FILLING AREA CONJECTURE AND OVALLESS REAL HYPERELLIPTIC SURFACES

VICTOR BANGERT[!], CHRISTOPHER CROKE[+], SERGEI V. IVANOV[†], AND MIKHAIL G. KATZ[*]

Abstract. We prove the filling area conjecture in the hyperelliptic case. In particular, we establish the conjecture for all genus 1 fillings of the circle, extending P. Pu's result in genus 0. We translate the problem into a question about closed ovalless real surfaces. The conjecture then results from a combination of two ingredients. On the one hand, we exploit integral geometric comparison with orbifold metrics of constant positive curvature on real surfaces of even positive genus. On the other, we exploit an analysis of the combinatorics on unions of closed curves, arising as geodesics of such orbifold metrics.

Contents

Key words and phrases. filling area, hyperelliptic surfaces, integral geometry, real surfaces, systolic inequality.

[!]Partially Supported by DFG-Forschergruppe 'Nonlinear Partial Differential Equations: Theoretical and Numerical Analysis'.

[+]Supported by NSF grant DMS 02-02536 and the MSRI.

[†]Supported by grants CRDF RM1-2381-ST-02, RFBR 02-01-00090, and NS-1914.2003.1.

[*]Supported by the Israel Science Foundation (grants no. 620/00-10.0 and 84/03).

1. To fill a circle: an introduction

Consider a compact manifold N of dimension $n \geq 1$ with a metric d. Here d is a distance function which is not necessarily Riemannian. The notion of the Filling Volume, $FillVol(N^n, d)$, of N was introduced in [Gr83], where it is shown that when $n \geq 2$,

$$FillVol(N^n, d) = \inf_g vol(X^{n+1}, g) \tag{1.1}$$

where X is any fixed manifold such that $\partial X = N$. Here one can even take a cylinder $X = N \times [0, \infty)$. The infimum is taken over all complete Riemannian metrics g on X for which the boundary distance function is $\geq d$, *i.e.* the length of the shortest path in X between boundary points p and q is $\geq d(p,q)$. In the case $n = 1$, the topology of the filling X^2 can affect the infimum, as is shown by example in [Gr83, Counterexamples 2.2.B].

The precise value of the filling volume is not known for any Riemannian metric. However, the following values for the canonical metrics on the spheres (of sectional curvature $+1$) is conjectured in [Gr83], immediately after Proposition 2.2.A.

Conjecture 1.1. *$FillVol(S^n, can) = \frac{1}{2}\omega_{n+1}$, where ω_{n+1} represents the volume of the unit $(n+1)$-sphere.*

This conjecture is still open in all dimensions. The case $n = 1$ can be broken into separate problems depending on the genus of the filling. The filling area of the circle of length 2π with respect to simply connected fillings (*i.e.* by a disk) is indeed 2π, by Pu's inequality (2.1) applied to the real projective plane obtained by identifying opposite points of the circle. Here one may need to smooth the resulting metric, with an arbitrarily small change in the two invariants involved.

In Corollary 1.6 below, we prove the corresponding result when the filling is by a surface of genus one. For fillings by an orientable surface X_1 of positive genus, we apply the same technique of glueing antipodal points together to form a nonorientable surface X_2 (whose metric might again need to be smoothed as before). We are thus reduced to proving a conjectured relative version of Pu's inequality, see discussion around formula (2.3). This conjecture is most easily stated by passing to the orientable double cover X_3 with deck transformation τ, as follows.

Conjecture 1.2. *Let X be an orientable surface with a Riemannian metric g that admits a fixed point free, orientation reversing, isometric involution τ. Then there is a point $p \in X$ with*

$$\frac{dist_g(p,\tau(p))^2}{area(g)} \le \frac{\pi}{4}.$$

The original proof of Pu's inequality passed via conformal techniques (namely, uniformisation), combined with integral geometry (see [Iv02] for another proof). It seems reasonable also to try to apply conformal techniques to the proof of the conjectured relative version of Pu's inequality. This works well in the hyperelliptic case.

Let X be an orientable closed aspherical surface. (By a surface in this context we mean one that comes with a fixed conformal structure and maps are conformal maps.)

Definition 1.3. A *hyperelliptic involution* is a holomorphic map $J : X \to X$ satisfying $J^2 = 1$, with $2s+2$ fixed points, where s is the genus of X. A surface X admitting such an involution will be called hyperelliptic.

The involution J can be identified with the nontrivial element in the center of the (finite) automorphism group of X (*cf.* [FK92, p. 108]) when it exists, and then such a J is unique, *cf.* [Mi95, p.204]. Note that the quotient map

$$Q : X \to S^2 \tag{1.2}$$

of such an involution J is a conformal branched 2-fold covering.

Definition 1.4. The $2s+2$ fixed points of J are called *Weierstrass points*. Their images in S^2 under the ramified double cover Q of formula (1.2) will be referred to as *ramification points*.

Our result about hyperelliptic surfaces is the following. Given a circle $C_0 \subset S^2$ and a point $w_0 \in S^2 \setminus C_0$, we can consider the orbifold metric with equator C_0 and poles at w_0 and at the image of w_0 under inversion in C_0, *cf.* Section 4. Given a hyperelliptic surface $Q : X \to S^2$ together with a circle $C \subset X$ double covering a circle $C_0 \subset S^2$, and a point $w \in X \setminus C$, denote by

$$AF(C,w) \tag{1.3}$$

the pullback to X of the orbifold metric for $C_0 = Q(C)$ and $w_0 = Q(w)$. Given an involution $\iota : X \to X$ fixing a circle, denote by $X^\iota \subset X$ its fixed circle.

A closed oriented Riemann surface X is called *ovalless real* if it admits a fixed point free, antiholomorphic involution τ, see Appendix A for a more detailed discussion.

Theorem 1.5. *Let (X, τ, J) be an ovalless real hyperelliptic surface of even genus s, and g a Riemannian metric in its conformal class. Then there is a point $p \in X$ with*

$$\frac{dist_g(p, \tau(p))^2}{area(g)} \leq \frac{\pi}{4}. \tag{1.4}$$

Specifically, there exists a curve joining p and $\tau(p)$, of length at most $\left(\frac{\pi}{4} area(g)\right)^{1/2}$, which consists of arcs of at most $s+1$ special curves γ_i. Each of the γ_i is a geodesic of the singular constant (positive) curvature metric

$$AF(X^{\tau \circ J}, w_i),$$

where w_i is one of the Weierstrass points, while $X^{\tau \circ J}$ is the fixed circle of the involution $\tau \circ J$.

The notion of an ovalless real hyperelliptic surface is reviewed in Appendix A. The two-step proof is sketched in section 3. An orbifold metric on the sphere that plays a key role in the proof is described in Section 4. The two steps appear, respectively, in Section 6 and Section 7.

Since every genus 2 surface is hyperelliptic [FK92, Proposition III.7.2, page 100], we obtain the following corollary.

Corollary 1.6. *All orientable genus 1 fillings of the circle satisfy the filling area conjecture.*

2. Pu's long shadow

In the companion paper [BCIK04], we study certain optimal systolic inequalities for a Riemannian manifold (X, g), depending on a pair of parameters, n and b. Here n is the dimension of X, while b is its first Betti number. The definitions of the systolic invariants $stsys_1(g)$, $confsys_1(g)$, and $sys\,\pi_1(g)$ can be found in the survey [CK03], while the Abel-Jacobi map in [Li69], [Gr99, p.249], *cf.* [Mi95, p. 249], [BK04, (4.3)], [IK04]. The proof of the inequalities involves streamlining and generalizing the techniques pioneered in [BI94, BI95] and [Gr99, pp. 259-260] (*cf.* [CK03, inequality (5.14)]), resulting in a construction of Abel-Jacobi maps from X to its Jacobi torus $\mathbb{T}^b$, which are area-decreasing, with respect to suitable norms on the torus. The norms in question are the stable norm, the conformally invariant norm, as well as other L^p norms.

The study of the boundary case of equality in the case $n = b + 2$ depends on the filling area conjecture and/or a suitable generalisation of Pu's inequality, as we now discuss. The inequality of P. Pu [Pu52] (*cf.* [Iv02]) can be generalized as follows: every surface (S, g) which is not a 2-sphere satisfies the inequality

$$sys\,\pi_1(g)^2 \leq \frac{\pi}{2} area(g), \tag{2.1}$$

where the boundary case of equality is attained precisely when, on the one hand, the surface S is a real projective plane, and on the other, the metric g is of constant Gaussian curvature. See [IK04] for a more detailed discussion.

Now let S be a nonorientable closed surface. Let $\phi : \pi_1(S) \to \mathbb{Z}_2$ be an epimorphism, corresponding to a map $\hat{\phi} : S \to \mathbb{R}P^2$ of absolute degree $+1$. We define the "1-systole relative to ϕ", denoted $\phi sys_1(g)$, of a metric g on S, by minimizing length over loops γ which are not in the kernel of ϕ:

$$\phi sys_1(g) = \min_{\phi([\gamma])\neq 0 \in \mathbb{Z}_2} length(\gamma). \tag{2.2}$$

Question 2.1. Does every nonorientable surface (S, g) and map $\hat{\phi}$: $S \to \mathbb{R}P^2$ of absolute degree one, satisfy the following relative version of Pu's inequality: $\phi sys_1(g)^2 \leq \frac{\pi}{2} area(g)$?

This question appeared in [CK03, conjecture 2.7]. Let

$$\sigma_2 = \sup_{(S,g)} \frac{\phi sys_1(g)^2}{area(g)}, \tag{2.3}$$

where the supremum is over all triples (S, g, ϕ) as above. Thus we ask whether $\sigma_2 = \frac{\pi}{2}$. The answer is affirmative in the class of metrics g whose underlying conformal structure (on the associated orientable surface) is hyperelliptic, *cf.* Theorem 1.5. An analogue for higher dimensional manifolds of the relative 1-systolic ratio was studied in [Bab].

It was shown in [IK04] that $\sigma_2 \in [\frac{\pi}{2}, 2]$, as well as a theorem which in the case of the manifold $\mathbb{R}P^2 \times \mathbb{T}^b$ states that every metric g on $\mathbb{R}P^2 \times \mathbb{T}^b$ satisfies the following inequality:

$$stsys_1(g)^b \, sys\,\pi_1(g)^2 \leq \sigma_2 \gamma_b^{\frac{b}{2}} \, vol_{b+2}(g), \tag{2.4}$$

where σ_2 is the optimal systolic ratio from (2.3).

Answering Question 2.1 in the affirmative would allow one to characterize the boundary case of equality in equation (2.4). Our Theorem 1.5 can be thought of as a partial result in the this direction. Calculating σ_2

depends on calculating the filling area of the Riemannian circle. Removing the hyperellipticity assumption of Theorem 1.5 would amount to proving the filling area conjecture [Gr83], as explained in section 1.

3. Outline of proof of Theorem 1.5

We will prove Theorem 1.5 in a way modeled on Pu's proof. The idea is to obtain estimates for an arbitrary metric in the conformal class of a hyperelliptic surface, by applying integral geometry to the quotient metric (by the hyperelliptic involution) on the 2-sphere. We start with a metric in the conformal class of an ovalless real hyperelliptic surface (X, τ, J) of even genus s and area A, and then proceed as follows.

Step 1: In Lemma 6.2, we will show that we may assume that the commuting involutions J and τ are isometries of our metric. Hence there is an induced metric $g_1 = f^2(x)g_0$ of area $A/2$ on S^2 (here $f(x)$ is a nonnegative square integrable function) which pulls back (away from the Weierstrass points) to our original metric on X. The involution τ induces an orientation reversing isometry, denoted τ_0, on S^2 which can be assumed to be the map that leaves the equator fixed and exchanges the northern and southern hemispheres. Clearly, τ_0 maps ramification points to ramification points, and hence there are $s+1$ of them in each hemisphere.

Step 2: There is a loop γ on S^2 of g_1-length $\leq \sqrt{\frac{\pi}{4}A}$ through a point p on the equator (hence p is fixed by τ_0) such that a lift of γ starting at a preimage, $\tilde{p}$, of p back on X is not closed.

We call a loop whose g_1-length $\leq \sqrt{\frac{\pi}{4}A}$ an *optimally short loop.* The loop we find lies in a hemisphere. In fact, we find a loop with a lift that is not closed, *i.e.* the sum of the winding numbers about the ramification points in the hemisphere is odd.

Note that this proves the theorem since the endpoints of such a lift would have to be $\tilde{p}$ and $\tau(\tilde{p})$ since the fact that p is fixed by τ_0 means that τ (which has no fixed points) must exchange the two (distinct) preimages of p.

4. Near optimal surfaces and the football

It is often useful in the proof of sharp inequalities to keep in mind how the argument applies to spaces that are optimal or near optimal. We present such a discusson below, and also find some optimally short loops on these optimal surfaces.

A special role here is played by the simply connected 2-dimensional orbifold $AF(\pi, \pi)$, which we will refer to henceforth as an (American) football, see Figure 5.1. It is the orbifold of constant curvature with a pair of conical singularities, each with total angle π. We denote the metric g_{AF}.

Denote by g_0 the round metric of Gaussian curvature $+1$ on the 2-sphere. Here $S^2 = \mathbb{C} \cup \{\infty\}$, with the usual coordinates $(x, y) = (\Re(z), \Im(z))$, $z \in \mathbb{C}$. We have

$$g_0 = \frac{4}{(1+x^2+y^2)^2}\left(dx^2+dy^2\right), \tag{4.1}$$

cf. [Ri1854, p. 292].

Proposition 4.1. *The football admits the following three equivalent descriptions:*

1. *The orbifold $AF(\pi,\pi)$ is obtained from a round hemisphere by folding its boundary in two, i.e. using an identification on the boundary which is a reflection with 2 antipodal fixed points.*
2. *The rotation by π around the z-axis in $\mathbb{R}^3$ gives rise to a ramified double cover $D : (S^2, \bar{g}_0) \to AF(\pi,\pi)$, which is a local isometry away from the poles.*
3. *The orbifold metric g_{AF} can be expressed in terms of the standard metric by means of the following conformal factor:*

$$g_{AF} = \frac{1+\mathcal{O}(r)}{4r} g_0,$$

where $\mathcal{O}(r) \to 0$ as $r \to 0$, while $r = |z| = \sqrt{x^2+y^2}$.

Proof. The equivalence of the first two definitions is clear. Let us relate the first definition to the third. We represent a hemisphere defining the orbifold as the upperhalf plane $H^2 \subset \mathbb{C}$. Then the orbifold can be viewed as a quotient of (H^2, g_0), where the positive and negative rays of the x-axis are identified. Denote by w the complex coordinate in H^2.

By the uniqueness of the underlying conformal structure, there exists a conformal map $c : (H^2, w) \to (S^2, z)$, identifying the orbifold with the 2-sphere, with complex coordinate z. In our coordinates, we can represent c by the map $z = w^2$. Here the singular points of the orbifold correspond to 0 and ∞. Since $\frac{dc}{dw} = 2w$, the metric g_0 pulls back under c to the quadratic differential $c^*(g_0)(w) = \mu(|w|^2)g_0$, where

$$\mu(t) = 4t\frac{(1+t)^2}{(1+t^2)^2} = 4t(1+\mathcal{O}(t)).$$

Note that g_0 restricted to H^2 is precisely the orbifold metric $c^*(g_{AF})$ we are looking for. Thus

$$\begin{aligned} c^*(g_{AF}) &= \frac{1}{\mu(|w|^2)} c^*(g_0) \\ &= \frac{1}{\mu(|z|)} c^*(g_0) \\ &= c^* \left(\frac{1+\mathcal{O}(|z|)}{4|z|}(g_0) \right), \end{aligned}$$

as required. $\square$

Remark 4.2. The function $\frac{1}{\sqrt{r}}$ is locally square integrable in $\mathbb{C}$, which is important for our applications, *cf.* Proposition B.1 of Appendix B.

Example 4.3. The football $AF(\pi,\pi)$ arises in the typical example where the associated systolic ratio is close to being optimal. Thus, start with a round $\mathbb{R}P^2$. Attach a small handle. The orientable double cover X can be thought of as the unit sphere in $\mathbb{R}^3$, with two little handles attached at north and south poles, *i.e.* at the two points where the sphere meets the z-axis. Then one can think of the hyperelliptic involution J as the rotation of X by π around the z-axis. The six Weierstrass points are the six points of intersection of X with the z-axis. The orientation reversing involution τ on X is the restriction to X of the antipodal map in $\mathbb{R}^3$. The composition $\tau \circ J$ is the reflection fixing the xy-plane, in view of the following matrix identity:

$$\begin{pmatrix} -1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & -1 \end{pmatrix} \begin{pmatrix} -1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{pmatrix}. \tag{4.2}$$

Meanwhile, the induced orientation reversing involution τ_0 on S^2 can just as well be thought of as the reflection in the xy-plane. This is because, at the level of the 2-sphere, it is "the same thing as" the composition $\tau \circ J$. Thus the fixed circle of τ_0 is precisely the equator, *cf.* formula (A.3). Then one gets a quotient metric on S^2 which is roughly that of the western hemisphere, with the boundary longitude folded in two, *cf.* Proposition 4.1(2). The metric has little bulges along the z-axis at north and south poles, which are leftovers of the small handle.

5. Finding a short figure eight geodesic

In our case, the optimal metric is not round, so it is natural to consider geodesics on the football $AF(\pi,\pi)$ instead of the round sphere. The branched double cover $D : S^2 \to AF(\pi,\pi)$ of Proposition 4.1(2)

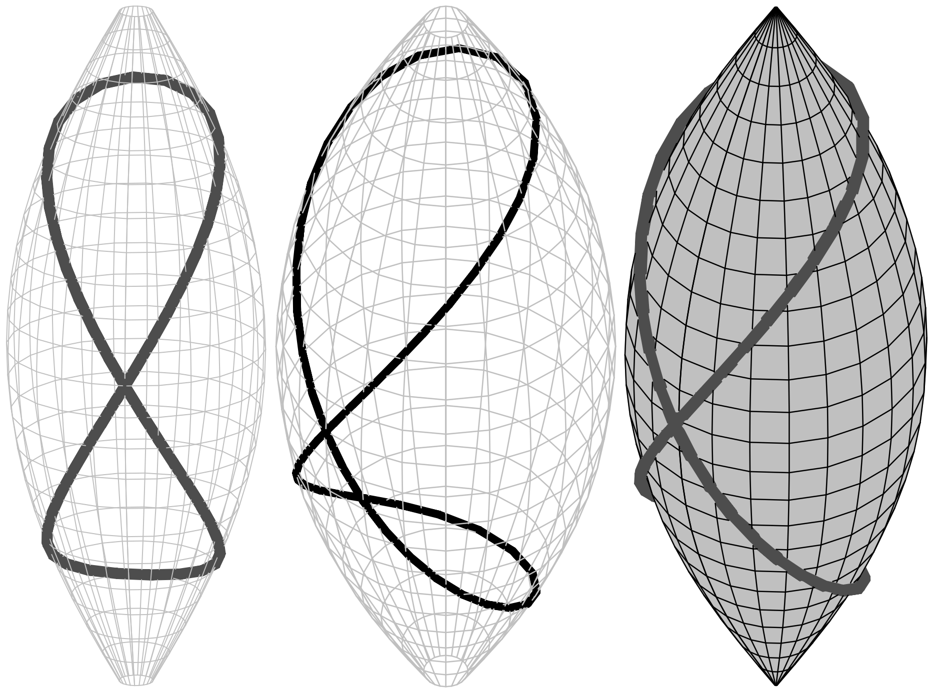

FIGURE 5.1. Figure eight geodesics on an American football

pulls back g_{AF} to the standard metric g_0 on S^2. This allows us easily to see the shape of the g_{AF}-geodesics. They are of figure eight type with their point of self intersection on the equator (except the ones through the poles), see Figure 5.1. Note also that the loop in each of the northern and southern hemispheres winds around the pole precisely once.

Starting with a square integrable function $f(\sigma)$ on $AF(\pi, \pi)$, we can pull it back to S^2 and apply Lemma B.1 to obtain the following.

Lemma 5.1. *There is a g_{AF}-geodesic γ satisfying the following estimate:*

$$\left(\int_\gamma f(\gamma(t)) dt \right)^2 \leq \pi \int_{S^2} f^2(\sigma) d\sigma = 2\pi \, area(f^2 g_{AF}).$$

The factor of 2 comes from the fact that the area of the pull back of the (possibly singular) metric $f^2 g_{AF}$ in S^2 (under the double cover $S^2 \to AF(\pi, \pi)$) has twice the area of the original metric. Lemma 5.1 implies the following. Assume that our function f is symmetric with respect to the inversion in the equator. Then with respect to the metric $f^2 g_{AF}$, the length L of each of the two hoops of the figure eight curve, satisfies the desired inequality $L \leq \sqrt{\frac{\pi}{2} A(f^2 g_{AF})}$. This inequality is sharp, since $A(f^2 g_{AF})$ is half the area A of the original hyperelliptic surface.

6. Proof of circle filling: Step 1

Let X be an ovalless real hyperelliptic surface. Let $J : X \to X$ the hyperelliptic involution, *cf.* Definition 1.3. Let τ be the antiholomorphic involution defining the real structure, *cf.* Appendix A.

Lemma 6.1. *The involution τ commutes with J and descends to S^2. The induced involution $\tau_0 : S^2 \to S^2$ is an inversion in a circle $C_0 = Q(X^{\tau \circ J})$. The set of ramification points is invariant under the action of τ_0 on S^2.*

Proof. By the uniqueness of J (*cf.* Appendix A), we have the commutation relation

$$\tau \circ J = J \circ \tau, \tag{6.1}$$

cf. relation (4.2). Therefore τ descends to an involution τ_0 on the sphere. Suppose τ_0 were fixed point free. Then it would be conjugate to the antipodal map of S^2. Therefore, the set of the $2s+2$ ramification points on S^2 is centrally symmetric. Since there is an odd number, $s+1$, of ramification points in a hemisphere, a generic great circle $A \subset S^2$ has the property that its inverse image $Q^{-1}(A) \subset X$ is connected. Thus both involutions τ and J, as well as $\tau \circ J$, act fixed point freely on the loop $Q^{-1}(A) \subset X$, which is impossible. Therefore τ_0 must fix a circle. □

The next lemma allows us to assume that J and τ are isometries of the metric.

Lemma 6.2. *Let F denote a conformal involution of X, and let $\bar{g} = \frac{1}{2}\{g + F^* g\}$ be the averaged metric. Then for all $p, q \in X$ we have*

$$\frac{dist_{\bar{g}}(p,q)^2}{area(\bar{g})} \geq \frac{\left(\frac{1}{2}\{dist_g(p,q) + dist_g(F(p), F(q))\}\right)^2}{area(g)}. \tag{6.2}$$

and hence if Theorem 1.5 holds for $\bar{g}$ then it holds for g, as well.

Remark 6.3. Applying Lemma 6.2 to $F = \tau$ and $q = \tau(p)$, we obtain a τ-invariant metric g^-, conformal to g, such that

$$\frac{dist_{g^-}(p,q)^2}{area(g^-)} \geq \frac{dist_g(p,q)^2}{area(g)}.$$

If we apply Lemma 6.2 to g^- and $F = J$, $q = \tau(p)$, we obtain a J-invariant metric $\bar{g}$, also conformal to g. Now inequality (6.2) implies that if Theorem 1.5 holds for $\bar{g}$, then it holds for g as well. Moreover, since J and τ commute, $\bar{g}$ is also τ-invariant.

Proof. Since F is a conformal diffeomorphism, we have $F^*g = f^2g$ for some smooth positive f. Then $\bar{g} = \frac{1}{2}(f^2+1)g$. (Note that a similar proof works also with $\bar{g} = \left(\frac{1}{2}(f+1)\right)^2 g$ as an average.) Let dx represent the volume element of g. We have

$$\begin{aligned} area(\bar{g}) &= \int \frac{1}{2}(f^2(x)+1)dx \\ &= \frac{1}{2}\left(\int f^2(x)dx + \int 1dx\right) \\ &= \frac{1}{2}(area(F^*(g)) + area(g)) \\ &= area(g), \end{aligned}$$

and thus the denominator is unchanged. As for the numerator, start with any curve γ from p to q, then $F(\gamma)$ is a curve from $F(p)$ to the point $F(q)$. The length element for $\bar{g}$ is $\sqrt{\frac{f^2+1}{2}} \geq \frac{f+1}{2}$, so that length of the curve γ for the averaged metric is bounded below by the average of the length of the curves γ and $F \circ \gamma$ for the original metric. Thus taking γ to be a $\bar{g}$ minimizing path, we obtain

$$\begin{aligned} dist_{\bar{g}}(p,q) &= L_{\bar{g}}(\gamma) \\ &\geq \frac{1}{2}(L_g(\gamma) + L_g(F \circ \gamma)) \\ &\geq \frac{1}{2}\{dist_g(p,q) + dist_g(F(p),(F(q)))\}, \end{aligned}$$

proving inequality (6.2). □

Now if inequality (1.4) holds for $\bar{g}$ then there is a p such that $\frac{\pi}{4} \geq \frac{dist_{\bar{g}}(p,\tau(p))^2}{area(\bar{g})}$. Hence either $\frac{\pi}{4} \geq \frac{dist_g(p,\tau(p))^2}{area(g)}$, or $\frac{\pi}{4} \geq \frac{dist_g(F(p),\tau(F(p)))^2}{area(\bar{g})}$, and the inequality follows for g, as well. From now on we will assume that J and τ are isometries.

7. Proof of circle filling: Step 2

Let $C_0 = Q(X^{\tau \circ J}) \subset S^2$ be the fixed circle of the involution τ_0. Let $H \subset S^2$ be a connected component of $S^2 \setminus C_0$. Since the set of Weierstrass points is invariant under τ, there is an odd number of ramification points, namely $s+1$, in the "hemisphere" H.

We need to find a optimally short loop (*i.e.* a loop satisfying $L \leq \sqrt{\frac{\pi}{4}A}$) based at a point on the "equator" C_0 whose lift does not close up, and hence connects an orbit of τ.

Lemma 7.1. *For every ramification point $w \in H$, there is an optimally short loop (half of a figure 8 loop) in H with unit winding number around w.*

Proof. By using a Mobius transformation if necessary, we can assume that C_0 is a great circle and our ramification point $w \in S^2$ is the polar point of the great circle C_0.

Recall that S^2 carries a metric induced from X. We then can use a conformal diffeomorphism from $AF(\pi, \pi)$ to S^2 which takes the ramification points to the poles, to pull the metric back to $AF(\pi, \pi)$.

Namely, we compare our surface X to $AF(X^{\tau\circ J}, w)$, *cf.* formula (1.3). Here $AF(X^{\tau\circ J}, w)$ double covers $AF(\pi, \pi)$, while the circle $X^{\tau\circ J}$ double covers C_0 under Q. Recall that the conformal factor of our metric may be assumed invariant under τ. Hence Lemma 5.1 produces an optimally short loop (half of a figure 8 curve) with unit winding number around w, based at a point of C_0.

By placing ourselves in a generic situation, we can assume that the loops we work with do not pass through ramification points, and thus the winding number is well defined. Namely, we assume that the geodesic we pick on $AF(\pi, \pi)$ is generic. This worsens the resulting estimate by an arbitrarily small $\epsilon > 0$. But if we prove that the optimal systolic ratio is within ϵ of $\pi/2$ for arbitrary ϵ, then it certainly follows that it equals $\pi/2$. □

Note in the above we will need to take a different Mobius transformation for each w.

If a loop produced by Lemma 7.1 encircles no other ramification points, then its lift to the ovalless real hyperelliptic Riemann surface (X, τ, J) connects an orbit of τ, and we are done with Step 2. Otherwise we argue as follows.

Fix a finite set S of points in the hemisphere. In the end, S will be the set of ramification points, but the argument involves subsets, too. The loops involved in the argument are not assumed simple. Given a loop b, denote by $W_S(b)$ the set of points of S for which the winding number of b is odd.

Lemma 7.2. *Let b and c be smooth optimally short loops in a disk, such that the set $|W_S(c) \setminus W_S(b)|$ is odd. Then there is an optimally short loop h (with the same basepoint as either b or c) such that $|W_S(h)|$ is odd.*

Proof. Consider a double ramified cover of the disk, with ramification points at S. Given a loop, the number of points with odd winding number is even if and only if the lift of the loop to the cover closes up.

Choose a fixed lift of the loop b, and a fixed lift of the loop c. Suppose both lifts are closed. The points of the intersection $b \cap c$ fall into two types: "crossroads", where their lifts intersect, as well; and "bridges", where the lifts do not intersect, *i.e.* lie in different sheets of the cover.

It suffices to find two intersection points of different types. Indeed, if p and q are such points, consider the arc of b between p and q not containing the basepoint, and the similar arc of c. Exchanging these arcs produces a new pair of loops of the same combined length, whose lifts do not close up. Hence one of them is optimally short, as required.

We need the fact that a self-intersection of one of the loops remains a self-intersection after lifting. Suppose it does not. Then we have a sub-loop, not containing the basepoint, whose lift is not closed. Remove this sub-loop from the original loop. Then the remaining loop has the desired property: it is optimally short, and its lift is not closed.

To find a pair of intersection points of different types, we argue by contradiction. Assume that they are all crossroads (if bridges, change the lift of b). Thus, we have a (homeomorphic) lift of the union $b \cup c$. Hence every simple loop contained in this union lifts to a closed curve, and therefore encircles an even number of ramification points. Consider the graph Γ in the disk, defined by the union of the two loops. Let F_i be the faces (*i.e.* connected components of the complement) of Γ containing points from $W_S(c) \backslash W_S(b)$. Consider the subgraph $\cup_i F_i$ of Γ, and note that

$$W_S(c) \setminus W_S(b) \subset \cup_i F_i. \tag{7.1}$$

Since $|W_S(c) \backslash W_S(b)|$ is odd, formula (7.1) implies that one of the faces, say F_{i_0}, must contain an odd number of points w_j from $W_S(c) \setminus W_S(b)$. By construction, for each such $w_j \in F_{i_0}$, the loop c has odd winding number, while the loop b has even winding number. Any ramification point $p \in F_{i_0}$ can be connected to a w_j by a path disjoint from $b \cup c$. Hence the loops b and c have the same winding number with respect to p as with respect to w_j. Hence in fact we have $p \in W_S(c) \setminus W_S(b) \subset S$.

We conclude that F_{i_0} contains an odd number of ramification points. Hence the lift of its boundary does not close up. This contradicts the assumption that all intersection points are crossroads, and proves the lemma. □

Lemma 7.3. *Let S be an odd pointset in the interior of a Riemannian disk. Consider a family of smooth simple loops $\gamma_1, \dots, \gamma_k$ based at distinct points of the boundary. Let $L > 0$, and assume each loop is of length at most L, and that each point of S is in the interior of at least one of the loops. Then the union of the loops γ_i contains a subloop γ,* *also based at a point of the boundary and of length at most L, such that $W_S(\gamma)$ is odd.*

Proof. Let $|S| = 2n+1$. The proof is by induction on n. Let $b = \gamma_1$. If $W_S(b)$ is odd, the desired short loop is $\gamma = b$. Assume $W_S(b)$ is even. By the inductive hypothesis applied to the odd set

$$S_0 = S \setminus W_S(b) \tag{7.2}$$

in place of S, there is a short loop c such that $W_{S_0}(c)$ is odd. *A priori* the loop c might have odd winding number with respect to additional points in $W_S(b)$, but at any rate the equality

$$W_S(c) \setminus W_S(b) = W_{S_0}(c)$$

allows us to apply Lemma 7.2 to the pair of loops b and c, completing the proof of Lemma 7.3 and Step 2. □

Appendix A. Ovalless reality and hyperellipticity

A.1. **Hyperelliptic surfaces.** For a treatment of hyperelliptic surfaces, see [Mi95, p. 60-61]. By [Mi95, Proposition 4.11, p. 92], the affine part of a hyperelliptic surface X is given by a suitable equation of the form

$$w^2 = f(z) \tag{A.1}$$

in $\mathbb{C}^2$, where f is a polynomial. On this affine part the map J is given by $J(z,w) = (z,-w)$, while the hyperelliptic quotient map $Q : X \to S^2$ is represented by the projection onto the z-coordinate in $\mathbb{C}^2$.

A slight technical problem here is that the map $X \to \mathbb{C}P^2$, whose image is the compactification of the solution set of (A.1), is not an imbedding. Indeed, there is only one point at infinity, given in homogeneous coordinates by $[0 : w : 0]$. This point is a singularity. A way of desingularizing it using an explicit change of coordinates near infinity is presented in [Mi95, p. 60-61].

To explain what happens "geometrically", note that there are two points on our affine surface "above infinity". This means that for a large circle $|z| = r$, there are two circles above it satisfying equation (A.1) where f has even degree $2s+2$ (for a Weierstrass point we would only have one circle). To see this, consider $z = re^{ia}$. As the argument a varies from 0 to 2π, the argument of $f(z)$ will change by $(2s+2)2\pi$. Thus, if $(re^{ia}, w(a))$ represents a continuous curve on our surface, then the argument of w changes by $(2s+2)\pi$, and hence we end up where we started, and not at $-w$ (as would be the case were the polynomial of odd degree). Thus there are *two* circles on the surface over the circle $|z| = r$. We conclude that to obtain a smooth compact surface, we will need to add two points at infinity, *cf.* discussion around [FK92, formula (7.4.1), p. 102].

Thus, the affine part of X, defined by equation (A.1), is a Riemann surface with a pair of punctures p_1 and p_2. A neighborhood of each p_i is conformally equivalent to a punctured disk. By replacing the punctured disk by the full one, we obtain the desired compact Riemann surface X. The point at infinity $[0 : w : 0] \in \mathbb{C}P^2$ is the image of both p_i.

A.2. **Ovalless surfaces.** In our case, the hyperelliptic Riemann surface X admits an antiholomorphic involution τ. In the literature, the components of the fixed point set of τ are sometimes referred to as "ovals". Since in our situation τ is fixed point free, we introduce the following terminology.

Definition A.1. A hyperelliptic surface (X, J) of even genus $s > 0$ is called *ovalless real* if one of the following equivalent conditions is satisfied:

1. X admits an imaginary reflection, *i.e.* a fixed point free, orientation reversing involution τ;
2. the affine part of X is the locus in $\mathbb{C}^2$ of the equation
$$w^2 = -P(z), \tag{A.2}$$
where P is a monic polynomial, of degree $2s + 2$, with real coefficients, no real roots, and with distinct roots.

Lemma A.2. *The two ovalless reality conditions of Definition A.1 are equivalent.*

Proof. A related result appears in [GrH81, p. 170, Proposition 6.1(2)]. To prove the implication $2 \Longrightarrow 1$, note that complex conjugation leaves the equation invariant, and therefore it also leaves invariant the locus of (A.2). A fixed point must be real, but P is positive hence (A.2) has no real solutions. There is no real solution at infinity, either, as there are two points at infinity, which are not Weierstrass points since P is of even degree, as discussed above. The desired imaginary reflection τ switches the two points at infinity, and, on the affine part of the Riemann surface, coincides with complex conjugation $(z, w) \mapsto (\bar{z}, \bar{w})$ in $\mathbb{C}^2$.

To prove the implication $1 \Longrightarrow 2$, note that by Lemma 6.1, the induced involution τ_0 on $S^2 = X/J$ may be thought of as complex conjugation, by choosing the fixed circle of τ_0 to be the circle
$$\mathbb{R} \cup \{\infty\} \subset \mathbb{C} \cup \{\infty\} = S^2. \tag{A.3}$$

Since the surface is hyperelliptic, it is the smooth completion of the locus in $\mathbb{C}^2$ of *some* equation of the form (A.2), as discussed above.

Here P is of degree $2s+2$ with distinct roots, but otherwise to be determined. The set of roots of P is the set of (the z-coordinates of) the Weierstrass points. Hence the set of roots must be invariant under τ_0. Thus the roots of the polynomial either come in conjugate pairs, or else are real. Therefore P has real coefficients. Furthermore, the leading cofficient of P may be absorbed into the w-coordinate by extracting a square root. Here we may have to rotate w by i, but at any rate the coefficients of P remain real, and thus P can be assumed monic.

If P had a real root, there would be a ramification point fixed by τ_0. But then the corresponding Weierstrass point must be fixed by τ, as well! This contradicts the fixed point freeness of τ. Thus all roots of P must come in conjugate pairs. $\square$

Appendix B. A double fibration and integral geometry

In the proof of Pu's theorem, as well as our argument here (see Lemma 5.1), one uses the following integral geometry result. It has its origin in results of P. Funk [Fu16] determining a symmetric function on the two-sphere from its great circle integrals, *cf.* [He99, Proposition 2.2, p. 59], as well as Preface therein.

B.1. **A double fibration of $SO(3)$.** Note that the sphere S^2 is the homogeneous space of the Lie group $SO(3)$, so that $S^2 = SO(3)/SO(2)$. Denote by $SO(2)_\sigma$ the fiber over (stabilizer of) a typical point $\sigma \in S^2$. The projection

$$p : SO(3) \to S^2 \tag{B.1}$$

is a Riemannian submersion for the standard metric of constant sectional curvature $\frac{1}{4}$ on $SO(3)$. The total space $SO(3)$ admits another Riemannian submersion, which we denote

$$q : SO(3) \to \widetilde{\mathbb{R}P^2}, \tag{B.2}$$

whose typical fiber γ is an orbit of the geodesic flow on $SO(3)$ viewed as the unit tangent bundle of S^2.

Each orbit γ projects under p to a great circle on the sphere. We think of the base space $\widetilde{\mathbb{R}P^2}$ of q as the configuration space of oriented great circles on the sphere, with measure $d\gamma$. The diagram of Figure B.1 illustrates the maps defined so far.

B.2. **Integral geometry on S^2.** Let f be a square integrable function on S^2, which is positive and continuous except possibly for a finite number of points where f either vanishes or has a singularity of type $\frac{1}{\sqrt{r}}$, *cf.* Remark 4.2.

$$\begin{tikzcd}
SO(1) \arrow[dr] & & \left(\widetilde{\mathbb{R}P^2}, d\gamma\right) \\
& SO(3) \arrow[ur, "q"] \arrow[dr, "p"] & \\
\gamma \arrow[ur] \arrow[rr, "\text{great circle}"'] & & (S^2, d\sigma)
\end{tikzcd}$$

FIGURE B.1. Integral geometry on S^2, *cf.* (B.1) and (B.2)

Proposition B.1. *There is a great circle γ such that the following two equivalent inequalities are satisfied:*

1. *we have* $\left(\int_\gamma f(\gamma(t))dt\right)^2 \leq \pi \int_{S^2} f^2(\sigma)d\sigma$, *where t is the arclength parameter and $d\sigma$ is the Riemannian measure on the standard unit sphere.*
2. *there is a great circle of length L in the metric $f^2 g_0$, so that $L^2 \leq \pi A$, where A is the Riemannian surface area of the metric $f^2 g_0$.*

In the boundary case of equality in either inequality, the function f must be constant.

Proof. The proof is an averaging argument and shows that the average length of great circles is short.

Denote by $E_g(\gamma)$ the energy, and by $L_g(\gamma)$ the length, of a curve γ with respect to a (possibly singular) metric $g = f^2 g_0$. We apply Fubini's theorem [Ru87, P. 164] twice, to both p and q, to obtain

$$\begin{aligned}
\frac{\left(\int_{\widetilde{\mathbb{R}P^2}} L_g(\gamma)^2\right)}{2\pi} &\leq \int_{\widetilde{\mathbb{R}P^2}} E_g(\gamma)d\gamma \\
&= \int_{\widetilde{\mathbb{R}P^2}} d\gamma \left(\int_\gamma f^2 \circ p \circ \gamma(t)dt\right) \\
&= \int_{SO(3)} f^2 \circ p \\
&= \int_{S^2} f^2 \, d\sigma \left(\int_{SO(2)_\sigma} 1\right) \\
&= 2\pi \; area(g),
\end{aligned}$$

proving the formula since $area(\widetilde{\mathbb{R}P^2}) = 4\pi$. In the boundary case of equality, length and energy must be equal and therefore the function f

must be constant along every great circle, hence constant everywhere on S^2. □

Acknowledgments

The authors are grateful to Hershel Farkas, Shmuel Krushkal, and Steve Shnider for helpful discussions of hyperelliptic surfaces; to Mathieu Dutour for a helpful discussion of the combinatorial arguments of section 7; and to Franz Auer for the Super Bowl figures of section 5.

References

[Bab] Babenko, I. K.: Nature topologique des systoles unidimensionnelles, preprint.

[BCIK04] Bangert, V; Croke, C.; Ivanov, S.; Katz, M.: Boundary case of equality in optimal Loewner-type inequalities, in preparation.

[BK04] Bangert, V; Katz, M.: An optimal Loewner-type systolic inequality and harmonic one-forms of constant norm. `arXiv:math.DG/0304494`.

[BI94] Burago, D.; Ivanov, S.: Riemannian tori without conjugate points are flat. *Geom. Funct. Anal.* **4** (1994), no. 3, 259–269.

[BI95] Burago, D.; Ivanov, S.: On asymptotic volume of tori. *Geom. Funct. Anal.* **5** (1995), no. 5, 800–808.

[CK03] Croke, C.; Katz, M.: Universal volume bounds in Riemannian manifolds, *Surveys in Differential Geometry* **VIII**, Lectures on Geometry and Topology held in honor of Calabi, Lawson, Siu, and Uhlenbeck at Harvard University, May 3- 5, 2002, edited by S.T. Yau (Somerville, MA: International Press, 2003.) pp. 109 - 137. See `arXiv:math.DG/0302248`

[FK92] Farkas, H. M.; Kra, I.: Riemann surfaces. Second edition. Graduate Texts in Mathematics, 71. Springer-Verlag, New York, 1992.

[Fu16] Funk, P.: Uber eine geometrische Anwendung der Abelschen Integralgleichung. *Math. Ann.* **7**7 (1916), 129-135.

[Gr83] Gromov, M.: Filling Riemannian manifolds, *J. Diff. Geom.* **1**8 (1983), 1-147.

[Gr99] Gromov, M.: Metric structures for Riemannian and non-Riemannian spaces. Progr. in Mathematics, vol 152, Birkhäuser, Boston, 1999.

[GrH81] Gross, B. H.; Harris, J.: Real algebraic curves. *Ann. Sci. Ecole Norm. Sup.* (4) **1**4 (1981), no. 2, 157–182.

[He99] Helgason, S.: The Radon transform. 2nd edition. Progress in Mathematics, 5. Birkhäuser, Boston, Mass., 1999.

[Iv02] Ivanov, S.: On two-dimensional minimal fillings, *St. Petersbg. Math. J.* **1**3 (2002), no. 1, 17–25.

[IK04] Ivanov, S.; Katz, M.: Generalized degree and optimal Loewner-type inequalities, preprint.

[Jo48] John, F.: Extremum problems with inequalities as subsidiary conditions. Studies and Essays Presented to R. Courant on his 60th Birthday, January 8, 1948, 187–204. Interscience Publishers, Inc., New York, N. Y., 1948.

[Li69] Lichnerowicz, A.: Applications harmoniques dans un tore. *C.R. Acad. Sci., Sér. A* **2**69, 912–916 (1969).

[Mi95] Miranda, R.: Algebraic curves and Riemann surfaces. *Graduate Studies in Mathematics*, **5**. American Mathematical Society, Providence, RI, 1995.

[Pu52] Pu, P.M.: Some inequalities in certain nonorientable Riemannian manifolds, *Pacific J. Math.* **2** (1952), 55–71.

[Ri1854] Riemann, B.: Hypothèses qui servent de fondement à la géométrie, in Oeuvres mathématiques de Riemann. J. Gabay, 1990.

[Ru87] Rudin, W.: Real and complex analysis. Third edition. McGraw-Hill Book Co., New York, 1987.

Mathematisches Institut, Universität Freiburg, Eckerstr. 1, 79104 Freiburg, Germany

E-mail address: `bangert@mathematik.uni-freiburg.de`

Department of Mathematics, University of Pennsylvania, Philadelphia, PA 19104-6395 USA

E-mail address: `ccroke@math.upenn.edu`

Steklov Math. Institute, Fontanka 27, RU-191011 St. Petersburg, Russia

E-mail address: `svivanov@pdmi.ras.ru`

Department of Mathematics and Statistics, Bar Ilan University, Ramat Gan 52900 Israel

E-mail address: `katzmik@math.biu.ac.il`