A local systolic inequality and Gromov’s filling area conjecture

Olaf Müller∗

September 8, 2020

Abstract

The article treats some questions around Gromov’s filling area conjecture. It is shown that any filling with volume < 2\pi would not be attained. This result indicates the possibility of proof strategies using classical compactness arguments. The article also shows an a priori lower estimate on the total volume via a local systolic inequality, which is of independent interest. Furthermore, it is shown that one can assume that the boundaries in a vol-infimizing sequence of fillings are geodesic and that a lower bound for the systole holds. Finally, it is shown that any vol-infimizing sequence of fillings within a fixed conformal structure converges in L^1.

1 Introduction and results

For a Riemannian manifold (M, g), let d^g denote the induced metric as a function on M \times M. For a Riemannian manifold (N, h), an orientable manifold-with-boundary (M, g) is called a filling of (N, h) iff

- (\partial M, g|_{\partial M \otimes \partial M}) is (as a Riemannian manifold) isometric to (N, h), and
- (\partial M, d^g|_{\partial M \times \partial M}) is (as a metric space) isometric to (N, d^h).

Rephrasing the second condition, a filling of N cannot offer any shortcut for any pair of points on N. Let F be the set of fillings of the one-dimensional unit sphere S^1. In 1983, M. Gromov conjectured in a seminal article [2] that the volume of every element of F is greater or equal to 2\pi (the area of the round hemisphere). Let F_m be the set of metrics on a compact orientable surface M_m of genus m and one open disc removed (recall that any two such are diffeomorphic by a classical result of Radó), then obviously inf{vol(M, g)|g ∈ F} = inf_{m∈\mathbb{N}} inf_{g∈F_m} vol(M_m, g). In the article cited above, Gromov showed that inf_{g∈F_0} vol(M_0, g) = 2\pi. In 2001, S. Ivanov [3] proved the same assertion

∗Institut für Mathematik, Humboldt-Universität zu Berlin, Unter den Linden 6, 10099 Berlin (Germany), Email: mullerol@math.hu-berlin.de

1This is indeed a set if we define manifolds-with-boundary by means of gluing data.
where \( F_0 \) is replaced with the set of all Finsler metrics on the disc \( M_0 \). For genus 1 the result has also been shown \(^\text{1}\).

Here we want to contribute two aspects to the solution of the still open genus \( \geq 2 \) part of this conjecture. The first result is that if the infimum is not attained at the round hemisphere, then it is not attained at all:

**Theorem 1** If \( \text{vol} \) attains a minimum in \( F_m \) at a \( C^2 \) Riemannian manifold \((M,g)\) then \( m = 0 \) and \( \text{vol}(M,g) = 2\pi \).

As a consequence of the first theorem, one strategy to attack the conjecture could be to assume that we are given a sequence in \( F_m \) whose volumes converge to a number \( a < 2\pi \) and construct from it a convergent sequence whose volumes still converge to a number \( b < 2\pi \). Then an important tool would be a priori estimates. Therefore we want to derive a lower volume estimate like in Gromov’s article \(^\text{2}, 5.5A\ c\), where the antipodal map \( f := -1_{S} \) on \( S = \partial M \) is used to define the projection \( q : M \to M := M/f \). If we assume \( \text{genus}(M) = 0 \), we get that \( M \) is homeomorphic to a real projective plane, and as \( M \in F_0 \), we know that \( \text{sys}(M,p) \geq \pi \) for every \( p \in q(\partial M) \). But \( M \setminus q(\partial M) \) is contractible (as \( \text{genus}(M) = 0 \)), so each noncontractible curve in \( M \) intersects \( q(\partial M) \), thus

\[
\pi^2 \leq \inf \{ \text{sys}^2(M,p) | p \in q(\partial M) \} = \text{sys}^2(M) \leq \frac{\pi}{2} \text{vol}(M),
\]

where the last inequality is Pu’s systolic inequality which, in turn, uses as an essential ingredient conformal relatedness to a homogeneous metric over whose isometry group the conformal factor can then be averaged, whereas it is easy to see that for \( m > 1 \) the isometry group of a generic hyperbolic metric is empty. Instead, we apply the following theorem, which is of independent interest. For a Riemannian manifold \((X,g)\), \( A \subset X \) and \( p \in X \) we denote by \( B(p,r,g) \) the ball around \( p \) of radius \( r \) w.r.t. \( g \) and \( \mu_n(A,g) \) the \( n \)-dimensional Hausdorff measure, (omitting in each case the last slot where confusion is not possible), with \( l := \mu_1 \) and \( \text{vol} := \mu_n \) and define

\[
\Omega(X,A) := \{ c \in C^{0,1}(S^1,X) | c \text{ noncontractible } \land c(S^1) \cap A \neq \emptyset \},
\]

\[
\text{sys}(X,A) := \inf \{ l(c) | c \in \Omega(X,A) \}
\]

with \( \text{sys}(X,p) := \text{sys}(X,\{p\}) \) for \( p \in X \).

**Theorem 2** (Local systolic inequality) Let \( X \) be a two-dimensional manifold and let \( p \in X \). Then \( \text{vol}(B(p,R)) \geq R^2 \) for all \( R \in [0; \frac{1}{2} \text{sys}(X,p)] \).

**Remark.** Prop. 5.1.B of \(^\text{2}\) reads (without explicit proof) \( \text{vol}(B(p,\frac{1}{2} \text{sys}(X,p))) \geq \frac{1}{2} \text{sys}^2(X,p) \), which is apparently a misprint: A counterexample is a sequence of surfaces \( X_n \) of round spheres of radius \( r \) with some handles of total volume \( < 1/n \) attached in a ball of radius \( < 1/n \) around some point \( q \). Let \( p \) be a point of distance \( > 2\pi r - 1/n \) from \( q \), then \( \text{sys}(M,p,g_n) \to_{n \to \infty} 2\pi r \) and...
\( \text{vol}(M, g_n) \to n \to \infty 4\pi^2, \text{ so } \frac{\text{vol}(M, g_n)}{\text{sys}(M, p)} \to n \to \infty \frac{1}{\pi} < \frac{1}{\pi} \), contrary also to the other claim \( \text{vol}(X) \geq \frac{1}{2} \text{sys}(X, p) \) from the chain of inequalities in [2]. As the proof in [2] does not yield any bound of the volume in terms of the local systole (and as it is, other hand, difficult to derive lower volume estimates by classical Jacobi estimates in the absence of any control over the injectivity radius), we will modify Gromov’s proof via an appropriate refinement of the notions of tension and height. The inequality \( \text{vol}(B(p, \frac{1}{2}\text{sys}(X, p))) \geq \frac{1}{2}\text{sys}^2(X) \) in the same line, however, does follow in a straightforward manner from the rest of Prop. 5.1.B. in [2] taking into account \( h(v) \geq 2R - \frac{1}{\pi}(\partial B^+(v, R)) \) in Gromov’s terminology of height \( h \) (see below for a precise definition).

As a corollary, we obtain:

**Theorem 3 (A-priori volume estimate for \( g_n \))** \( g \in F_n \Rightarrow \text{vol}(g) \geq \pi^2/4. \)

**Proof.** Taking into account \( \text{vol}(M) = \text{vol}(\tilde{M}) \) and the fact that every simple closed geodesic through \( p \in q(\partial M) \) lifts via \( q \) to a geodesic between antipodal points on \( \partial M \), this follows from the fact that \( \text{sys}(\tilde{M}, p) \geq \pi \) for all \( p \in q(\partial M) \) and from Lemma 2.

Of course every vol-infimizing sequence satisfies a vol-bound from above. Furthermore, for each \( \epsilon > 0 \), by choosing an appropriate conformal factor on the parts with distance \( > \pi + \epsilon \) from the boundary, we can assume a diameter bound from above by \( \pi + \epsilon \). It should be noted that we have the equivalence \( g \in F_m \Leftrightarrow \text{sys}(\tilde{M}, g, p) < \pi \) for all \( p \in q(\partial M) \), because if there is a shortcut there is also a shortcut between antipodal points on the boundary.

If we are interested in a proof by contradiction as sketched above, we focus on a fixed genus: We put \( v(m) := \inf\{\text{vol}(g) | g \in F_m\} \) and \( m_0 := \min\{m \in \mathbb{N} | v(m) < v(0)\} \). Then the following proposition gives a lower bound on \( \text{sys}(g) \) for \( g \in F_{m_0} \):

**Proposition 4** \( \forall g \in F_{m_0} : \text{sys}(g) > \frac{v(0) - \text{vol}(g)}{4\pi}. \)

Furthermore, we can assume that the boundaries are geodesic:

**Proposition 5** There is a vol-infimizing sequence in \( F \) with geodesic boundary.

The proposition enables us to glue two copies of a \( C^1 \)-approximation of each metrics and so facilitates the application of the classical Koebe-Poincaré uniformization result, allowing for the conclusion that each metric is conformally related to a hyperbolic metric with geodesic boundary. And in view of Proposition 4 and the version of Mumford’s compactnes theorem with boundary [4] it seems a good idea to express each metric \( g_n \in F_{m_0} \) as conformal multiple of its hyperbolic counterpart \( g_n^0 \) with geodesic boundary as above. One can glue two copies of \( (M_n, g_n^0) \) to each other, obtaining a hyperbolic manifold \( N_n \) without boundary in which the boundary is mapped to a closed geodesic \( k \). In
the homotopy class of \( k \) there is no other closed geodesic, so \( k \) is minimal in its homotopy group, and consequently, \((M_m, g^n_0) \in F_m\).

For the purpose of careful interpretation of the following theorem, remind that for a Riemannian surface \((M, g)\) and \( f \in L^p(M)\), there is in general no well-defined notion of length of curves in \((M, fg)\).

**Theorem 6** For each conformal class \([g]\), the set \( A := \{ f \in C^0(M) | f^2 g_0 \in F_m \}\) is convex. Moreover, for any vol-infimizing sequence \( n \mapsto f^2_n \cdot g_0 \in F_m\), the sequence \( n \mapsto f_n \) converges in \( L^2(g)\). In particular, each conformal class in \( F \) contains at most one integrable vol-minimizer.

Note however that in general an infimizing sequence will not converge in \( C^0\). For \( f_n \in F_m \) we have \( h_n \geq f_n \Rightarrow h_n \in F_m \), thus for \( k_n \in C^0(M, \{0; \infty\}) \) with \( \sup(k_n) \geq n \) and \( |k_n| \leq 1/n \), the sequence \( n \mapsto f_n + k_n \) is \( I_m \) is still an infimizing sequence in \( F_m \) but does not converge in the sup norm.

## 2 Proofs

In the following, all curves are assumed to be Lipschitz. We denote the \( n\)-dimensional Hausdorff measure by \( \mu_n \), length of a curve \( c \) resp. volume of open subsets w.r.t. a Riemannian metric \( g \) by \( \text{vol}_g \) resp. \( l(c, g) \) or vol resp. \( l(c) \), if no confusion is possible.

**Proof of Theorem 6** Let \( N \subset M \) be a maximal Fermi (i.e., normal geodesic) neighborhood of \( S \) in \( M \), i.e., on \( \partial N \setminus S \) there is a non-regular point of the geodesic distance \( d_S \) to \( S \), whereas \( d_S \) is regular in the interior of \( N \).

**Lemma 7** There is an isometric \( S^1 \)-action on \((N, g)\).

**Proof of Lemma 7** In Fermi coordinates, \( N \) can be written \((\{0; a\} \times S^1, dt^2 + f(t, s)ds^2\) ). Each action as above must leave the geodesic distance to \( S \) invariant, therefore such an action exists if and only if \( f(t, s) = f(t, u) \) for all \( t \in [0; a) \) and all \( s, u \in S^1 \). On \( N \) we define a sequence of metrics: For all \( n \in \mathbb{N} \), let \( \psi_n \in C^\infty(\{0; a\}, [0; 1]) \) such that \( \psi_n(t) = 0 \) for all \( t \in [0; a - \frac{1}{n}] \) and \( \psi_n(t) = 1 \) for all \( t \in [a - \frac{1}{n}; a) \). Then we put

\[
f_n(t, s) := (1 - \psi_n(t)) \int_{S^1} f(t, u)du + \psi_n(t) \cdot f(t, s), \quad g_n(t, s) := dt^2 + f_n(t, s)ds^2.
\]

where the integral on \( S^1 \) is w.r.t. the Haar measure. Now obviously \( g_n \) can be extended to a smooth metric \( \tilde{g}_n \) by \( g_{|M \setminus N} \) as on a neighborhood of \( \partial N \), it coincides with the original metric. Furthermore, \( \tilde{g}_n \in F_m \) for all \( n \in \mathbb{N} \) can be shown via \( l(c, \frac{1}{2}(f_1 + f_2)g) \geq \frac{1}{2}((l(c, f_1) + l(c, f_2)) \) for every continuous piecewise \( C^1 \) curve \( c \), as an elementary arithmetic calculation reveals that for all \( a, b, r, u \in \mathbb{R}^+ \) we get
The last step of the proof is to realize that also the second fundamental form is of constant length for all $\tau \in [0; a]$, and by smoothness of $\exp$ this also holds for $\tau = a$, so $\gamma := \rho^a$ is of constant length. If $\gamma$ is nonconstant, there is no $S$-focal point in the image of $\gamma$, thus there are $x, y \in S^1$ with $\gamma(x) = \gamma(y) =: p$. Let $c$ resp. $k$ the minimal geodesic curve $[0; a] \to M$ from $r(x)$ resp. $r(y)$ to $p$, parametrized by arc length. Then

**Lemma 8** If $\gamma$ is nonconstant, then $c'(a) = -k'(a)$.

**Proof of Lemma 8** By the first variational formula for geodesics and the minimality of $a$, we obtain $k'(a), c'(a) \in (\gamma'(y))^\perp \cap (\gamma'(x))^\perp$, which implies that $c'(a)$ and $k'(a)$ are collinear. Uniqueness for the geodesic equation shows that they are not identical.

The last step of the proof is to realize that also the second fundamental form (this is the point where we need $C^2$ regularity of the metric) of $\rho^a$ is spherically invariant, so it only depends as a function $h(\tau)$ on $\tau$, for all $\tau \in [0; a)$, and then as above smoothness of $\exp$ and $r$ implies that also $\gamma = \rho^a$ has a constant second fundamental form. At $p$, two different branches of $\gamma$ osculate each other with the corresponding normal vectors $k'(a)$ and $c'(a) = -k'(a)$, thus $h(a) = -h(a)$ and therefore $\gamma$ is a geodesic, and according to the above, the curve $\gamma$ is an even multiple of a simple closed geodesic. Therefore the closure of $N$ is homeomorphic to a Klein bottle without a disc. But then $N$ cannot be embedded into an orientable surface, contradiction. So $\gamma$ is constant, and $M$ is homeomorphic to a disc.

**Proof of Theorem** First, for $p \in A \subset X$ and $c, k : S^1 \to X$ noncontractible with $c(0) = p$ we define $c \sim_A k$ iff there is a homotopy $H : I \times S^1 \to X, c \sim k$ with $H(t, s) = c(s)$ for all $s \in S^1 \setminus c^{-1}(A)$ and $H(t, s) \in A$ for all $t \in c^{-1}(A)$ (in other words, the homotopy is only allowed to move the part in $A$ and is itself constrained to $A$). Then we define

\[
\text{tension}(c, A) := \sup\{l(c_1) | c \sim_A c_1\},
\]
\[
\text{h}(p, A) := \inf\{\text{tension}(c, A) | c \text{ noncontractible } \land c(0) = p\}.
\]
For $r > 0$, we put $h(p, r) := h(p, B(p, r))$. Of course $h(p, r) \geq h(p, s)$ whenever $r \geq s$. Now let $r < \sys(X, p)/2$. Then let $c$ be a noncontractible curve of minimal length through $p$, i.e. $\ell(c) = \sys(M, p)$. For every homotopy $H$ from $c$ as in the definition of $h(p, r)$, we have that $k := H(1, \cdot)$ meets $B(p, r)$ but cannot be constrained to it as $k$ is noncontractible whereas $B(p, r)$ is. Thus $\ell(k) \geq \sys(X, \partial B(p, r))$. Thus tension($c$) \leq \sys(X, p) - \sys(X, \partial B(p, r)) and

\[ h(p, r) \leq \sys(X, p) - \sys(X, \partial B(p, r)). \] (1)

Furthermore, we see as above that $B(p, \sys(X, p)/2)$ does not contain a noncontractible curve, so any $c$ as in the definition of $\sys(X, \partial B(p, r))$ has to run twice between $\partial B(p, r)$ and $\partial B(p, \sys(X, p)/2)$, thus we get

\[ \sys(X, \partial B(p, r)) \geq 2(\frac{1}{2}\sys(X, p) - r). \] (2)

Let $r < \frac{1}{2}\sys(X, p)$, then $B(p, r)$ is contractible in $X$, thus for every connected component $S_i$ of $\partial B(p, r)$, there is a $D_i \subset X$ homeomorphic to a disc with $\partial D_i = S_i$, and then

\[ B^+(p, r) := B(p, r) \cup \bigcup_{i \in I} D_i \]

(where $I$ parametrizes the connected components of $\partial B(p, r)$) is contractible in itself, thus homeomorphic to a disc. Note that $B(p, r) \subset B^+(p, r)$ whereas $\partial B^+(p, r) \subset \partial B(p, r)$ is homeomorphic to a one-sphere. Now let $c : [0; 1] \to X$ be a noncontractible closed curve through $p = c(0) = c(1)$. As $B^+(p, r)$ is contractible, $K := c^{-1}(\partial B^+(p, r)) \neq \emptyset$, so we define $t_\pm \in [0; 1]$ by $t_- := \min K$ and $t_+ := \max K$. Let $x_\pm := c(t_\pm)$, then $\partial B^+(p, r) \setminus \{x_+\} \setminus \{x_\pm\}$ consists of two connected components $L_1, L_2$ with $\mu_1(L_1) = \mu_1(\partial B^+(p, r))$, w.l.o.g. assume $l_1 := \mu_1(L_1) \leq \mu_1(\partial B^+(p, r)) \geq 2\mu_1(L_2)$. Then there is a homotopy in $X$ from $c$ to $\tilde{c} := c|_{[t_-, t_+]} \circ k$ where $k$ is a curve running along $L_1$ (changing only the parts in the ball), thus

\[
tension(c) \geq \ell(c) - \ell(\tilde{c}) = \ell(c|_{[0, t_-]}) + \ell(c|_{[t_+, 1]})) - l_1 \geq 2r - l_1 \geq 2r - \frac{1}{2}\ell(\partial B^+(p, r)).
\]

Therefore,

\[ l(\partial B^+(p, r)) \geq 4r - 2h(p, r). \] (3)

As $d(p, \cdot)$ is a Lipschitz function with gradient of norm 1 where it is differentiable, we finally can apply Federer’s coarea formula of geometric measure theory and obtain
\[ \text{vol}(B(p, R)) = \int_0^R \mu_1(\partial B(p, r))dr \]
\[ \geq \int_0^R \mu_1(\partial B^+(p, r))dr \]
\[ \geq_{\text{Eq.} 5} \int_0^R (4r - 2h(p, r))dr \]
\[ \geq_{\text{Eq.} 1} \int_0^R (4r - \text{sys}(M, p) + \text{sys}(V, \partial B(M, p)))dr \]
\[ \geq_{\text{Eq.} 2} \int_0^R (4r - \text{sys}(M, p) + \text{sys}(M, p) - 2r)dr \]
\[ = R^2, \]

which concludes the proof. \(\square\)

**Proof of Prop. 4.** Let \( g \in F_{m_0} \) with \( v = v(0) - \text{vol}(g) > 0 \) and let \( c \) be a noncontractible closed curve. If we cut \( M \) along \( c \) and fill the new boundary components with two hemispheres, the so obtained manifold \( \hat{M} \) has smaller genus than \( M \) and volume \( v(g) + 4\pi l(c) \), which consequently must be larger or equal to \( v(0) \). \(\square\)

**Proof of Prop. 5.** Let \( n \mapsto g_n \) be a vol-inifirmizing sequence sequence in \( F_m \) with \( v(g_n) < v(m) + 1/n \). Then for each \( n \in \mathbb{N} \), we have a Fermi neighborhood \( U_n \) of \( \partial S \) isometric to \( (S^1 \times [0; \epsilon_n], dt^2 + u_n(s, t)ds^2 \) for \( s \in S^1, t \in [0; \epsilon_n] \). Via the extension procedure in [5], we can smoothly extend \( (U_n, g_n) \) by \( \hat{U}_n := (S^1 \times [-(a_n + 2/n); \epsilon_n], dt^2 + \hat{u}_n(s, t)ds^2 \) for \( \hat{u}_n \in C^\infty(S^1 \times [-(a_n + 2/n); \epsilon_n]) \) with \( u_n(s, t) \geq 1 - 1/n\nu(s, t) \in \mathbb{R} \times [-(a_n + 2/n); 0] \) and \( u_n(s, t) = 1 - 1/n\nu(s, t) \in \mathbb{R} \times [-(a_n + 2/n); -a_n] \). Then for \( \hat{M}_n \) constructed via the above isometry out of \( \hat{M}_n \) and \( \hat{U}_n \), the boundary \( \partial \hat{M}_n \) is geodesic, and every possible shortcut between two points on \( \partial \hat{M}_n \) would have to meet \( \partial \hat{M}_n \) at two points, still being a shortcut for those, thus \( \hat{M}_n \in F_m \) for all \( n \). Furthermore, as we can choose \( a_n \) as small as desired, the increase in volume can be chosen smaller than \( 3/n \). \(\square\)

**Proof of Theorem 6.** Convexity is readily shown: For \( f_1, f_2 \in F_m \) and \( \alpha \in [0; 1] \), then for any curve \( c \in M \) we get \( l_{\alpha f_1 + (1-\alpha) f_2}(c) = \alpha l_{f_1}(c) + (1- \alpha) l_{f_2}(c) \) and \( (\alpha f_1 + (1-\alpha) f_2)_M = \hat{f}_1 g \phi_M \), thus \( (\alpha f_1 + (1-\alpha) f_2)_M g \in F_m \). The second assertion follows from a standard convexity argument. Let \( \epsilon > 0 \) and let \( u, w \in C^0(M) \) with \( f_{(M, g)} w^2, f_{(M, g)} w^2 \leq (1 + \epsilon)v(m) \). Then from \( v(m) \leq \frac{1}{2}(u + w)^2 = \frac{1}{2}(|u|^2 + |w|^2 + \langle u, w \rangle) \) we conclude
\[ |u - w|^2 = |u|^2 + |w|^2 - 2\langle u, w \rangle \leq 2(|u|^2 + |w|^2) - 4 \leq 4(1 + \epsilon)^2 - 4 \]
\[ = 8\epsilon + 4\epsilon^2 \to_{\epsilon \to 0} 0, \]

which concludes the proof. \(\square\)
References

[1] Victor Bangert, Christopher B. Croke, Sergei Ivanov, Mikhail G. Katz: *Filling area conjecture and ovalless real hyperelliptic surfaces*, Geom. Funct. Anal. 15 (3): 577 — 597 (2005). arXiv:math/0405583

[2] Mikhail Gromov: *Filling Riemannian Manifolds*, J. Diff. Geom. 18 (1): 1 — 147 (1983)

[3] Sergei V. Ivanov: *Filling minimality of Finslerian 2-discs*, Proc. Steklov Inst. Math. 273 (1): 176 — 190 (2011). arXiv: 0910.2257

[4] Lixin Liu, Athanase Papadopoulos, Weixu Su, Guillaume Thret: *Length spectra and the Teichmüller metric for surfaces with boundary*, Monatsh. Math (2010) 161:295 — 311. arXiv: 0904.2370

[5] Olaf Müller: *Cheeger-Gromov compactness for manifolds with boundary*, arXiv:1808.06458