The homology systole of hyperbolic Riemann surfaces

Hugo Parlier

Abstract   The main goal of this note is to show that the study of closed hyperbolic surfaces with maximum length systole is in fact the study of surfaces with maximum length homological systole. The same result is shown to be true for once-punctured surfaces, and is shown to fail for surfaces with a large number of cusps.

1 Introduction

There is a natural function on moduli space, called the systole function, which associates to a hyperbolic surface the length of its shortest non-trivial closed curve. Unless the hyperbolic surface is topologically a pair of pants, the curve in question is a simple closed geodesic generally called the systole or the systolic loop. Although there exist surfaces with arbitrarily small systole, the systole function is bounded over any given moduli space (of complete finite area surfaces of a given signature). Furthermore via Mumford's compactness theorem, there is (at least) a surface in each moduli space which realizes the maximum length systole. The study of these surfaces, and more generally the study of critical points of the systole function, largely initiated by Schmutz Schaller [15], has generated quite a bit of interest and can be thought of as a type of hyperbolic sphere packing problem. Notable results include a sharp upper bound on the systole among all surfaces of genus 2 [9], the fact that principal congruence subgroups of $PSL_2(\mathbb{Z})$ give rise to global maxima in their respective signatures

H. Parlier (✉)
Department of Mathematics, University of Fribourg, Fribourg, Switzerland
e-mail: hugo.parlier@gmail.com

[1,16] and Akrout's theorem that the systole function is in fact a topological Morse function [2] (partial results previously due to Schmutz Schaller [17]).

If we define sys_g (or more generally $\text{sys}_{g,n}$) to be the maximum length of a systole among all hyperbolic surfaces of genus g (resp. of genus g with n cusps), it is also an interesting question to ask how these constants grow as a function of topology. As it turns out, the more interesting question is how these constants grow as a function of genus, because for large enough n, they cease to grow [1,16]. Note that the systole length of a closed hyperbolic surface is exactly twice the length of the minimum injectivity radius. Thus if a surface has a systole of length ℓ, then around any point of the surface, there is an embedded open disk of radius $\frac{\ell}{2}$. By considering the area of a disk around a point in the hyperbolic plane, which grows roughly exponentially in radius, one immediately sees that the systole function is bounded by roughly $\log(g)$. In the more general case of closed Riemannian surfaces with area normalized to $4\pi(g-1)$, the result is also true and this is a theorem of Gromov [7]. Conversely, there are constructions of families of surfaces, one in each genus, where the systoles grow roughly like $\log(g)$. The first of these constructions was due to Buser and Sarnak [4], and there have been others since [3,11]. Both Buser–Sarnak, and Gromov in the more general setting of Riemannian metrics, also considered the homological systole $\text{sys}^h(S)$, i.e., the shortest homologically non-trivial curve on a surface S of genus g. Of course one has the obvious inequality $\text{sys}(S) \leq \text{sys}^h(S)$ for any surface S and it is easy to construct surfaces where the inequality is an equality, resp. where the inequality is strict. Notice that the embedded disk argument above does not necessarily give a homologically non-trivial curve. Nonetheless, as in the case of the homotopy systole, it is not too difficult to find a rough $\log(g)$ upper bound on the homology systole of a hyperbolic surface of genus g [4]. Again, the rough $\log(g)$ bound remains true in the setting of Riemannian metrics of normalized area [8, 2.C].

The goal of this note is to observe that in any signature, there are surfaces which realize the supremum of the homological systole function $\text{sys}^h_{g,n}$ ($\text{sys}^h_{g,0} = \text{sys}^h_g$ in the closed case) and that in the closed and once-punctured cases, these are the same surfaces that realize the maximum homotopy systole. Specifically, in Sect. 2, the following is shown.

Theorem 1.1 *If S is maximal for* sys *among all closed genus g (resp. genus g with one cusp) hyperbolic surfaces, then it is maximal for* sys^h*. Thus* $\text{sys}_g = \text{sys}^h_g$ *and* $\text{sys}_{g,1} = \text{sys}^h_{g,1}$*.*

However, it is clear that one cannot hope to generalize the above result to arbitrary signature. Indeed, via Buser's hairy torus examples [5,6], one can construct a family of surfaces of genus 1 with n cusps with homology systoles of length roughly $\sqrt{n}$. In contrast, the homotopy systole of a surface of genus 1, no matter how many cusps it has, is uniformly bounded. In Sect. 3, this example is adapted to arbitrary genus to show $\text{sys}^h_{g,n} > \text{sys}_{g,n}$ for $n \geq 25g$.

2 Proof of Theorem 1.1

We'll begin by showing that like the usual systole, the homological systole admits a maximum over the moduli space of genus g hyperbolic surfaces with n cusps. For the systole function, this is an immediate consequence of the continuity of the systole function and Mumford's compactness theorem [12] which states the set of surfaces with injectivity radius bounded below is a compact subset of moduli space. Here we need to be more careful because a priori surfaces could be arbitrarily close to the supremum of the homological systole and have a an arbitrarily small systole (which in this case would be homologically trivial). The following lemma will allow us to again apply Mumford's compactness theorem. Before stating the

lemma, we set $\mathcal{M}_{g,n}^{\varepsilon}$ to be the set of all surfaces of genus g with systole bounded below by ε and $\left(\mathcal{M}_{g,n}^{\varepsilon}\right)^c$ to be the set of all surfaces with systole strictly less than ε.

Lemma 2.1 *For each signature* (g, n)*, there is an* $\varepsilon_{g,n} > 0$ *such that*

$$\sup_{S \in \left(\mathcal{M}_{g,n}^{\varepsilon_{g,n}}\right)^c} \mathrm{sys}^h(S) < \sup_{S \in \mathcal{M}_{g,n}^{\varepsilon_{g,n}}} \mathrm{sys}^h(S).$$

Before giving the proof of the lemma, we recall the following lemma (see for instance [13, 14, 18]).

Lemma 2.2 [Length expansion lemma] *Let* S *be a surface with* $N > 0$ *disjoint simple closed geodesics* $\gamma_1, \ldots, \gamma_N$ *of lengths* $\ell_1, \ldots, \ell_N$*. For* $(\delta_1, \ldots, \delta_N) \in (\mathbb{R}^+)^N$ *with at least one* $\delta_k \neq 0$*, there exists a surface* $\tilde{S}$ *with* $\ell_{\tilde{S}}(\gamma_1) = \ell_1 + \delta_1, \ldots, \ell_{\tilde{S}}(\gamma_N) = \ell_N + \delta_N$ *and all simple closed geodesics of* $\tilde{S}$ *disjoint from* $\gamma_1, \ldots, \gamma_N$ *of length strictly greater than their length on* S*.*

Proof of Lemma 2.1 Set

$$\varepsilon_{g,n}' = \frac{1}{2} \sup_{S \in \mathcal{M}_{g,n}} \mathrm{sys}^h(S) = \frac{1}{2} \mathrm{sys}_{g,n}^h$$

and consider a surfaces in $\left(\mathcal{M}_{g,n}^{\varepsilon_{g,n}'}\right)^c$. Clearly, the supremum of sys^h among all such surfaces with homological systole less than $\varepsilon_{g,n}'$ will be strictly less than $\mathrm{sys}_{g,n}^h$. We only need to worry about surfaces with homologically trivial systole in $\left(\mathcal{M}_{g,n}^{\varepsilon_{g,n}'}\right)^c$. Let S be such a surface.

The collar lemma [10] ensures that if a simple closed geodesic $\gamma \subset S$ is sufficiently short, then any simple closed geodesic that intersects it transversally is long (where long depends only on how short the geodesic is). In particular, as sys^h is bounded above over $\mathcal{M}_{g,n}$, if γ is less than a certain constant $\varepsilon_{g,n}''$ (which depends only on the topology) then all of the surface's homological systoles are disjoint from γ. We can now apply Lemma 2.2 to increase the (homologically trivial) systoles of S by some small $\delta > 0$ which will give us a new surface with $\mathrm{sys}^h(S') > \mathrm{sys}^h(S)$ and $\mathrm{sys}(S') = \mathrm{sys}(S) + \delta$. This process can be repeated until either the systole of the surface is now $\varepsilon_{g,n}'$, or there is a homological systole which crosses each of the homologically trivial systoles. In the latter case, both the homologically trivial and non-trivial systoles are at least of length equal to the constant $\varepsilon_{g,n}''$. We set

$$\varepsilon_{g,n} = \min\left(\varepsilon_{g,n}', \varepsilon_{g,n}''\right)$$

and the lemma is proved. □

We've shown that $\sup_{S \in \mathcal{M}_{g,n}}(\mathrm{sys}_h(S)) = \sup_{S \in \mathcal{M}_{g,n}^{\varepsilon_{g,n}}}(\mathrm{sys}_h(S))$ thus via Mumford's compactness theorem, we can now conclude that there are surfaces which realize the maximum size homological systole in every signature.

A closed geodesic γ is said to be *straight* if for any two of its points p, q, any distance realizing path is a sub-arc of γ. Homological systoles *always* have this property [7]. The following is well known for closed surfaces, and we provide a proof that includes once-punctured surfaces.

Lemma 2.3 *On closed and once-punctured surfaces, systoles are straight.*
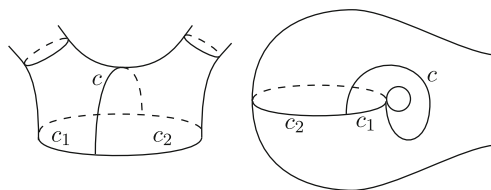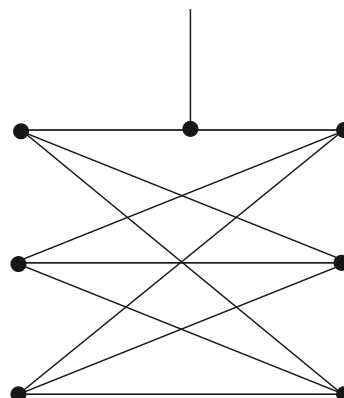
Fig. 1 The two cases for c, c_1, c_2

Fig. 2 The modified K_{33} graph

Proof of Lemma 2.3 Suppose that a systole σ is not straight, thus there is a path c between two of its points p and q of length less than the length of the shortest of the two paths c_1 and c_2 of σ between the same points (Fig. 1).

Consider the two homotopy classes $c \cup c_1$ and $c \cup c_2$. They are either non-trivial or are parallel to a cusp. If the surface has at most one cusp, one of them, say $c \cup c_1$ must be non-trivial. The geodesic $\tilde{\sigma}$ in the homotopy class satisfies $\ell(\tilde{\sigma}) < \ell(c) + \ell(c_1) \leq \ell(\sigma)$, a contradiction. □

Remark 2.4 The lemma above cannot be generalized to surfaces with multiple punctures. To see this, consider the following twice-punctured surface which is constructed as follows.

We begin with any finite trivalent graph G of girth at least 4 (the girth is the shortest non-trivial cycle). (An example of such a graph is the complete bipartite $K_{3,3}$ graph: in fact, it is the smallest trivalent graph with girth 4 and it has exactly 6 vertices.) A new graph $\tilde{G}$ is obtained by removing a single edge and replacing it with a tripod (see Fig. 2). This operation does not decrease the girth of the graph.

We view the resulting graph as the graph of a pants decomposition, and construct a surface by inserting (hyperbolic) pants with boundary lengths all equal to $4 \operatorname{arcsinh} 1$ in the usual way, without paying any attention to twist parameters. The resulting surface B has a single boundary geodesic of length $4 \operatorname{arcsinh} 1$. To obtain a surface with 2 cusps, we glue a pair of pants with 2 cusps and a boundary geodesic of length $4 \operatorname{arcsinh} 1$ along the boundary geodesic of B. As an illustration, if we choose the $K_{3,3}$ graph to begin with, the resulting surface has a total of 8 pairs of pants and is of genus 4.

Our claim is that the systoles of this surface are exactly the N ($= 11$ in the $K_{3,3}$ case) geodesics of length $4 \operatorname{arcsinh} 1$ which were the boundary geodesics of the pants. We'll denote these curves $\sigma_1, \ldots, \sigma_N$ with σ_1 being the geodesic that forms a pair of pants with the two cusps (Fig. 3).

4

Fig. 3 Adding a cusped pair of pants
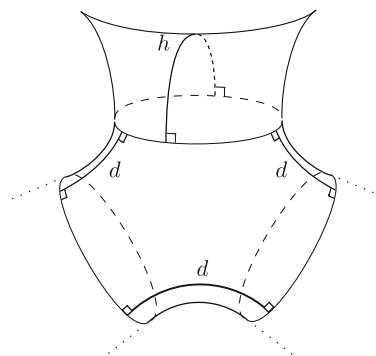
To see that these curves are indeed systoles, we will show that any geodesic that intersects one of these curves is strictly longer. We begin by taking any curve γ that crosses one of the curves σ_k and has an arc c that leaves and comes back on the same side of σ_k. (In the event where this curve is σ_1 we consider the arc of γ not contained in the pair of pants with 2 cusps.) We replace c with the unique shortest geodesic $\tilde{c}$ in the free homotopy class where the endpoints are allowed to slide on σ_k. Now $\tilde{c}$ separates σ_k into two arcs, one of which, say $\hat{c}$, has length less than $2\,\mathrm{arcsinh}1$. Via standard hyperbolic trigonometric arguments, the unique geodesic in the homotopy class of $\tilde{c} \cup \hat{c}$ has length less than $\ell(\tilde{c})$, thus less than $\ell(c)$, and less than $\ell(\gamma)$. It follows that we can restrict our attention to curves that don't have this "backtracking" property.

Such curves describe non-trivial cycles in the underlying graph, and have cycle length at least 4. It follows that they pass through at least 4 pairs of pants and as a consequence, their length is at least 4 times the shortest distance between boundary curves of the pants. Again, via standard hyperbolic trigonometry, this distance is $d = 2\,\mathrm{arcsinh}(\frac{1}{2})$. These curves are thus at least of length $8\,\mathrm{arcsinh}(\frac{1}{2}) > 4\,\mathrm{arcsinh}1$.

We conclude by observing that although the curve σ_1 is a systole, it is not straight. Indeed, on the pair of pants with σ_1 and the two cusps, there is a unique simple geodesic path h, perpendicular to σ_1 in both endpoints. Again, via hyperbolic trigonometry its length is exactly $2\,\mathrm{arcsinh}1$. It is thus a distance realizing path between the endpoints and is not contained in σ_1.

We now use Lemma 2.3 to show the following.

Lemma 2.5 *Let γ be the homological systole of a surface S with at most one puncture. Then if δ is the systole of S, then γ and δ intersect at most once.*

Proof of Lemma 2.5 We proceed by contradiction. Suppose γ and δ as above intersect more than once. If we cut the surface along γ then δ is cut into at least 2 arcs. One of these arcs, say c, has the property of being of length less than $\frac{1}{2}\ell(\delta) \leq \frac{1}{2}\ell(\gamma)$. This arc c can be of two types: either c has its endpoints on the two copies of γ, or it has both endpoints on the same copy.

In the first case, we consider the shortest arc $\tilde{c}$ of γ between the two endpoints of c. Clearly $\ell(\tilde{c}) < \ell(\gamma)$. Note that $c \cup \tilde{c}$ describe a simple closed path, and consider the geodesic $\tilde{\gamma}$ in the homotopy class of $c \cup \tilde{c}$. By construction we have $i(\gamma, \tilde{\gamma}) = 1$ and $\ell(\gamma) < \frac{1}{2}(\ell(\gamma) + \ell(\delta)) \leq \ell(\gamma)$. Observe that a curve that essentially intersects another curve exactly once is not only homotopically non-trivial, but also homologically non-trivial (a separating curve essentially intersects any other curve at least twice).

5

In the second case, consider as before the arc c, the shortest of the arcs of δ obtained by cutting along γ. Consider c_1 and c_2 the two arcs of γ between the endpoints of c. Consider the two geodesics γ_1 and γ_2 in the homotopy class of respectively $c \cup c_1$ and $c \cup c_2$. Observe that

$$\ell(\gamma_1), \ell(\gamma_2) < \ell(\gamma).$$

In homology, one can orient the curves so that γ can be written as the sum of γ_1 and γ_2. This is because the three are the boundary curves of a pair of pants. The boundary curves of a pair of pants, for a certain given orientation, form a multicurve which is trivial in homology. It follows that if γ is non-trivial in homology, then γ_1 and γ_2 cannot both be trivial. As both are of length strictly less than γ, we obtain a contradiction. □

We have the obvious inequalities $\mathrm{sys}_{g,n} \leq \mathrm{sys}_{g,n}^h$. We will now proceed to show that $\mathrm{sys}_g \geq \mathrm{sys}_g^h$, resp. $\mathrm{sys}_{g,1} \geq \mathrm{sys}_{g,1}^h$. Consider a maximal surface $S_{\max}$ for the homology systole in $\mathcal{M}_{g,n}$ for $n = 0$ or $n = 1$.

Claim *All systoles of $S_{\max}$ are homologically non-trivial.*

Remark 2.6 Note that the claim implies the desired inequalities for $n = 0, 1$:

$$\mathrm{sys}_{g,n}^h = \mathrm{sys}^h(S_{\max}) = \mathrm{sys}(S_{\max}) \leq \mathrm{sys}_{g,n}$$

and this concludes the proof of Theorem 1.1.

Proof of claim: We shall proceed by contradiction. Suppose $S_{\max}$ has a homologically trivial systole, i.e., a separating systole δ. As $S_{\max}$ is maximal, the curve δ must cross a homological systole γ, otherwise by the length expansion lemma, one can increase δ to strictly increase the length of all simple closed geodesics that do not cross δ, thus including all homological systoles. Now δ is separating and essentially intersects γ, thus δ must intersect γ at least twice which by Lemma 2.5 is a contradiction. □

3 Surfaces with punctures

One could ask what happens for multiply punctured surfaces. What fails in the proof is that for $n \geq 2$, the systole of a surface is no longer necessarily straight. Note that the equality may in fact hold for $n = 2$, only the method given here doesn't work. A general inequality of the form $\mathrm{sys}_{g,n}^h = \mathrm{sys}_{g,n}$ is deemed to fail however, as will be explained in what follows.

The basic reason for this is that $\mathrm{sys}_{g,n}^h$ is a strictly growing function of n and in contrast $\mathrm{sys}_{g,n}$ is uniformly bounded (by a function of g). One way of making this effective is by examining Buser's hairy torus examples (see [6]). We recall briefly the construction and features of these surfaces.

One begins by constructing a hyperbolic "square" (a right angled quadrilateral with equal length sides) with a cusp in the middle. Each side of the square can be taken to be of length $2\,\mathrm{arcsinh}1$. One then constructs an $m \times m$ checker board using m^2 of these squares (Fig. 4). The board now has its 4 sides of length $2m\,\mathrm{arcsinh}1$. A torus is then obtained by gluing the opposite sides in the obvious way. Now if one cuts along a non-homologically trivial curve on this surface, the genus must be reduced. Such a curve must cross either every horizontal line or every vertical line. As such, it must have length at least $2m\,\mathrm{arcsinh}1$.

To obtain surfaces of genus g surfaces which enjoy the same property, one can paste together g copies of a hairy torus. To paste two tori together, one can replace a cusp on each

Fig. 4 The schematics of a hairy torus—the circles represent cusps
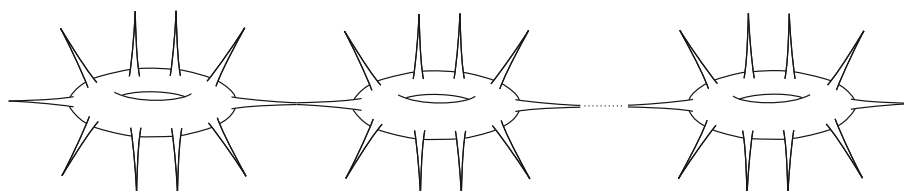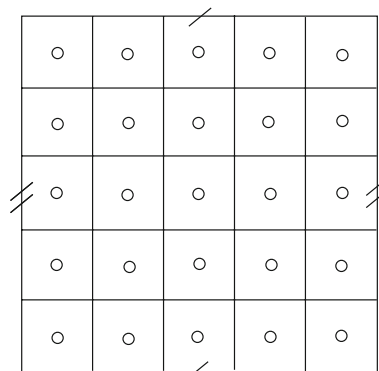

Fig. 5 The hairy genus g surface

torus by a very short geodesic in the standard way. The very short geodesic is homologically trivial, and again by the collar lemma, any curve that crosses it must be very long. In particular, a homology systole will not cross this curve and will thus remain in one of the two tori. One now repeats the construction to obtain a string of hairy tori (see Fig. 5). The resulting surface is of signature (g, gm^2) with homology systole at least $2m$ arcsinh1.

In contrast, note that for fixed genus, the systole length of a surface is uniformly bounded. More specifically, Schmutz Schaller [16] proved that the systole of a surface of signature (g, n) is bounded by $4 \operatorname{arccosh}((6g - 6 + 3n)/n)$ (which is uniformly bounded by a constant which depends on g). For each g, we can now compute the minimal m for which $2m \operatorname{arcsinh}1 > 4 \operatorname{arccosh}((6g - 6 + 3n)/n)$ with $n = gm^2$. The first m for which this occurs is $m = 5$. As a result of this construction, and the monoticity of $\operatorname{sys}_{g,n}^h$ in n, we obtain the following.

Proposition 3.1 *For all $n \geq 25g$, $\operatorname{sys}_{g,n}^h > \operatorname{sys}_{g,n}$.*

The bound in the above proposition is certainly far from being sharp. It might be interesting to find for any given genus g what the first n_g is for which we have $\operatorname{sys}_{g,n_g}^h > \operatorname{sys}_{g,n_g}$.

References

1. Adams, C.: Maximal cusps, collars, and systoles in hyperbolic surfaces. Indiana Univ. Math. J. **47**(2), 419–437 (1998)
2. Akrout, H.: Singularités topologiques des systoles généralisées. Topology **42**(2), 291–308 (2003)
3. Brooks, R.: Platonic surfaces. Comment. Math. Helv. **74**(1), 156–170 (1999)

4. Buser, P., Sarnak, P.: On the period matrix of a Riemann surface of large genus. Invent. Math. **117**(1), 27–56 (1994). With an appendix by J. H. Conway and N. J. A. Sloane

5. Buser, P.: Riemannshe flächen und längenspektrum vom trigonometrishen standpunkt, Habilitation Thesis, University of Bonn (1981)

6. Buser, P.: Geometry and spectra of compact Riemann surfaces, progress in mathematics, vol. 106. Birkhäuser Boston Inc., Boston, MA (1992)

7. Gromov, M.: Filling Riemannian manifolds. J. Differ. Geom. **18**(1), 1–147 (1983)

8. Gromov, M.: *Systoles and intersystolic inequalities*, Actes de la Table Ronde de Géométrie Différentielle (Luminy, 1992), Sémin. Congr., vol. 1, Soc. Math. France, Paris, 1996, pp. 291–362

9. Jenni, F.: Über den ersten Eigenwert des Laplace-Operators auf ausgewählten Beispielen kompakter Riemannscher Flächen. Comment. Math. Helv. **59**(2), 193–203 (1984)

10. Keen, L.: Collars on Riemann Surfaces, Discontinuous groups and Riemann Surfaces. (Proc. Conf., Univ. Maryland, College Park, Md., 1973), Princeton Univ. Press, Princeton, N.J., 1974, pp. 263–268. Ann. of Math. Studies, No. 79. MR 52 #738

11. Katz, M.G., Schaps, M., Vishne, U.: Logarithmic growth of systole of arithmetic Riemann surfaces along congruence subgroups. J. Differ. Geom. **76**(3), 399–422 (2007)

12. Mumford, D.: A remark on Mahler's compactness theorem. Proc. Am. Math. Soc. **28**, 289–294 (1971)

13. Parlier, H.: Lengths of geodesics on Riemann surfaces with boundary. Ann. Acad. Sci. Fenn. Math. **30**(2), 227–236 (2005)

14. Papadopoulos, A., Théret, G.: Shortening all the simple closed geodesics on surfaces with boundary. In: Proceedings of the AMS, vol. 113 (2010) (To appear)

15. Schmutz, P.: Riemann surfaces with shortest geodesic of maximal length. Geom. Funct. Anal. **3**(6), 564–631 (1993)

16. Schmutz, P.: Congruence subgroups and maximal Riemann surfaces. J. Geom. Anal. **4**(2), 207–218 (1994)

17. Schaller, P.S.: Systoles and topological Morse functions for Riemann surfaces. J. Differ. Geom. **52**(3), 407–452 (1999)

18. Thurston, W.: A spine for Teichmüller space, preprint