the symmetrization arguments that work in the plane do not work on the paraboloid. We will give a simple proof of our more general Theorem 3.1, which we discovered unaware of the work of Benjamini and Cao [BeC] and more recent work of Pansu [Pan, esp. Prop. 7], Topping [T2, T1], and Ritoré [Rit]. The following Theorem 1.2 summarizes some results of Theorem 3.1.

1.2. Theorem. *Consider the plane with smooth, rotationally symmetric, complete metric such that the Gauss curvature K is a strictly decreasing function of the distance r from the origin. Then a least-perimeter enclosure of prescribed area exists and must consist of one or two circles about the origin, bounding a disc, the complement of a disc, or an annulus. If the Gauss curvature is positive or the total Gauss curvature of every compact region is at most 2π, then the minimizer is a disc. If the total area is finite, then the minimizer is a disc or disc complement.*

Received by the editors July 10, 1998 and, in revised form, November 1, 1998.
2000 *Mathematics Subject Classification.* Primary 53Cxx, 53Axx, 49Qxx.
Key words and phrases. Isoperimetric problem.



Case C: Infinite total area (includes Case A)

FIGURE 1. In a smooth, complete surface of revolution, with decreasing Gauss curvature K, a least-perimeter enclosure of prescribed area must consist of one or two circles about the origin, bounding a disc, the complement of a disc, or an annulus. (These surfaces cannot in general be isometrically embedded in $\mathbf{R}^3$.)

See Figure 1. The easiest Case A of curvature $K \geq 0$, or more generally (as it turns out) with the integral of the positive part of K at most 2π, includes the paraboloid of revolution. In the general case, an annulus can occur.

Benjamini and Cao assume total Gauss curvature of regions at most 2π. Pansu considers mainly disc-type regions. Our Theorems 3.1, 3.3, 3.5, and 3.6 also consider regions in a sphere, a real projective plane, a closed disc, a plane which is not complete, and an annulus. Theorem 3.7 considers overlapping regions. Section 4 provides applications to comparison theorems, eigenvalues, and the minimal volume of $\mathbf{R}^2$.

All approaches involve integrating the Gauss-Bonnet Theorem as the prescribed area shrinks to 0. Benjamini and Cao (cf. [T2]) shrink by curvature-flow curve shortening after Grayson and Andrews [And]. Pansu (cf. [T1]) shrinks curves at a constant rate after the classical argument of Fiala. We consider a (not necessarily continuous) family of perimeter minimizers, as in some previous studies of the isoperimetric profile (see e.g. [BaP], [K]). For the existence and regularity of a minimizer inside a large ball, we use only the new, simple theory of [HM], not the machinery of geometric measure theory.

The basic idea is that a smooth perimeter-minimizing disc has constant geodesic curvature $\kappa = dL/dA$, the rate of change of length with respect to area. By the Gauss-Bonnet formula,

$$\left(\frac{1}{2}L^2\right)' = LL' = \int \kappa = 2\pi - \int K.$$

Therefore minimizers prefer large Gauss curvature K. One complication is that in general minimizers may bump up against themselves.

Theorem 2.1 thus shows that a minimizer in the category of unions of disjoint discs is a single round disc about the origin. Theorem 3.1 then uses a nesting argument to improve any enclosure to circles about the origin. Differential inequalities involving length, area, Gauss curvature, and geodesic curvature eventually identify discs, disc complements, and certain annuli as the only minimizers.

Along the way we (cf. [T1]) provide a new isoperimetric inequality (2.4) in any smooth Riemannian surface S, perhaps with boundary, subsuming the Bol-Fiala inequality (2). Consider a region of c components, Euler characteristic χ, perimeter L, and area A. Then

$$L^2 \geq 4\pi(\chi - c + 1)A - 2\int_0^A G(t)\,dt,$$

where $G(A)$ denotes the supremum of the total Gauss curvature over regions of area A. Equality holds for a disc about the origin in a surface of revolution of nonincreasing Gauss curvature.

This work was partly inspired by a more difficult question we heard from J. C. C. Nitsche about the soap film between a large wire boundary and a small, moveable loop of thread. The thread wants to position itself to minimize the area of the soap film outside it. If the thread were constrained to lie in a fixed surface bounded by the wire (which unfortunately is not the case), then the thread would want to be an isoperimetric curve in that surface.

Nitsche asked whether the thread always touches the wire. We found that Spivak [Sp, Vol. 4, p. 449] reports on experimental evidence that it does, but our own experiments with R. Kusner and J. Sullivan indicate that it does not.

For a recent survey of these and other new and old results, see "The isoperimetric problem on surfaces" [HHM].

We would like to thank Colin Adams, Gene Calabi, Pierre Pansu, and Renato Pedroza for helpful conversations. Morgan's work is partially supported by a National Science Foundation grant and was partly carried out while he was 250-Anniversary Visiting Professor for Distinguished Teaching at Princeton University.

2. Isoperimetric discs are round discs about the origin

Th following new Isoperimetric Theorem 2.1 for certain surfaces of revolution considers only regions comprised of embedded discs. The proof integrates the Gauss-Bonnet theorem over a family of perimeter-minimizing discs. Pansu [Pan, Prop. 7] and Topping [T1] independently obtain almost the same result with a different proof, using equidistant curves and real-analytic approximation after Fiala. The new Isoperimetric Inequality 2.4 holds in any smooth Riemannian surface with or without boundary. Sections 3 and 4 will not depend on the latter sections 2.4, 2.6, 2.7.

2.1. Isoperimetric Theorem (embedded discs). *Consider a plane, sphere, real projective plane, or closed disc S with smooth, rotationally symmetric metric such that the Gauss curvature is a nonincreasing function of the distance from the origin. Then among disjoint unions of embedded discs of given area, a round disc centered at the origin minimizes perimeter.*

It is unique, except of course that a circle in the interior of a ball about the origin of constant Gauss curvature may be replaced by a congruent circle in that ball.

We need not assume that the plane is complete. The total area may be finite or infinite. For smoothness, it is sufficient that the metric be C^1 and piecewise C^2. It follows that a C^2 curve has a well-defined, continuous geodesic curvature, that the Gauss curvature is piecewise continuous, and that the Gauss-Bonnet formula holds. We may even allow the origin to be a conical point with positive contribution to the total Gauss curvature.

Remarks. A sphere of revolution of nonincreasing Gauss curvature must be a (round) sphere of constant Gauss curvature. Suppose there were another of area 4π, so that $G(4\pi) = 4\pi$ and for $0 < A < 4\pi$, $G(A) > A$, where $G(A)$ denotes the total Gauss curvature of a disc of area A centered at the origin. Then equality in 2.3(3) below for circles about the origin implies that $L(A)^2$ hits 0 before A reaches 4π, a contradiction. (Section 3.6 has some results on more general spheres.)

There are, however, smooth real projective planes of revolution of decreasing Gauss curvature, including for example ellipsoids with antipodal points identified.

Our proof depends on Lemmas 2.2 and 2.3. Lemma 2.2 applies some of the new existence and regularity results of [HM].

2.2. Lemma. *In a smooth compact, connected Riemannian surface S, perhaps with convex boundary, given $0 < A <$ area S, $c_0 > 0$, there is a perimeter minimizer among collections of $c \leq c_0$ disjoint C^1 embedded discs of total area A, except that the boundary of a disc may bump up against itself at finitely many points or along finitely many geodesic arcs, still remaining C^1. It may also bump up against ∂S. Elsewhere the perimeter curve C has constant geodesic curvature κ_0. For A bounded and c_0 large, no component bumps up against itself on the inside, and the geodesic curvature κ satisfies $\kappa \leq \kappa_0$ everywhere (weakly).*

These results extend to more general categories of subsurfaces of bounded topological type.

For use in 2.4–2.6, we note that if ∂S is merely piecewise smooth and piecewise convex, the regularity can fail only at the exceptional points, where the additional contribution to $\int \kappa$ (in the Gauss-Bonnet formula) is nonpositive.

We assume S connected to avoid the degenerate case when a perimeter-minimizing region is a component of S, its perimeter lies entirely in ∂S, and κ_0 is undefined.

Proof of Lemma 2.2. The existence and regularity results of [HM, 3.1, 3.3] extend immediately to surfaces with convex boundary and provide a minimizer, possibly with bumping as asserted, C^1 away from ∂S, constant curvature κ_0 away from bumping. (As long as the boundary is convex, it does not interfere with the proof, a convexification argument.) Two discs will not bump up against each other, or they could profitably be combined into one disc.

If $\kappa_0 > 0$, the curve could bump up against ∂S, even on a Cantor set of positive measure. A small inward variation at a boundary point, decreasing the area A enclosed and possibly decreasing the length L, must satisfy $dL/dA \leq \kappa_0$; otherwise a simultaneous variation elsewhere to compensate for the loss in area with $dL/dA = \kappa_0$ would produce a net decrease in length. Thus $0 \leq \kappa \leq \kappa_0$ weakly on ∂S.

Ordinarily if $\kappa_0 < 0$, a component could bump up against itself on the inside with $\kappa = 0 > \kappa_0$. Now suppose that $A \leq A_0$ and that c_0 is large. If the minimizer actually has c_0 components, one component must have small area. By Gauss-Bonnet, κ_0 is positive and there is no inner bumping. Likewise if the minimizer has fewer than c_0 components, no component can bump up against itself on the inside, since otherwise we could cut the boundary curve where it bumps and re-glue, to obtain two curves which violate regularity. In any case, $\kappa \leq \kappa_0$.

Finally we consider an isolated point p where ∂S fails to be smooth and convex. Since in a deleted neighborhood of p C has bounded curvature, only finitely many pieces of C reach p.

If ∂S is not convex at p, there may be several nested pieces of C with alternating orientations and bounded curvature except perhaps for discontinuous changes at p in the unit tangent of magnitudes

$$\alpha_1 \leq -\alpha_2 \leq \alpha_3 \leq \cdots \leq 0.$$

The additional contribution to $\int \kappa$ is $(\alpha_1 - \alpha_2) + (\alpha_3 - \alpha_4) + \cdots$ and hence nonpositive. (If ∂S has an angle of $-\pi$ at p, there is the additional innocuous possibility of two pieces of curve bumping each other on the outside there, but remaining C^1.)

Lemma 2.3, a standard kind of result (see [K, p. 46]), will let us integrate the key differential inequality in the proof of Theorem 2.1.

2.3. Lemma. *Let f and g be continuous functions on $[a, b]$. Suppose the lower left derivative of f satisfies*

$$f'_{L_*}(t) = \liminf_{\Delta t \to 0^-} \frac{f(t + \Delta t) - f(t)}{\Delta t} \geq g(t). \tag{1}$$

Then $f(b) - f(a) \geq \int_a^b g(t)\, dt$.

Proof. It suffices to consider the case $g = 0$, $[a, b] = [-1, 0]$. It suffices to show that for all x in $[-1, 0]$, $f(0) - f(x) \geq \varepsilon x$. If not, there is a least $x_0 > -1$ such that $f(0) - f(x) \geq \varepsilon x$ for x in $[x_0, 0]$. Now (1) yields a smaller x_0, the desired contradiction.

Remark. The continuity hypothesis on g may be relaxed to uppersemicontinuity from the left:

$$\limsup_{x \to x_0^+} g(x) = g(x_0).$$

We may allow $g(x) = -\infty$ (and $f'_{L_*} = +\infty$) as long as $f'_{L_*} > -\infty$ and $g < +\infty$. Since any L^1 function can be approximated within ε in measure from below by an uppersemicontinuous function bounded from above, and from above by a lowersemicontinuous function bounded from below (Vital-Carathéodory [Rud, 2.24]), it follows that a continuous BV function with (one-sided) upper derivative *everywhere* less than $+\infty$ and (one-sided) lower derivative everywhere greater than $-\infty$

is absolutely continuous (cf. [Rud, 8.11]). The need to assume BV is shown by sums of examples such as $x^2 \sin x^{-2}$ [Rud, 8.20a]. Of course absolutely continuous implies BV.

Incidentally, the continuity hypothesis on f in Lemma 2.3 may be relaxed to lowersemicontinuity from the right.

Proof of Theorem 2.1. By replacing the plane by a large closed disc, we may assume that S is compact. By attaching continuously differentiably an annulus of very negative constant curvature, we may assume ∂S is convex. (Such annuli are given by metrics of the form $dr^2 + P(r)^2\,d\theta^2$, $0 \le r \le r_0$, $\theta \in \mathbf{R}/\mathbf{Z}$, $P(r) = ae^{kr} + be^{-kr}$, $k > 0$. By differentiating (4) below, the Gauss curvature $K = -P''/P = -k^2$. The parameters a and b may be chosen to keep the metric C^1; for k large, a and b are positive. The geodesic curvature $\kappa(r)$ of a circle of revolution satisfies $\kappa(r) = dP/dA = P'/P$, which approaches k as $r \to \infty$, so that if r_0 is large, the boundary is convex.)

For $0 < A < A_1$, c_0 large, let $L(A)$ denote the length of a shortest curve of $c \le c_0$ components, enclosing c disjoint discs of total area A, which exists by Lemma 2.2. $L(A)$ is continuous by a compactness argument. Recall that one geometric interpretation of geodesic curvature is the rate of change of length with respect to area under perturbations of the given minimizer (see [M2, Chapt. 2]). For $\Delta A < 0$, the new minimizers must do at least as well as perturbations of the old one, and the lower left derivative (see 2.3(1)) satisfies $L'_{L_*}(A) \ge \kappa_0$.

By the Gauss-Bonnet formula, the total Gauss curvature of the enclosed region equals

$$2\pi\chi - \int \kappa \ge 2\pi - L(A)\kappa_0 = 2\pi - L(A)L'_{L_*}(A).$$

Let $G(A)$ denote the total Gauss curvature of a disc of area A centered at the origin. Since the Gauss curvature is a nonincreasing function of radius, any other region with the same area has no more total Gauss curvature. So we have

$$\begin{aligned} &2\pi - L(A)L'_{L_*}(A) \le G(A), \\ &L(A)L'_{L_*}(A) \ge 2\pi - G(A). \end{aligned} \tag{2}$$

By Lemma 2.3,

$$L(A_1)^2 \ge 2\int_0^{A_1} L(A)L'_{L_*}(A) \ge 4\pi A_1 - 2\int_0^{A_1} G(A). \tag{3}$$

On the other hand, the right-hand side of (3) equals the length squared of a circle of area A_1 centered at the origin, as we can see by integrating the Gauss-Bonnet formula for circles centered at the origin with area A from 0 to A_1. Therefore a round disc centered at the origin minimizes perimeter.

To prove uniqueness, note that equality holds in (3), L is differentiable everywhere, and equality holds in (2). Therefore a minimizer encloses Gauss curvature $G(A)$ and must be a circle about the origin, at least unless C lies strictly inside a maximal set E of constant Gauss curvature K_0. Any other minimizer C' must lie in E. If r denotes distance from the origin, then $E = \{r_1 \le r \le r_2\}$ for some $0 \le r_1 < r_2 \le \infty$.

Of course if $r_1 = 0$, then round circles in E about other points yield other minimizers. Conversely, the only constant-curvature curves in E are such circles.

Suppose $r_1 > 0$. We will show that tangential stretching of E by some factor $\lambda > 1$ is also contained in a surface of revolution of nonincreasing Gauss curvature, and then we will derive a contradiction. Let $P(r)$ and $\kappa(r)$ denote the perimeter and geodesic curvature of a circle about the origin bounding a disc D. Since $dP/dr = (dP/dA)(dA/dr) = \kappa P$, by the Gauss-Bonnet formula,

$$P' = \kappa P = 2\pi - \iint_D K, \tag{4}$$

and consequently,

$$P'' = -KP. \tag{5}$$

Choose a small interval $[r_3, r_4] \subset (0, r_1)$ on which K decreases and $|K| > 0$. We may assume $|P'| > 0$, because on any subinterval where $P' = 0$, $K = 0$ by (5). On $[r_3, r_4]$, we may view K as a function of P, and (5) becomes

$$dP'^2/dP = -2PK(P). \tag{6}$$

Hence by altering K slightly in $(P(r_3), P(r_4)]$ to a still nonincreasing function $\widetilde{K}$ we can obtain a new initial piece of surface $\widetilde{S}$ with $\widetilde{P}'$ at $\widetilde{P} = P(r_4)$ arbitrary in a neighborhood of $P'(r_4)$. Choose $r_5 > r_4$ very close to r_4 with $\lambda = P(r_4)/P(r_5) > 1$. Let $S(r_5)$ denote the portion of S with $r \geq r_5$ stretched tangentially by a factor λ (with circumference λP and the same Gauss curvature and other local properties as S). We may select $\widetilde{S}$ such that $\widetilde{S}$ and $S(r_5)$ fit together continuously differentiably to form a new surface S_1 of nonincreasing Gauss curvature.

Tangential stretching f of E by the factor λ maps it into S_1. This map f stretches C (which is completely tangential) by the factor λ onto a round circle in S_1. This same map f would stretch any other minimizer C' (which is not completely tangential) by a factor less than λ. Moreover, f stretches all area by the factor λ. Thus in comparison with the round circle $f(C)$, $f(C')$ would be shorter and enclose the same area, the desired contradiction. We conclude that C is unique.

If the Gauss curvature decreases discontinuously at say r_5, the uniqueness argument simplifies: for $\widetilde{S}$ just take a slight homothetic expansion by $\lambda > 1$ of $\{r \leq r_5\}$. The Gauss curvature of the final S_1 will still be nonincreasing, with a smaller jump down at r_5.

Remark. As noted in [BeC, pp. 361–362], the hypothesis of nonincreasing Gauss curvature is necessary. Take a unit ball in hyperbolic space and attach a flat conical end. Then the ball has greater perimeter than a Euclidean disc of the same area in the conical end.

2.4. Isoperimetric Inequality. *Let S be a smooth Riemannian surface, perhaps with boundary. Let $G(A)$ denote the supremum of the total Gauss curvature of a region of area A. Then the length L of the boundary of a region of c components, Euler characteristic χ, and area A satisfies*

$$L^2 \geq 4\pi(\chi - c + 1)A - 2\int_0^A G(t)\, dt. \tag{1}$$

Equality holds for a disc about the origin in a surface of revolution of nonincreasing Gauss curvature.

Remarks. Inequality (1) subsumes the Bol-Fiala inequality 1.1(2) for a disc and its generalization [Os, (4.30), (4.29)] to connected regions of lower Euler characteristic. Inequality (1) has been given independently by Topping [T1].

Proof. The proof parallels the proof of Theorem 2.1. First we treat the case of S connected. By replacing S by a large, compact, connected subsurface, we may assume S compact. If S has boundary, we may assume the boundary piecewise smooth and piecewise convex, by adding a thin collar to S which alters $\int G(A)$ by less than any given $\varepsilon > 0$.

For $0 < A < A_1$, and c_0 large, let $L(A)$ denote the infimum of perimeters of embedded regions with $\chi' - c' + 1 \geq \chi - c + 1$ and $c' \leq c_0$. By Lemma 2.2, a minimizer exists (possibly with some bumping). (Degeneracies from curves shrinking away do not decrease $\chi' - c'$.) If the minimizer actually has c_0 components, one component has only short curves, all of which lie in a disc bounded by one of them. By Gauss-Bonnet, κ_0 is nonnegative and there is no inside bumping. Likewise if the minimizer has fewer than c_0 components, there is no inside bumping, since otherwise we could cut the boundary curve where it bumps and re-glue (which would not decrease $\chi' - c'$, although it might increase c'), to obtain two curves which violate regularity. Outside bumping occurs only if κ_0 is positive. In any case, $\kappa \leq \kappa_0$. Hence by the Gauss-Bonnet formula,

$$\begin{aligned} L(A)L'_{L_*}(A) &\geq 2\pi\chi' - G(A) \geq 2\pi(\chi' - c' + 1) - G(A) \\ &\geq 2\pi(\chi - c + 1) - G(A). \end{aligned}$$

By Lemma 2.3, integrating yields (1). Equality holds as asserted.

If S is not connected, (1) holds for subregions of perimeter L_i and area A_i in components of S. Since $L^2 \geq \sum L_i^2$ and

$$\int_0^{\sum A_i} G(A) \leq \sum \int_0^{A_i} G(A),$$

it suffices to show that

$$\sum (\chi_i - c_i + 1)A_1 \geq (\chi - c + 1) \sum A_i$$

or that $\chi_i - c_i + 1 \geq \chi - c + 1$, i.e., that $\sum_{j \neq i}(\chi_j - c_j) \leq 0$, which holds because $\chi_j \leq c_j$.

2.5. Proposition. *Let S be a smooth Riemannian disc with piecewise smooth and piecewise convex boundary. Suppose that the total Gauss curvature of any region is at most 2π. In a category of regions of bounded topological type, a perimeter-minimizing region must be a union of discs. The minimum perimeter L is a nondecreasing function of the area A.*

Proof. By Lemma 2.2, in a category of regions of prescribed area and bounded topological type, a perimeter-minimizing region R exists and its boundary has constant geodesic curvature κ_0 (except for bumping). Since at bumps $\kappa \leq \kappa_0$, applying the Gauss-Bonnet formula to any boundary component C and the disc it bounds yields $\kappa_0 \geq 0$. Hence $L'_{L_*}(A) \geq 0$ and L is nondecreasing (Lemma 2.3). It follows that a minimizer must be a union of discs, or deleting an interior curve would increase A and decrease L.

The following corollary of the Isoperimetric Inequality 2.4 extends the Bol-Fiala isoperimetric inequality 1.1(2) to hyperbolic space $\mathbf{H}^n$ and other ambient manifolds of dimension greater than two. We want to thank P. Pansu for a helpful discussion.

2.6. Corollary (higher codimension). *Let M be a smooth, complete n-dimensional Riemannian manifold with sectional curvature $K \leq K_0$. Then for a (possibly branched) minimal disc of area A and boundary of length L,*

$$L^2 \geq 4\pi A - K_0 A^2. \tag{1}$$

Proof. By approximation we may assume that the boundary is smooth, and hence the mapping $f : D \to M$ is smooth up to the boundary, perhaps with finitely many branch points where Df vanishes to finite order [HH]. By the Gauss equations, the pull-back metric on D has Gauss curvature at most K_0 (though undefined at the branch points). In the almost conformal coordinates provided by f, with metric $ds^2 = g\, ds_0^2$ vanishing at the branch points, the Gauss curvature is given by

$$K = -(1/2g)\Delta \log g = \frac{|\nabla g|^2 - g\Delta g}{2g^2}.$$

Adding a small positive constant ε to g yields a nonvanishing smooth metric with curvature

$$K_\varepsilon = -(1/2(g+\varepsilon))\Delta \log(g+\varepsilon) = \frac{|\nabla g|^2 - g\Delta g - \varepsilon\Delta g}{2(g+\varepsilon)^2}.$$

On $\{\Delta g \leq 0\}$,

$$K_\varepsilon \leq K \leq K_0.$$

On $\{\Delta g > 0 \text{ and } g \geq \sqrt{\varepsilon}\}$,

$$K_\varepsilon \leq \frac{g^2}{(g+\varepsilon)^2} K \leq K_0 + o(1),$$

where $o(1)$ denotes a quantity which tends to 0 as $\varepsilon \to 0$. Finally, on the small set $E = \{\Delta g > 0 \text{ and } g < \sqrt{\varepsilon}\}$, $K_\varepsilon \leq \frac{g^2}{(g+\varepsilon)^2} K$. If K_0 is positive, $K_\varepsilon \leq K_0$. If K_0 is nonpositive, $K_\varepsilon \leq 0$ and for any region R,

$$\int_{R\cap E} K_\varepsilon \leq 0 \leq \int_{R\cap E} K_0 + |K_0| \text{ area } E \leq \int_{R\cap E} K_0 + o(1).$$

Therefore in 2.4 the associated function $G_\varepsilon(A)$ satisfies

$$G_\varepsilon(A) \leq K_0 A + o(1),$$

$$-2\int_0^A G_\varepsilon(t)\, dt \geq -K_0 A^2 + o(1),$$

and 2.4(1) yields 2.6(1).

Remarks. The existence of least-area minimal discs, under the additional hypothesis of curvature and injectivity radius bounded below, is provided by Morrey [Mo, Thm. 9.4.3]. Of course, a surface minimizing area in a larger category still satisfies (1) if $K_0 \leq 0$.

An alternative proof of Corollary 2.6, integrating a differential inequality for L^2 as in 2.1 and 2.4, would require the existence and regularity of a least-perimeter least-area disc D of area A. Existence holds for M compact. For regularity, to show that ∂D has bounded curvature, we would like to know that the first variation δD is nonzero (somewhere at its boundary), but we do not see how to prove that!

The following Proposition 2.7 yields and upper bound on least perimeter from a curvature bound. Bray [Br] has already applied such arguments to surfaces in three-dimensional manifolds. For a generalization to general dimension, see [MJ, Thm. 3.5].

2.7. Upper bound on least perimeter. *Let S be a smooth, closed two-dimensional Riemannian surface of Gauss curvature $K \geq K_0$ and area A_0. For prescribed area $0 < A < A_0$, the least perimeter L satisfies*

$$L^2 \leq 4\pi A - 2\int_0^A G_0(t)\,dt \leq 4\pi A - K_0 A^2, \tag{1}$$

even among connected regions. Here $G_0(A)$ denotes the infimum of the total Gauss curvature of any region of area A.

Proof. The proof parallels the proofs of 2.1 and 2.4. Among connected regions of area A (and Euler characteristic $\chi \leq 1$), a perimeter minimizer exists, possibly with bumping. (Cf. Lemma 2.2. Existence without an *a priori* bound on the number of boundary components follows as in [HM, Thm. 3.4]. We did not use this result of [HM] before in proving isoperimetric inequalities, because its proof depends on having some kind of isoperimetric inequality.)

Any bumping must be on the interior, since re-gluing at any exterior bumping would leave the region connected and contradict regularity. The geodesic curvature κ is a constant κ_0, except at interior bumping, where $\kappa = 0$. If there is interior bumping, $\kappa_0 < 0$. In any case, $\kappa_0 \leq \kappa$ and the upper right derivative of the least perimeter function $L(A)$ satisfies

$$\frac{1}{2}(L^2)^{*\prime}_R \leq \kappa_0 L \leq \int \kappa = 2\pi\chi - \int K \leq 2\pi - G_0(A).$$

By the analog of Lemma 2.3,

$$\frac{1}{2}L^2 \leq 2\pi A - \int_0^A G_0(t)\,dt \leq 2\pi A - \frac{1}{2}K_0 A^2.$$

Remark. If equality holds in the first inequality of (1) for some A_1, then S has constant curvature K_0.

Proof of remark. Suppose that the curvature K_2 at some point p satisfies $K_2 > K_1 > K_0$. The perimeter L_1 of a small geodesic disc about p satisfies the well-known classical formula $L_1^2 = 4\pi A - K_2 A^2 + o(A^2) < 4\pi A - K_1 A^2$ for A small. Meanwhile, for A small, any minimizer is bounded by short curves, κ_0 is large, and there can only be one curve, bounding a disc D. Since equality holds in the first inequality of (1) for A_1 and hence for all $A \leq A_1$, therefore $K < K_1$ on D. By the Isoperimetric Inequality 2.4 applied to the surface D,

$$L^2 \geq 4\pi A - K_1 A^2 > L_1^2,$$

the desired contradiction.

3. Isoperimetric regions are discs, disc complements, and annuli

Our Main Theorem 3.1 seeks isoperimetric curves of any finite number of components in a surface of revolution of nonincreasing Gauss curvature. See Figure 1. Theorem 3.3 treats $\mathbf{R}P^2$ in greater detail. Corollary 3.5 and Theorem 3.6 consider the annulus and more general spheres. Theorem 3.7 considers overlapping regions.

3.1. Main isoperimetric theorem. *Consider the plane S with smooth, rotationally symmetric, complete metric such that the Gauss curvature K is a nonincreasing function of the distance r from the origin. Then a least-perimeter enclosure of prescribed area exists and must consist of one or two circles about the origin, bounding a disc, the complement of a disc, or an annulus.*

A. *If the Gauss curvature is nonnegative or the total Gauss curvature of every compact region is at most 2π, then the perimeter of a circle about the origin is a nondecreasing function of the radius, the total area is infinite, and the minimizer is a disc.*

B. *If the total area A_0 of the plane is finite, then the disc complement is minimizing for prescribed area $0 < A \leq A_0/2$, the disc is minimizing for $A_0/2 \leq A < A_0$, and there are no other minimizers.*

C. *If the total area is infinite, then a minimizer is either a disc, or an annulus with boundary of positive geodesic curvature.*

As in Theorem 2.1, there is one trivial exception to uniqueness. If a circle lies in the interior of a ball about the origin of constant Gauss curvature, it may of course be replaced by a congruent circle in that ball.

The first paragraph with the added possibility of the annulus complement holds as well for a plane which is not complete, a closed disc (with boundary), a sphere, or a real projective plane (see 3.3), except that if the outer boundary of a minimizing annulus coincides with the boundary of the closed disc, the outer boundary could have negative curvature.

As remarked after Theorem 2.1, for smoothness, it is sufficient that the metric be C^1 and piecewise C^2.

Remarks. Example 3.4 shows that minimizing annuli can occur in case C. Without the hypothesis of decreasing Gauss curvature, even existence may fail. Consider a surface which for large r comes from revolving $y = 1/x$ about the x-axis. Annuli going off to infinity enclose given area with arbitrary small perimeter. (By differentiating (1) below, this surface has Gauss curvature $K = -P''/P = -2x^6/(1+x^4)^2$).

Proof. We begin with the easy Case A, then prove the first paragraph in the general case with the help of Lemma 3.2, and then take up the other cases. First note that by the Gauss-Bonnet formula, the perimeter $P(r)$ and geodesic curvature $\kappa(r)$ of a circle about the origin of radius r bounding a disc D satisfy

$$P' = \kappa P = 2\pi - \iint_D K. \tag{1}$$

Case A. Note that if K is nonnegative, then $\iint K \leq 2\pi$, since otherwise by (1) the perimeter $P(r)$ of a circle about the origin eventually satisfies $P' \leq -a < 0$ and P hits 0. Now by (1), $P'(r)$ is nonnegative, P is nondecreasing, and the total area is infinite. The rest follows immediately from Proposition 2.5 and Theorem 2.1. (Of course since the total area is infinite, any candidate lies in some large disc.)

General case. Consider any collection of disjoint simple closed curves enclosing area A_1. Each curve may be nested a number of levels from infinity. Curves at the same level bound disjoint discs. Form a new enclosure as follows: for each level, replace the union of all the curves at that level with a single circle about the origin that encloses the same total area. By Theorem 2.1, the new enclosure, consisting of m circles, is shorter.

In the category of collections of at most m circles about the origin, a minimizer exists. Indeed, if the total area is finite, no area can disappear to infinity in the limit. If the total area is infinite, we may assume that K goes negative, since otherwise Case A applies. Now the length $P(r)$ of the circle approaches infinity with the radius, because $P' = \kappa P$ approaches infinity by (1). (If the plane is not complete and has infinite area, then $\limsup P = \limsup P' = \infty$, and by (1) K goes negative; hence $\lim P' = \infty$ and $\lim P = \infty$.)

The circles must have the same geodesic curvature, alternating in sign: $\kappa, -\kappa, \kappa, -\kappa, \ldots$ (except that if the last circle coincides with ∂S, it may have smaller curvature). If $\kappa = 0$, no two circles can lie on a flat cylindrical segment, or moving them left (cf. Figure 1) would eventually reduce perimeter. Since the Gauss curvature is nonincreasing, it follows from (1) that there are at most three circles, because κ changes sign at most twice. If there is one circle, the minimizer is a disc or disc complement. If there are two circles, the minimizer is an annulus (of nonnegative geodesic curvature by Lemma 3.2(C)) or complement (to be ruled out in Cases B and C). If there were three circles, then κ would be positive, and replacing the annulus between the first two circles (regardless of whether it is part of the enclosed region or its complement) by a disc of the same area would reduce perimeter by Lemma 3.2(C).

Case B: *the total area is finite.* In this case there are at most two circles, of curvature $\kappa_0 > 0$, $-\kappa_0$, because κ changes sign only once; otherwise by (1) the limit as r approaches infinity of P' is positive and the total area would be infinite. Two such circles are ruled out by Lemma 3.2(C), so there is actually just one circle.

We finish this case by comparing the disc and the disc complement. Since κ changes sign only once, by Lemma 3.2(A) $\kappa' < 0$; P increases to a geodesic circle and then decreases. Since $\kappa = dP/dA$, it follows from Lemma 3.2(B) that up through the area A_1 of the disc bounded by the geodesic circle, the disc complement of the same area has smaller perimeter. By complementation, for area greater than $A_0 - A_1$, the disc has smaller perimeter than the disc complement of the same area. Hence the only two candidates are a disc complement for $0 < A \leq A_0 - A_1$ and a disc for $A_1 \leq A < A_0$ (and of course $A_1 \leq A_0/2$). Since the first has nonnegative geodesic curvature and the second has nonpositive geodesic curvature $\kappa = dP/dA$, the disc complement continues to have smaller perimeter precisely until the unique area for which their perimeters agree. (Note that a circle of curvature κ has curvature $-\kappa$ as the boundary of a region to its right.) By symmetry, this occurs when $A = A_0/2$. Therefore the disc complement has smaller perimeter than the disc when $0 < A \leq A_0/2$, and the disc has smaller perimeter than the disc complement when $A_0/2 \leq A < A_0$.

Case C: *the total area is infinite.* Case C follows from the general case and the trivial observation that since S has infinite area, there is no minimizing disc complement or annulus complement.

3.2. Lemma.

A. *The geodesic curvature $\kappa(r)$ of a circle about the origin has at most one interval $[r_1, r_2]$ of minima ($0 < r_1 \leq r_2 \leq \infty$). For $0 < r < r_1$, $\kappa'(r) < 0$, and for $r_2 < r, \kappa'(r) > 0$. If $P(r)$ goes to 0, then $\kappa'(r)$ remains nonpositive.*

B. *Consequently the perimeter $P(r)$ has at most one interval I_{max} of maxima and then at most one minimum. For given perimeter, the circle on the left (cf. Figure 1) of I_{max} has larger $|\kappa|$ than the circle between I_{max} and any minimum.*

C. *An annulus $\{r_3 \leq r \leq r_4\}$ with $\kappa(r_4) \leq -\kappa(r_3) < 0$ has greater perimeter than the disc about the origin of the same area.*

Proof. A. Of course initially κ' is negative. Let r_1 be a first zero of κ'. Since $dP/dr = (dP/dA)(dA/dr) = \kappa P$, differentiating 3.1(1) yields

$$\kappa' P + \kappa^2 P = -PK, \quad \kappa' = -K - \kappa^2, \tag{1}$$

(the Riccati equation [Ch2, (3.40)] or the second variation formula [BaP, §7]). On any interval $[r_1, R)$, $|\kappa|$ is bounded by some $M > 0$, so that

$$\kappa'' = (-K - \kappa^2)' = -2\kappa\kappa' - K' \geq -2M|\kappa'| - 0.$$

It follows that κ' remains nonnegative and, once positive, stays positive. Indeed, if $\kappa' > 0$, then $(\log \kappa')' \geq -2M$, so $\log \kappa'$ cannot go to $-\infty$ and κ' stays positive. Similar if κ' becomes negative, it was always negative. (If the metric is merely C^1 and piecewise C^2, the above inequality should be written $\Delta(\kappa') \geq -2M|\Delta\kappa|$. Both $-K$ and κ' may have discontinuous increases. r_1 is the first place where the right-hand derivative of κ is nonnegative.)

If $P(r)$ approaches 0, then eventually $\kappa(r)$ is negative. Suppose that eventually $\kappa'(r) > 0$. If $P(r) \to 0$ in finite time, $\log P \to -\infty$, $\kappa = (\log P)' \to -\infty$, a contradiction. If $P(r) \to 0$ as $r \to \infty$, by (1) κ' is increasing, a contradiction.

B. Since $dP/dr = (dP/dA)(dA/dr) = \kappa P$, differentiating 3.1(1) yields

$$\frac{d\kappa}{dr} P + \kappa^2 P = -PK,$$

$$\frac{d\kappa/dr}{dP/dr} = -\frac{\kappa^2 + K}{\kappa P},$$

$$d\log P = -\frac{\kappa\, d\kappa}{\kappa^2 + K}. \tag{2}$$

Hence, as you move away from I_{max}, towards circles on either side, for given $|\kappa|$ the circle on the right is shorter.

C. If $\kappa(r)$ is nonincreasing for $r \leq r_4$, then moving the annulus to the left, keeping area constant, decreases perimeter, eventually yielding a disc of less perimeter. Therefore we may assume that $\kappa(r)$ has a minimum before r_4. Suppose such an annulus has no greater perimeter than the disc about the origin of the same area, and choose such an annulus to maximize $\kappa(r_3)$. The curvature of the boundary of a disc of the same area must be greater than $-\kappa(r_3)$, or by (B) its perimeter would be less. Reduce r_3 slightly to produce an annulus of slightly greater area and less perimeter than a disc of the same area. Since $\kappa(r_3)$ increases, by choice of our annulus, now $-\kappa(r_3) < \kappa(r_4) < 0$. Continuously reduce r_3 and r_4, keeping the enclosed area constant. As long as $-\kappa(r_3) < \kappa(r_4)$, the perimeter decreases. Since $\kappa(r_3)$ is increasing as r_3 decreases, at no time can $-\kappa(r_3) = \kappa(r_4)$, by choice of the original annulus. Hence the process continues until you get a disc, with less perimeter than the annulus, a contradiction.

Remarks. Incidentally, on $[r_1, r_2]$, K is constant. If K is eventually constant and $P \to 0$, then κ is eventually the constant $-\sqrt{-K}$. Indeed, by 3.1(1), $P'' = -PK, P = Ce^{-r\sqrt{(-K)}}$, $\kappa = P'/P = -\sqrt{-K}$.

If $P(r) \to \infty$, $\kappa'(r)$ might remain negative (e.g., $P(r) \sim r^{1/2}$) or might become positive (e.g., $P(r) \sim e^{r^2}$).

3.3. Real projective plane. *Consider the real projective plane $\mathbf{R}P^2$ with smooth, rotationally symmetric metric, such that the Gauss curvature K is a nonincreasing function of the distance r from an origin. Let A_0, A_1 denote the total area and the area inside the first geodesic circle about the origin.*

A. *If $A_1 \geq A_0/2$, then the round disc about the origin is minimizing for $A \leq A_0/2$, the disc complement (Möbius strip) is minimizing for $A \geq A_0/2$, and there are no other minimizers.*

B. *If $A_1 < A_0/2$, then there is an $0 < A_2 < A_1$ such that the round disc is minimizing for $A \leq A_2$ and $A_0/2 \leq A \leq A_0 - A_2$, and the disc complement (Möbius strip) is minimizing for $A_2 \leq A \leq A_0/2$ and $A_0 - A_2 \leq A$. There are no other minimizers, unless there is a flat cylindrical segment, in which case both the disc and the disc complement are minimizing for $A_2 \leq A \leq A_0 - A_2$.*

As in Theorem 2.1, there is one trivial exception to uniqueness. If a circle lies in the interior of a ball about the origin of constant Gauss curvature, it may of course be replaced by a congruent circle in that ball.

As remarked after Theorem 2.1, for smoothness, it is sufficient that the metric be C^1 and piecewise C^2.

Proof. By Theorem 3.1, a minimizer exists and consists of a disc or annulus about the origin, or complement. The initial nesting argument, however, deserves some further comment in the case of $\mathbf{R}P^2$. Given any collection of disjoint simple closed curves enclosing area A_1, a typical curve bounds a disc on one side and a Möbius strip on the other side. There could however be one curve which is not a boundary (any two would have to intersect); just discard it. Now the discs are nested: discs not contained in other discs at the first level, discs contained only in level-1 discs at the second level, and so on.

Note that the points farthest from the origin comprise a geodesic circle. By Lemma 3.2(A, C), a minimizer cannot be an annulus (or complement). That leaves just the disc and disc complement, which we now compare by cases.

Case A. If $A_1 = A_0$, by Lemma 3.2(A) the perimeter of a circle about the origin is nondecreasing and the geodesic curvature nonincreasing all the way to the final (geodesic) circle. In this subcase, the result follows easily.

Therefore we may assume $A_0/2 \leq A_1 < A_0$. Since $A < A_1$, the disc boundary lies to the left of the first large geodesic circle C_0; the disc complement boundary lies to the right, but encloses less area with C_0. Since $|dA/dP| = |1/\kappa|$ is greater on the right by Lemma 3.2(B), the change in perimeter from C_0 must be less, and the disc complement has more perimeter. Therefore for $0 < A < A_0/2$, the disc has less perimeter than the disc complement. By complementation, for $A_0/2 < A < A_0$, the disc complement has less perimeter than the disc. Of course they tie at $A = A_0/2$.

Case B. If there is a flat cylindrical segment, it must extend from the first long geodesic circle onward, and the theorem follows easily. Henceforth we will assume there is no such flat cylindrical segment.

The disc is minimizing for small prescribed area A, but at $A = A_1 < A_0/2$ the disc complement has less perimeter. Let A_2 be the first area for which the disc complement and disc have the same perimeter. Since $|dA/dP| = |1/\kappa|$ is greater on the right by Lemma 3.2(B), the disc complement continues to beat the disc up through A_1 and then obviously past A_1 until they coincide at $A_0/2$. The case $A \geq A_0/2$ is complementary.

3.4. Annular region example. We now give an example of a surface S of revolution with decreasing Gauss curvature and an area $A > 0$, for which an annulus has less perimeter than the disc about the origin. One could argue that such examples must occur because surfaces of Case B of Theorem 3.1 with hyperbolic cusps and minimizing disc complements rather than discs arise as limits of surfaces of Case C with hyperbolic flared ends with thin necks. Instead we give an explicit example.

For convenience of construction we'll take Gauss curvature piecewise constant: S will be a hyperbolic flared end with a spherical cap, as in Case C of Figure 1.

For a large, consider the portion $\{x \geq -a\}$ of the cylinder C with hyperbolic metric

$$ds^2 = dx^2 + \frac{\cosh^2 x}{\cosh^2 a}\, d\theta^2 \qquad (\theta \text{ in } \mathbf{R}/\mathbf{Z})$$

and cap it off with a large northern spherical cap whose boundary circle has the same length $L = 1$ and the same geodesic curvature $\kappa = L'/L$ or equivalently the same $L' = \tanh a$. This is possible because $\tanh a < 1 < 2\pi$. This metric is C^1 and piecewise C^2.

The portion of the cylinder between -1 and 1 has perimeter

$$P_1 = 2(\cosh 1)/(\cosh a)$$

and area

$$A = \int_{x=-1}^{1} \frac{\cosh x}{\cosh a}\, dx = 2\frac{\sinh 1}{\cosh a}.$$

A small northern spherical cap of area A has perimeter P_2 satisfying

$$P_2 > \sqrt{2\pi A} > P_1$$

by choice of a large.

3.5. Corollary of Theorem 3.1 (the annulus). *Consider an annulus S (without boundary) with smooth, rotationally symmetric, complete metric of nonincreasing Gauss curvature. Suppose that the left end has finite area. Then the total area is infinite and a circle of revolution minimizes perimeter for given area.*

Proof. Let $-\infty < r < \infty$ denote distance from left to right perpendicular to circles of revolution. By Gauss-Bonnet, for $r_1 < r_2$,

$$P'(r_1) = P(r_1)\kappa(r_1) = P(r_2)\kappa(r_2) + \int_{r_1}^{r_1} K\, dA.$$

Suppose K were nonnegative on some interval $(-\infty, r_0]$. Since $P(r) \to 0$ as $r \to -\infty$, there would be an $r_2 \leq r_0$ with $\kappa(r_2) > 0$. Then for all $r_1 < r_2$, $P'(r_1) > P(r_2)\kappa(r_2) > 0$, which is impossible. Therefore since K is nonincreasing, K is negative. Consequently $P'(r)$ is increasing and $P'(r_1)$ must approach 0 as r approaches $-\infty$. In particular, for small r_1, $P'(r_1) < 2\pi = (2\pi r)'$, and we may replace $\{r \leq r_1\}$ by a small spherical cap to obtain a plane of revolution $S(r_1)$.

Now any curve C, perhaps of several components, enclosing prescribed area in the annulus S, lies in some $S(r_1)$. By Theorem 3.1(A), C is at least as long as a circle of revolution enclosing the same area.

Remark. Under the hypotheses of Corollary 3.5, the geodesic curvature $\kappa(r)$ of a circle of revolution satisfies $\kappa'(r) \geq 0$, similarly to Lemma 3.2(A).

Remark. The completeness hypothesis is necessary. For example, a circle of revolution in an annulus in a flat circular cone with vertex solid angle greater than 2π is unstable. The proof breaks down because we cannot cap off the left end with a spherical cap.

3.6. Theorem (the sphere)**.** *Let S be a smooth sphere of revolution with left-right reflective symmetry and Gauss curvature nonincreasing half way. If the distance between the two spherical caps (of revolution) of area A and perimeter P is at least $P/2$ for all $A \leq A_0$, then a spherical cap minimizes perimeter among (disjoint unions of) embedded discs of prescribed total area $A \leq A_0$.*

Remarks. In an earlier version of this paper, we conjectured that Theorem 3.6 holds for all A. For strictly decreasing curvature, this has been proved by Ritoré [Rit, Cor. 3.5] (who actually proves that the circles of revolution are the only stable, closed, constant-curvature curves). It follows (as in Theorem 3.1) that among multiply-connected regions of prescribed area, a spherical cap, convex annulus, or complement minimizes perimeter. Annuli do sometimes beat spherical caps, as pointed out in [BeC, p. 363]; the surface of Example 3.4 can be truncated and capped to produce an example.

The proof for all A under the additional assumption of nonnegative Gauss curvature in [BeC, Thm. B] is wrong, because equation [BeC, (4.6)] is wrong. For sizable area, it is hard to rule out a disc containing lots of Gauss curvature from both ends of the sphere.

Proof. For c_0 large, let $L(A)$ denote the least perimeter of a curve of $c \leq c_0$ components enclosing c disjoint regions of total area A. Let $G(A)$ denote the Gauss curvature of a spherical cap of area A. Choose a small positive $A_1 \leq A_0$ such that if $0 < A \leq A_1$,

$$4\pi - 2G(A/2) \geq 2\pi - G(A). \tag{1}$$

If $c = 1$, the hypothesis on A_0 implies that the perimeter-minimizing region overlaps at most one spherical cap and hence has no more Gauss curvature. As in the proof of Theorem 2.1, by Gauss-Bonnet, $LL'_{L_*} \geq 2\pi - G(A)$. Likewise if $c \geq 2$,

$$LL'_{L_*} \geq 2\pi c - 2G(A/2) \geq 4\pi - 2G(A/2) \geq 2\pi - G(A)$$

by (1). Integrating yields an inequality on L^2, with equality for a spherical cap, under the hypothesis $A \leq A_1$. Moreover, as long as some minimizer has $c = 1$ and $A \leq A_0$, the spherical cap is minimizing.

Suppose that for some $A_1 \leq A_2 \leq A_0$, a minimizer has $c \geq 2$, and choose the least such A_2. Its perimeter is no less than the sums of the perimeters of spherical caps of the same area, which is greater than the perimeter of a single spherical cap of area A_2, because by Lemma 3.2(A) the geodesic curvature κ of the boundary of spherical caps is decreasing or negative all the way to the geodesic in the middle, so that it is cheaper to enlarge the first cap than to add a second cap. From this contradiction we conclude that a spherical cap is the minimizer for $0 < A \leq A_0$.

At a talk on this paper, E. Calabi asked about perimeter minimizers in the larger class of immersed systems of curves bounding regions with positive multiplicity.

3.7. Isoperimetric Theorem (overlapping regions). *Let S be a plane, real projective plane, or closed disc with smooth, rotationally symmetric metric such that the Gauss curvature K is a nonincreasing function of the distance r from the origin. Given $A > 0$, there is a shortest oriented finite system of immersed curves bounding regions with positive multiplicities, of total area A (counting multiplicities). It is a sum of standard (non-overlapping) minimizers from Theorem* 3.1*, which must overlap pairwise, possibly plus copies of S. We give more detail in two interesting cases:*

A. *If S is a complete plane of nonnegative Gauss curvature, then a (round) disc (about the origin, of multiplicity* 1*) still minimizes perimeter.*

B. *If S is a complete plane of finite total area A_0, let $A = kA_0 + A'$, with $0 \leq A' < A_0$. If $0 < A' \leq A_0/2$, the disc complement plus k copies of S is minimizing. If $A' \geq A_0/2$, the disc plus k copies of S is minimizing.*

As usual, for smoothness, it is sufficient that the metric be C^1 and piecewise C^2.

Proof. Any region with multiplicity can be decomposed (e.g. as in [M1, Lemma p. 98] or [Fed, 4.5.17]) into a sum of regions of multiplicity 1 of the same perimeter. By Theorem 3.1, each summand would be at best a disc or annulus about the origin, or complement (or possibly a copy of S), bounded by circles. Most of them must be tiny circles at the left and/or (if $P(r) \to 0$) at the right. Since by Lemma 3.2(A) in any case, $\kappa(r)$ is nonincreasing, all but a bounded number of circles on the left or right may be profitably replaced by one circle. In this category a minimizer exists.

If a pair of summands does not overlap, combine them into one, until all summands overlap.

Case A: *complete plane of nonnegative curvature.* By Theorem 3.1(A), a minimizer is a disc with multiplicity m. Suppose $m > 1$. Since the geodesic curvature is nonincreasing by 3.2(1), moving one circle right and another left would decrease perimeter.

Case B: *complete plane of finite area.* By Theorem 3.1(B), a minimizer is a disc or disc complement with multiplicity m (not both since they do not overlap), possibly plus copies of S. Since the geodesic curvature is nonincreasing by Lemma 3.2(A), moving one circle right and another left would decrease perimeter (unless κ remains constant, in which case continuation leads to the asserted minimizer without increasing perimeter, although the minimizer is not unique).

Remarks. Finding a minimizer among (limits of) immersed discs is harder, unless the minimizer of 3.7 happens to be a disc or a limit of immersed discs, such as a disc with multiplicity.

Incidentally, existence and regularity is relatively trivial for the overlapping regions problem, since curves do not bump up against each other.

3.8. Higher multiplicity discs occur. We now give an example of a complete, planar surface S in which there are perimeter-minimizing discs of arbitrarily high multiplicity. Consider the flared end with circumference

$$P(r) = e^r e^{e^r}, \quad P'(r) = (1 + e^r)P(r), \quad P''(r) = (1 + 3e^r + e^{2r})P(r),$$

($0 \le r < \infty$), circles of geodesic curvature $\kappa = P'/P = 1 + e^r$, and (by differentiating 3.1(1)) Gauss curvature $K = -P''/P = -(1 + 3e^r + e^{2r})$, which decreases rapidly. Since $P'(0) = 2e < 2\pi = (2\pi r)'$, the surface can be closed off continuously differentiably with a spherical cap ($r_0 \le r \le 0$).

Since $\kappa(r)$ is decreasing on the spherical cap and increasing on the flared end, by Theorems 3.1(A) and 3.6, a minimizer is bounded by multiples of a circle or two circles $\{r = r_1\}, \{r = r_2\}$ with the same geodesic curvature, one on the cap and one on the end. Suppose there are two. Note that by 3.2(1)

$$\frac{d\kappa}{dA} = \frac{d\kappa/dr}{dA/dr} = \frac{-K - \kappa^2}{P}.$$

On the spherical cap,

$$\frac{d^2P}{dA^2}(r) = \frac{d\kappa}{dA}(r) < -\frac{\kappa^2}{P}(r) < -\frac{\kappa^2}{P}(0) = -\frac{4}{e}.$$

On the flared end,

$$\frac{d^2P}{dA^2}(r) = \frac{d\kappa}{dA}(r) = \frac{(1 + 3e^r + e^{2r}) - (1 + e^r)^2}{P(r)} = \frac{1}{e^{e^r}} < \frac{1}{e}.$$

Therefore the net second order effect is negative, and moving one circle right and the other left would reduce perimeter, the desired contradiction.

Now it suffices to show that a large circle is longer than a smaller circle of multiplicity two enclosing the same area (counting multiplicity). The area inside a large circle is given by $A(r) = e^{e^r} + c$. Suppose that the disc of multiplicity two out to r_1 has the same area as the disc out to r_2:

$$2(e^{e^{r_1}} + c) = e^{e^{r_2}} + c.$$

Then $e^{e^{r_1}} = (e^{e^{r_2}} - c)/2$ and for r_1, r_2 large,

$$\begin{aligned} e^{r_1} &= \log(e^{e^{r_2}}/2) + \log(1 - ce^{-e^{r_2}}) \\ &\le e^{r_2} - \log 2 + 2|c|e^{-e^{r_2}} < e^{r_2} - .5. \end{aligned}$$

Therefore the perimeters satisfy

$$\begin{aligned} P_1 &= 2e^{r_1}e^{e^{r_1}} \\ &< 2(e^{r_2} - .5)(e^{e^{r_2}} + |c|)/2 \\ &< 2e^{r_2}e^{e^{r_2}} = P_2 \end{aligned}$$

for r_1, r_2 large.

3.9. **Negative multiplicities.** Theorem 3.7 and proof have analogs allowing negative as well as positive multiplicities. A minimizer is a finite integral combination of minimizers from Theorem 3.1, which overlap pairwise if they have multiplicities of the same sign. Of course if S has finite area and no boundary, the more general minimizers are just the minimizers of Theorem 3.6 minus multiples of S. In particular, if mD is a multiplicity-m, perimeter-minimizing disc of Section 3.7, inside a very large open ball S, it follows easily that $(S - D) - (m - 1)D$ is a perimeter-minimizing region, consisting of two subregions of multiplicity $+1$ and $-(m + 1)$. On the other hand, it follows from Theorem 3.6(B) that a minimizer of prescribed area $A > 0$ in a complete plane of finite area and decreasing Gauss curvature has no regions of negative multiplicity.

3.10. **Constant curvature curves.** In an earlier draft of this paper, we conjectured that in a plane of revolution of strictly decreasing Gauss curvature, the circles of revolution are the only simple closed curves of constant geodesic curvature. Ritoré [Rit] has proved this result and thus provided another approach to some of our results. (It followed from Schmidt [Sch, Satz XXXIII p. 601, footnote p. 602, remark bottom p. 603] that a constant-curvature curve must enclose the origin. This is trivial for nonpositive constant geodesic curvature, since rotating any presumed counterexample until it first touched itself would contradict the maximum principle.)

In any smooth, compact, Riemannian surface, given small prescribed area, a perimeter minimizer is a small embedded disc near a point p of maximum Gauss curvature, as follows from Lemma 2.2 and standard estimates (see [Pan, sect. 3]). If p is a nondegenerate maximum, then, by the implicit function theorem and hard estimates, some small deleted neighborhood has a unique foliation by constant-curvature curves ([Ye1], [Ye2]). They are probably all perimeter minimizing for given area if p is a unique maximum. Perhaps any smooth Riemannian disc with boundary of constant geodesic curvature contains a critical point of the Gauss curvature.

A surface of revolution of strictly decreasing Gauss curvature always contains nonsimple closed curves of constant geodesic curvature. A curve of large geodesic curvature κ follows a loop-de-loop path around a circle of revolution; for some values of κ it closes up.

4. Comparison theorems

Benjamini and Cao [BeC] state their results as comparison theorems for more general surfaces. Similarly, our results yield our more general Comparison Theorem 4.1, results on eigenvalues (4.2), and a new simple proof of the minimal volume of $\mathbf{R}^2$ (4.3).

4.1. Comparison Theorem. *Let S_0 be a plane, real projective plane, or closed disc with smooth, rotationally symmetric metric of nonincreasing Gauss curvature, or the round sphere. Let S_1 be any smooth Riemannian surface, perhaps with boundary, such that*

(1) *a region of area A' has no more Gauss curvature than the round disc about the origin of area A' in S_0,*

for $0 < A' \leq A <$ area S_0. Then any disc (or union of discs) of area A in S_1 has at least as much perimeter as the round disc about the origin in S_0.

Suppose further that S_1 is a plane, real projective plane, closed disc, or sphere with area $S_1 =$ area S_0 and that hypothesis (1) holds for all $0 < A' \leq A <$ area S_1. Then any region of area A in S_1 has at least as much perimeter as the perimeter-minimizing region of the same area in S_0.

As remarked after Theorem 2.1, for smoothness, it is sufficient that the metric be C^1 and piecewise C^2.

Remarks. The round sphere is the only rotationally symmetric sphere of nonincreasing Gauss curvature (see remarks after Theorem 2.1).

For the first conclusion of Theorem 4.1, flow through interior equidistants may be used to show that equality holds only if the disc in S_1 is isometric to the round disc in S_0. Indeed, the curves must all be minimizing and provide the polar coordinate system for the rotational symmetry.

For general regions, it is essential to restrict the topological type of S_1, or a long, skinny torus for example would give a counterexample. If S_1 is a sphere, to satisfy (1), S_0 and S_1 must both be congruent round spheres. Indeed, S_0 cannot be a complete plane or $\mathbf{R}P^2$ (too little total curvature). If S_0 is a plane which is not complete, by the theorem the perimeter of a circle in S_0 must go to 0 as the area increases, so S_0 must be a sphere (possibly with a removable singularity at the end), hence round, say of area 4π. Since equality holds at $A = 4\pi$, the spheres must be congruent.

Proof. The first conclusion follows immediately from the Isoperimetric Inequality 2.4.

The second conclusion follows easily from Theorems 3.1 and 3.3. As usual take any collection of disjoint simple closed curves in S_1 enclosing area A. Since S_1 is a plane, $\mathbf{R}P^2$, or sphere, each curve bounds a disc (except possibly for one superfluous curve in an $\mathbf{R}P^2$). As usual the discs are nested. By the first conclusion of this theorem, discs at the same level, of total area A', have no less perimeter than a single disc about the origin in S_0 of area A'. (This is where we need hypothesis (1) for all A', which may exceed A.) Therefore that total perimeter is no less than that of a region of the same area in S_0 bounded by circles. If that region includes the complement of the outermost circle, it still has the right area, because area $S_1 =$ area S_0.

4.2. **Eigenvalues.** Benjamini and Cao [BeC, §3] follow Faber and Krahn (see [Bér, Chap. IV, esp. §31] and [Ch1, Chapt. IV]) in deducing estimates on the first eigenvalue of the Laplacian from isoperimetric estimates. The same arguments show that if S_0 and S_1 of Theorem 4.1 are closed discs, then the first eigenvalue of S_1 is greater than or equal to the first eigenvalue of S_0.

4.3. **Minimal volume of $\mathbf{R}^2$.** Our comparison theorems provide an alternative proof of Gromov's suggestion that *the infimum A_0 of areas of smooth metrics on $\mathbf{R}^2$ with Gauss curvature $|K| \leq 1$ is attained by a unit sphere with a hyperbolic cusp*, as proved by Bavard and Pansu [BaP] and later by Bowditch [Bow]. Indeed, let S_0 denote this piecewise smooth surface of revolution of nonincreasing Gauss curvature. Assume that there is a smooth competitor S_1 of area $A_1 < A_0$ and Gauss curvature $|K| \leq 1$. The inequality of Cohn-Vossen (see e.g. [Wh, Cor. 2]) implies that $\iint_{S_1} K \leq 2\pi = \iint_{S_0} K$. Consequently the total curvature of a region of S_1 is no more than the total curvature of a disc about the origin of S_0. Therefore by Theorem 4.1, the least perimeter L_i of discs in S_i of area A satisfies $L_0(A) \leq L_1(A)$. Since the injectivity radius of S_1 goes to zero at infinity, $\underline{\lim}_{A\to A_1} L_1(A) = 0$. Consequently $\underline{\lim}_{A\to A_1} L_0(A) = 0$, a contradiction of $A_1 < A_0$.

References

[And] Ben Andrews, *Contraction of convex hypersurfaces by their affine normal*, J. Diff. Geom. **43** (1996), 207–230. MR **97m**:58045

[BaP] Christophe Bavard and Pierre Pansu, *Sur le volume minimal de* $\mathbb{R}^2$, Ann. Sci. École Norm. Sup. (4) **19** (1986), 479–490. MR **88b**:53048

[BeC] Itai Benjamini and Jinguo Cao, *A new isoperimetric theorem for surfaces of variable curvature*, Duke Math. J. **85** (1996), 359–396. MR **97m**:58046

[Bér] Pierre Bérard, *Spectral Geometry: Direct and Inverse Problems*, Lecture Notes in Math., No. 1207, Springer-Verlag, New York, 1986. MR **88f:**58146

[Bow] B. H. Bowditch, *The minimal volume of the plane*, J. Austral. Math. Soc. **55** (1993), 23–40. MR **94i:**53034

[Br] Hubert L. Bray, *The Penrose inequality in general relativity and volume comparison theorems involving scalar curvature*, Ph.D. Dissertation, Stanford Univ., August, 1997.

[Ch1] Isaac Chavel, *Eigenvalues in Riemannian Geometry*, Academic Press, 1984. MR **86g:**58140

[Ch2] Isaac Chavel, *Riemannian Geometry—A Modern Introduction*, Cambridge Univ. Press, 1993. MR **95j:**53001

[Fed] Herbert Federer, *Geometric Measure Theory*, Springer-Verlag, New York, 1969. MR **41:**1976

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

[HH] Erhard Heinz and Stefan Hildebrandt, *Some remarks on minimal surfaces in Riemannian manifolds*, Comm. Pure Appl. Math. **23** (1970), 371–377. MR **41:**4398

[HHM] Hugh Howards, Michael Hutchings, and Frank Morgan, *The isoperimetric problem on surfaces*, Amer. Math. Monthly **106** (1999), 430–439. CMP 99:15

[K] Bruce Kleiner, *An isoperimetric comparison theorem*, Invent. Math. **108** (1992), 37–47. MR **92m:**53056

[M1] Frank Morgan, *Geometric Measure Theory: a Beginner's Guide*, 2nd ed., Academic Press, 1995, 3rd ed., 2000. MR **96c:**49001

[M2] Frank Morgan, *Riemannian Geometry: a Beginner's Guide*, 2nd ed., A.K. Peters, Ltd., 1998. MR **98i:**53001

[MJ] Frank Morgan and David L. Johnson, *Some sharp isoperimetric theorems for Riemannian manifolds*, preprint (1999).

[Mo] Charles B. Morrey, Jr., *Multiple Integrals in the Calculus of Variations*, Springer-Verlag, New York, 1966. MR **34:**2380

[Os] Robert Osserman, *The isoperimetric inequality*, Bull. Amer. Math. Soc. **84** (1978), 1182–1238. MR **58:**18161

[Pan] Pierre Pansu, *Sur la régularité du profil isopérimétrique des surfaces riemanniennes compactes*, Ann. Inst. Fourier (Grenoble) **48** (1998), 247–264. MR **99i:**53035

[Rit] Manuel Ritoré, *Constant geodesic curvature curves and isoperimetric domains in rotationally symmetric surfaces.*

[Rud] Walter Rudin, *Real and Complex Analysis*, McGraw-Hill, New York, 1966. MR **35:**1420

[Sch] Erhard Schmidt, *Über eine neue Methode zur Behandlung einer Klasse isoperimetrischer Aufgaben im Grossen*, Math. Z. **47** (1942), 489–642. MR7 (1946), p. 527. MR **7:**527h

[Sp] Michael Spivak, *A Comprehensive Introduction to Differential Geometry*, Vol. 4, Publish or Perish, Wilmington, 1979. MR **82g:**53003d

[T1] Peter Topping, *The isoperimetric inequality on a surface*, Manuscripta Math. **100** (1999), 23–33. CMP 2000:02

[T2] Peter Topping, *Mean curvature flow and geometric inequalities*, Journal für die Reine und Angewandte Math. **503** (1998), 47–61. MR **99m:**53080

[Wh] Brian White, *Complete surfaces of finite total curvature*, J. Diff. Geom. **26** (1987), 315–326. MR **88m:**53020; correction MR **89j:**53009

[Ye1] Rugang Ye, *Constant mean curvature foliation: singularity structure and curvature estimate*, Pacific J. Math. **174** (1996), 569–587. MR **97m:**53054

[Ye2] Rugang Ye, *Foliation by constant mean curvature spheres*, Pacific J. Math. **147** (1991), 381–396. MR **92f:**53030

Department of Mathematics, Williams College, Williamstown, Massachusetts 01267
E-mail address: `Frank.Morgan@williams.edu`

Department of Mathematics, Stanford University, Stanford, California 94305
E-mail address: `hutching@math.stanford.edu`

Department of Mathematics, Wake Forest University, Winston-Salem, North Carolina 27109
E-mail address: `howards@wfu.edu`