ISSN 1472-2739 (on-line) 1472-2747 (printed)

Algebraic & Geometric Topology
Volume 1 (2001) 349–368
Published: 2 June 2001

A characterization of shortest geodesics on surfaces

Max Neumann-Coto

Abstract Any finite configuration of curves with minimal intersections on a surface is a configuration of shortest geodesics for some Riemannian metric on the surface. The metric can be chosen to make the lengths of these geodesics equal to the number of intersections along them.

AMS Classification 53C22; 53C42,57R42

Keywords Surfaces, curves, geodesics, minimal intersections, metrics

If S is a closed surface, with some Riemannian metric, then each essential curve immersed in S is freely homotopic to a smooth geodesic in S which is shortest among all the curves in that homotopy class. So while closed geodesics represent the critical points of the length in each free homotopy class, these s*hortest geodesics* represent the absolute minima in each class. Shortest geodesics have topological properties which are not shared by all geodesics: Freedman, Hass, Rubinstein and Scott showed in [1] and [2] that shortest geodesics intersect minimally, i.e., they have the minimum number of intersections and self-intersections allowed by their free homotopy classes, unless they factor through coverings of other shortest geodesics. The curve in figure 1a, for example, represents a geodesic for some metric on S, but it can't represent a shortest geodesic. On the other hand, when a homotopy class allows different configurations with minimal intersections, as in figures 1b, 1c and 1d, it seems natural to ask which ones correspond to shortest geodesics for some Riemannian metric on S. This question was first considered by Shepard [5].

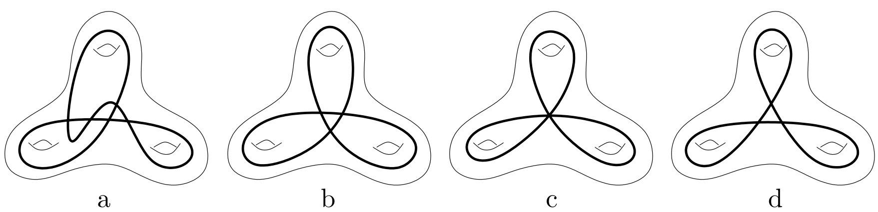

Figure 1

In this paper we prove that the shortest geodesics in a surface are characterized by the minimal intersection property, showing that any finite configuration of curves with minimal intersections in S is a configuration of shortest geodesics for some Riemannian metric g on S, and also that g can be chosen to make the lengths of these geodesics equal to the number of intersections along them. The proof starts by 'blowing up' the metric outside a regular neighborhood of the curves (an idea introduced by Bonahon in the context of least area surfaces in 3-manifolds) to transform the problem into a combinatorial one.

The main result implies that all minimal configurations in S can be extended to contain curves in any other homotopy classes, and gives conditions for the existence of 'absolute' inequalities relating the minimal lengths of curves in different homotopy classes (inequalities that hold for all Riemannian metrics on S). In the second part the idea of transmitting cut and paste instructions along a homotopy is combined with a result of Hass and Scott [3] to give a new proof of the minimal intersection property of [1] and to find some absolute inequalities involving minimal configurations.

1 Minimal configurations.

Two collections of immersed curves in S have *the same configuration* if there is an ambient isotopy that moves the image of one to the other. The curves in a configuration *intersect minimally* or have *minimal intersections* if they minimize the number of intersections and self-intersections among all transverse and self-transverse curves in their free homotopy classes. Following [1] and [2], the intersections and self-intersections are counted 'in the source', by counting how many curves one crosses when following a curve all the way around (so multiple intersections are counted with multiplicity and all the curves in figure 1 have 6 self-intersections).

As with all geodesics in a surface, shortest geodesics are transverse and self-transverse, unless they factor through coverings of other geodesics. Shortest geodesics may not be unique, and they may not be in general position as they may have points of multiple intersections.

The results of [1] and [2] can be summarized as follows:

(a) Shortest geodesics intersect minimally, unless they are coverings of other shortest geodesics.

(b) If α is an orientation-preserving curve in S, the shortest geodesics representing powers of α always cover a shortest geodesic representing α, but if α

is orientation-reversing and there are 2 different shortest geodesics representing α, then the shortest geodesics representing α^2 and the odd powers of α do not cover other shortest geodesics.

According to these results, the image of a collection of shortest geodesics in S is a configuration of essential curves that intersect transversely and minimally and do not represent proper powers of any orientation-preserving class. A finite configuration of essential curves in S with these properties will be called a *minimal configuration* in S.

Theorem 1.1 *Any minimal configuration of curves in a closed surface S is a configuration of shortest geodesics for some Riemannian metric g on S. g can be chosen so that the length of each curve in the configuration is equal to the total number of intersections along it.*

Remark The curves need not be in general position, some may be homotopic or represent proper powers of an orientation-reversing class. The curves with no intersections will have length 1.

Lemma 1.2 *If $c_1, c_2, ..., c_n$ is a collection of curves immersed transversely in S, and N is a regular neighborhood of $\cup c_i$, then there is a Riemannian metric g on S such that:*
(a) *Each c_i is a geodesic in S and a shortest geodesic in N.*
(b) *All essential curves in S that don't lie entirely in N are longer than every c_i.*
(c) *The lengths of the arcs of the configuration (the components of $\bigcup c_i -$ intersections) can be chosen to be any positive numbers.*

Proof The idea is to make the surface look like a landscape with the curves lying in the bottom of deep and narrow canyons and surrounded by large mountains. Start with any Riemannian metric g_S on S and regular neighborhoods $N^- \subset N \subset N^+$ of $\bigcup c_i$. Since there is a positive lower bound for the lengths of all essential curves in $S - N^-$ and all arcs running from $S - N$ to N^-, then by multiplying g_S by a constant k we can make that lower bound larger than the desired lengths of the curves (this makes the canyons deep and the mountains large). Now we want to modify the metric inside N^+. $\bigcup c_i$ is a union of arcs that meet at the multiple points, so N^+ is the union of (topological) rectangles and polygons (around the multiple points) as in figure 2a. Put a flat metric g_N on N^+ to make these rectangles and polygons Euclidean, so each c_i is a geodesic and no homotopy of c_i within N^+ reduces its length. Since the lengths of the rectangles can be chosen independently of each other, we can choose the length of each arc of the configuration to be any positive number, and the diameters of the polygons can be taken to be smaller than any prescribed number d (this makes the canyons long and narrow).

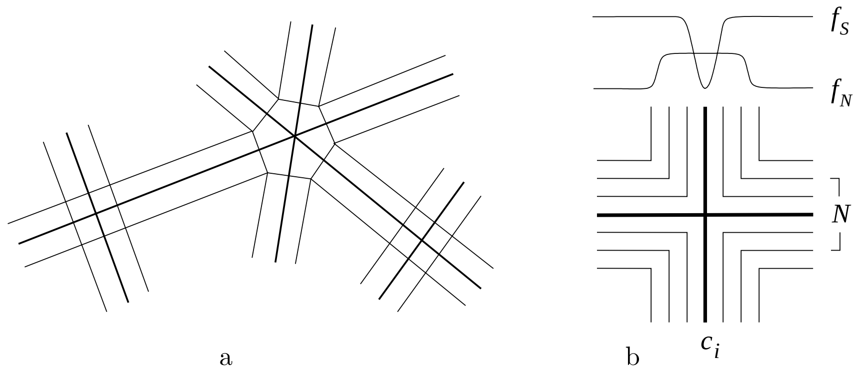

Figure 2

Let $g = f_S g_S + f_N g_N$ (in terms of the first fundamental forms), where f_S and f_N are smooth scalar functions such that:

$f_N = 1$ on N, $f_N > 0$ on $N^+ - N$ and $f_N = 0$ on $S - N^+$.

$f_S = 0$ on $\bigcup c_i$, $f_S > 0$ on $N^- - \bigcup c_i$ and $f_S = k$ on $S - N^-$.

See figure 2b. As $g \geq k g_S$ on $S - N^-$, with the metric g any essential curve in $S - N^-$ and any arc that crosses from N^- to $S - N$ is longer than c_i, so any essential curve which doesn't lie in N is longer than c_i. As c_i was a shortest geodesics in N with the metric g_N , and $g = g_N$ on $\bigcup c_i$ but $g > g_N$ on $N - \bigcup c_i$, the metric g makes every c_i a geodesic in S and shortest geodesic in N. □

Each essential curve c immersed in N is freely homotopic in N to a *polygonal* curve p made of arcs of the configuration (maybe repeated) and we may assume that p is reduced in the sense that no arc is followed immediately by the same arc in the opposite direction.

Lemma 1.3 *The metric g can be defined so that if a (reduced) polygonal curve p is longer than c_i, then all the curves homotopic to p in N are also longer than c_i.*

Proof This is accomplished by choosing d small (narrow canyons). If p has corners (i.e., if p is not one of the c_i's) then its length can be reduced by rounding the corners, but no homotopy within N can reduce its length by more than d multiplied by the number of corners of p (this is clear for the metric g_N, and $g = g_N$ along p and $g \geq g_N$ elsewhere in N), and the number of corners in p is bounded above by a linear function of its length. So if c is homotopic to p in N then $\frac{length(c)}{length(p)} > 1 - ld$, where l is the coefficient of the linear function, so by taking d small enough we can make this ratio as close to 1 as we want. But the set of lengths of the polygonal curves in N is discrete (because it is contained in the set of positive linear combinations of the lengths of the arcs), so $\frac{length(c_i)}{length(p)} < m < 1$ for all c_i 's and all longer p's. So by taking d small we can make $\frac{length(c)}{length(p)} > \frac{length(c_i)}{length(p)}$ for every p longer than c_i. □

The previous lemmas have no minimal intersection hypothesis: the metric g makes each c_i a geodesic in S, but not necessarily a shortest geodesic, because nonhomotopic curves in N may be homotopic in S. The c_i's are shortest geodesics for some metric g on S if and only if the lengths of the arcs of the configuration can be chosen so that all homotopic c_i's have the same length and all polygonal curves homotopic to c_i are longer than c_i.

Now let $c_1, c_2, ..., c_n$ be a collection of curves with minimal intersection and self-intersection in S. In order to choose the lengths of the arcs of the configuration, take a collection of *measuring curves* $\mu_1, \mu_2, ..., \mu_m$ in general position with respect to the c_i's, assign to each μ_j a positive *width* w_j, and define the length of each arc of the configuration as the sum of the widths of the curves μ_j that meet the arc. As the arcs of the configuration must have positive length, we need a collection $\{\mu_j\}$ whose union meets all the arcs.

Let's say that a measuring collection for $\{c_i\}$ is *good* if it intersects each c_i minimally but does not intersect any polygonal curve homotopic to some c_i minimally.

Lemma 1.4 *If $\{\mu_j\}$ is a good measuring collection, then for any choice of widths the assigned lengths make the c_i's shortest geodesics for a Riemannian metric in S.*

Proof As the μ_j's intersect c_i minimally, if a polygonal curve p is homotopic to c_i, then each μ_j must intersect p at least as many times as it intersects c_i, so p is at least as long as c_i, and it is longer than c_i if and only if the total number of intersections of the μ_j's with p is larger, i.e., if some μ_j does not intersect p minimally. This is clearly independent of the choice of widths. Now apply lemmas 1.2 and 1.3. □

Remark Notice that if a good measuring collection is extended in any way (by adding curves that intersect the c_i's minimally) then the resulting measuring collection is good.

Construction of good measuring collections

Let $\{c_i\}$ be a configuration of essential curves in a surface S. If $\chi(S) \leq 0$, the universal covering of S is a plane $\widetilde{S}$, and the cyclic coverings S^α of S corresponding to the subgroups generated by elements α of $\pi_1(S)$ are annuli or Moebius bands (depending on whether α is orientation preserving or orientation-reversing). So the preimage of $\{c_i\}$ in $\widetilde{S}$ is an infinite configuration of topological lines, while the preimage of $\{c_i\}$ in S^α is a configuration of lines and curves (the liftings of the c_i's representing powers of α, if any). The curves in S^α will be denoted by c_i^α and the lines in S^α or $\tilde{S}$ by $\widetilde{c}_i$.

According to [1] and [2], $\{c_i\}$ is a minimal configuration in S if and only if for each S^α the curves c_i^α intersect minimally and intersect the lines $\widetilde{c}_i$ minimally, that is:

(a) If α is orientation-preserving then the curves representing α in S^α are embedded and disjoint, and intersect each line in at most 1 point.

(b) If α is orientation-reversing then the curves representing α^r, r odd, have $r-1$ self-intersections and intersect the curves representing α^s (s odd, $s \geqslant r$) in r points. The curves representing α^2 are embedded and disjoint from all the other curves representing powers of α. A curve representing α^r ($r = 2$ or odd) intersects a line that crosses S^α in r points.

Case 1 *All c_i's are primitive and orientation-preserving.*

A natural candidate for a good measuring collection consists of a pair of 'parallel' curves μ_{i+} and μ_{i-} for each c_i, one to the right and one to the left of c_i and sufficiently close so that the immersed annulus determined by μ_{i+} and μ_{i-} intersects the curves of the configuration along arcs that cross the annulus, and the only multiple points of the configuration inside the annulus are the ones along c_i. So c_i and μ_{i+} intersect each c_j the same number of times (the arcs of intersection between the curves and the annulus give a one to one correspondence between the intersections along c_i and the intersections along μ_{i+}) and so μ_{i+} intersects each c_j minimally. By construction $\{\mu_{i+}, \mu_{i-}\}$ meets all the arcs of the configuration. Notice that taking all the widths equal to $\frac{1}{2}$ makes the length of each c_i equal to the number of intersections of the configuration along c_i (counted with multiplicity).

Define the *distance* between two lines in the configuration $\{\widetilde{c}_i\}$ in $\widetilde{S}$ as the minimum number of complementary regions that one has to cross to go from one line to the other (so the distance is 0 iff the lines meet). We will say that 2 -not necessarily different- curves c_i and c_j in S are *close neighbors* if two of their preimages $\widetilde{c}_i$ and $\widetilde{c}_j$ are at distance 1 in $\widetilde{S}$.

Claim In case 1, $\{\mu_{j+}, \mu_{j-}\}$ is a good measuring collection for $\{c_i\}$ if and only if every c_i has close neighbors on both sides.

Proof Observe that a polygonal curve p homotopic to c_i intersects $\{\mu_{i+}, \mu_{i-}\}$ minimally if and only if in the corresponding covering S^α, the curves p^α and c_i^α intersect the same measuring curves and lines (μ_{i+}^α, μ_{i-}^α, $\widetilde{\mu}_{j+}$, $\widetilde{\mu}_{j-}$) and do so the same number of times (once in this case). So p^α cannot cross or touch c_i^α or any other curve c_j^α (because p^α would intersect μ_{i+}^α, μ_{i-}^α, μ_{j+}^α or μ_{j-}^α), and the annulus bounded by c_i^α and p^α must intersect the lines $\widetilde{c}_j$ along arcs that cross the annulus (if a line $\widetilde{c}_j$ touches this annulus at one point or intersects it along an arc that starts and ends in p^α, then p^α intersects one of the lines $\widetilde{\mu}_{j+}$ or $\widetilde{\mu}_{j-}$ twice). In particular p^α must be made exclusively of arcs of lines $\widetilde{c}_j$ that cross c_i^α.

If all the c_i's representing α have close neighbors on both sides, then the curves c_i^α in S^α are just one complementary region away from other c_j^α's or from lines $\widetilde{c}_j$ that don't meet c_i^α. So any polygonal curve p^α must cross or at least touch one of these curves or lines, and so p cannot have minimal intersections with $\{\mu_{j+}, \mu_{j-}\}$.

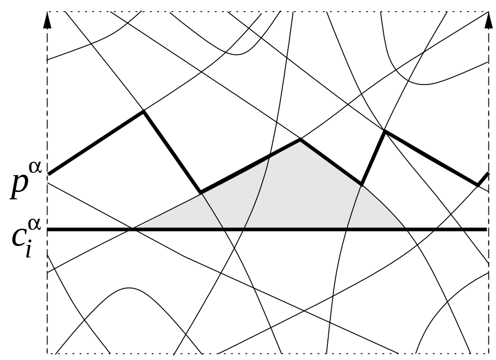

Figure 3

Now suppose that some c_i representing α doesn't have close neighbors on one side. Then for each complementary region R on one side of c_i^α, the lines adjacent to R must intersect c_i^α. These lines determine triangles with base in c_i^α that contain R, and the widest of these triangles (the one with maximal base in c_i^α) is crossed by the lines $\widetilde{c}_j$ along arcs that meet the base of the triangle. See figure 3. The union of these wide triangles on one side of c_i^α is an annulus whose boundaries are $\widetilde{c}_i$ and a polygonal curve p^α, and the lines $\widetilde{c}_j$ can intersect this annulus only along arcs that cross the annulus, so p^α projects to a polygonal curve p homotopic to c_i that intersects $\{\mu_{i+}, \mu_{i-}\}$ minimally, and p is the nearest polygonal curve with this property. □

In the configuration of figure 4a the curve c_1 has close neighbors on both sides but the curve c_2 doesn't. So a metric g that makes each arc of the configuration of length 1 can make c_1 (but not c_2) a shortest geodesic in the surface.

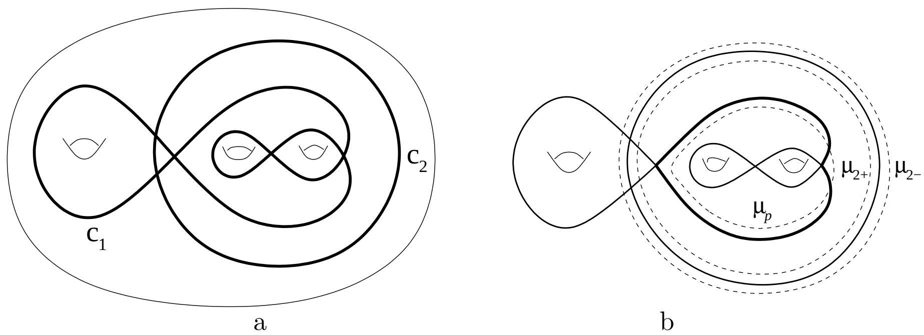

Figure 4

Now we want to extend $\{\mu_{i+}, \mu_{i-}\}$ to a good measuring collection in the case that some c_i's don't have close neighbors. The idea is given in figure 4b: if there is a polygonal curve p homotopic to c_i that intersects $\{\mu_{i+}, \mu_{i-}\}$ minimally, take a measuring curve μ_p that runs "quasiparallel" to p crossing each edge of p once, so its lifting to S^α looks like in figure 5, making it sufficiently close so that each c_j intersects the singular annulus determined by c_i and μ_p along arcs that cross the annulus.

To see that μ_p intersects each c_j minimally, it is enough to show that μ_p and c_i intersect each c_j the same number of times. This happens because in S^α each line $\widetilde{c}_j$ intersects the annulus determined by c_i^α and μ_p^α along arcs that cross the annulus (an arc of intersection of $\widetilde{c}_j$ with the annulus cannot start and end in μ_p^α, because then $\widetilde{c}_j$ would cross p^α twice or it would touch it at one point). So by adding a measuring curve μ_p for each short polygonal curve p we can extend $\{\mu_{i+}, \mu_{i-}\}$ to a good measuring collection. As only finitely many polygonal curves homotopic to c_i can intersect $\{\mu_{i+}, \mu_{i-}\}$ minimally (because

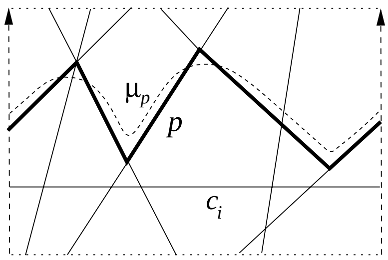

Figure 5

the number of arcs in such polygonals is bounded above by the number of intersections along c_i), we are done.

Case 2 *All orientation-reversing c_i's are primitive and no two of them are homotopic*

Choose the measuring curves corresponding to the orientation-preserving curves as in case 1. The orientation-reversing c_i's are one sided, so instead of two parallel curves μ_{i+} and μ_{i-} there is a single curve $\mu_{i\pm}$ homotopic to c_i^2 that runs on "both sides" of c_i. To see that $\mu_{i\pm}$ intersects each c_j minimally, look at the covering S^α of S corresponding to the class α represented by c_i. The only closed curve in S^α is c_i^α, which by construction does not meet $\mu_{i\pm}^\alpha$, and the lines $\widetilde{c}_j$ that cross S^α intersect c_i^α at a single point, so they must intersect $\mu_{i\pm}^\alpha$ at exactly two points. Now if p is any polygonal curve homotopic to c_i, then (as c_i is one sided) p^α must cross c_i^α, so p^α must intersect $\mu_{i\pm}^\alpha$, and so $\mu_{i\pm}$ doesn't intersect p minimally. Therefore this measuring collection is already good.

Observe that if an orientation-reversing c_i is nonprimitive, or is homotopic to another c_j, then $\mu_{i\pm}$ doesn't have minimal intersection with c_i (or c_j), and therefore $\mu_{i\pm}$ cannot be used as a measuring curve.

Case 3 *S is a projective plane*

All c_i's are homotopic to the unique nontrivial element of $\pi_1(S)$, so they are embedded and intersect each other in 1 point. For each c_i take a collection of measuring curves μ_{ix} each made of an arc that runs parallel to c_i all the way around and a small arc that crosses c_i at one point, as in figure 6a.

Make μ_{ix} sufficiently close to c_i so that the other c_i's intersect the singular strip determined by c_i and μ_{ix} along arcs that cross it from c_i to μ_{ix} (so μ_{ix} will intersect each c_j once) and the only multiple points of the configuration inside the band are the ones along c_i. Take two μ_{ix} for each arc of c_i, one crossing the arc in each direction. Now if p is any polygonal curve homotopic to c_i and p contains arcs of c_j, then some μ_{jx} crosses a corner of p twice (see figure 6b), so p doesn't have minimal intersection with that μ_{jx}.

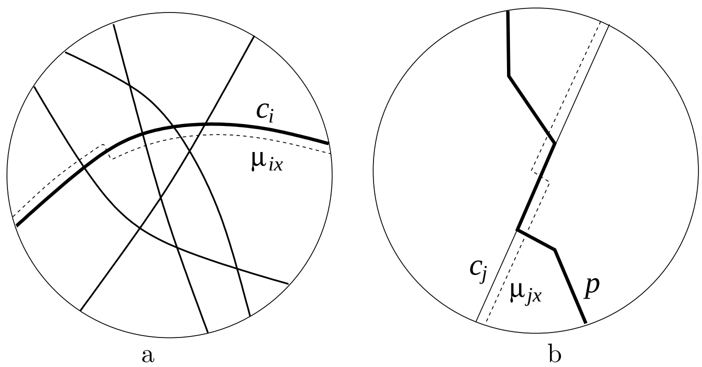

Figure 6

Case 4 *All orientation-reversing c_i's are primitive, but some are homotopic*

Choose the measuring curves for the orientation-preserving c_i's as in case 1, and those for the orientation-reversing c_i's that are not homotopic to other c_j's as in case 2.

Now consider an orientation-reversing class α represented by 2 or more c_i's. These c_i's lift to curves c_i^α in the Moebius band S^α that are embedded and intersect each other in 1 point. For each of these c_i's take a collection of measuring curves μ_{ix} as in case 3, each one made of an arc that runs parallel to c_i all the way around and a small arc that crosses c_i at one point, so μ_{ix} lifts to a curve μ_{ix}^α in S^α that intersects c_i^α in exactly one point as in figure 7a. Take again one μ_{ix} crossing each arc of c_i in each direction. As μ_{ix} and c_i intersect each $c_j \neq c_i$ the same number of times, then μ_{ix} intersects each $c_j \neq c_i$ minimally, and as μ_{ix} intersects c_i one more time than c_i intersects itself, then μ_{ix} also intersects c_i minimally.

Observe that a polygonal curve p that intersects these μ_{ix}'s minimally must be made exclusively of arcs of orientation-preserving c_j's, because if p contains an arc of some orientation-reversing c_j then one of the μ_{jx}'s crosses a corner of p twice. And one can show as in case 1 that such p has minimal intersection with the curves μ_{i+}, μ_{i-} and μ_{ix} if and only if p^α intersects c_i^α and every curve c_j^α homotopic to c_i^α in exactly one point, and each line $\widetilde{c}_j$ that intersects the singular annulus determined by c_i^α and p^α does so along one arc that crosses the annulus from p^α to c_i^α.

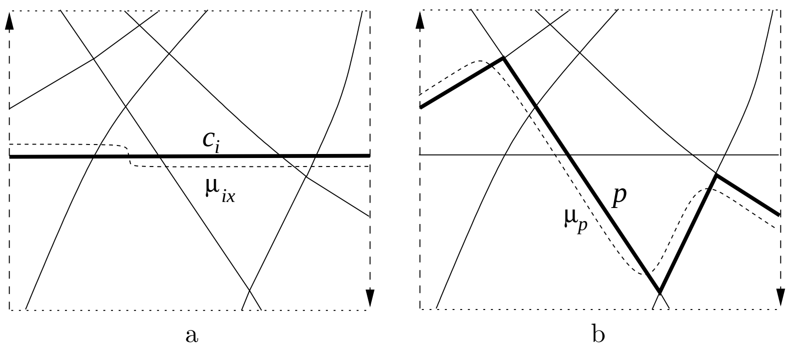

Figure 7

If there is a short polygonal curve p homotopic to α, take a measuring curve μ_p that runs "quasiparallel" to p crossing each edge of p once, so μ_p lifts to a curve μ_p^α in S^α that looks like in figure 7b. To see that μ_p intersects each c_j minimally, observe that μ_p^α intersects each curve c_j^α once (otherwise p^α would intersect c_j^α more than once) and that each $\widetilde{c}_j$ intersects the singular annulus determined by c_i^α and μ_p^α along arcs that cross it from c_i^α to μ_p^α (an arc of intersection cannot start and end in μ_p, because then $\widetilde{c}_j$ would intersect the singular annulus determined by c_i^α and p^α in an arc that starts and ends in p^α).

By construction the number of intersections between μ_p^α and p^α is equal to the number of corners of p, so μ_p intersects p minimally only when p has one corner. To deal with these short polygonal curves with only one corner, we need an extra measuring curve $\mu_{\alpha\pm}$ whose lifting to S^α runs parallel to the boundary of the region $V\alpha$ determined by all the curves c_i^α, as in figure 8a, so $\mu_{\alpha\pm}$ is homotopic to α^2. $\mu_{\alpha\pm}$ intersects each c_j minimally because the curves c_i^α are contained in the Moebius band bounded by $\mu_{\alpha\pm}$, and if a line $\widetilde{c}_j$ intersects this Moebius band along a nonessential arc then $\widetilde{c}_j$ intersects some c_i^α in two points.

We claim that if p is a short polygonal curve with one corner then p^α cannot be contained in $V\alpha$, so p^α intersects $\mu_{\alpha\pm}$ and so p doesn't intersect $\mu_{\alpha\pm}$ minimally. If p^α were contained in $V\alpha$ then its corner would be in the region determined by two curves c_i^α and c_j^α. As p^α is made of an arc of a line $\widetilde{c}_k$ that starts and ends at the corner, $\widetilde{c}_k$ would have to cross c_i^α or c_j^α twice (see figure 8b) contradicting the fact that c_k intersects c_i and c_j minimally.

Case 5 *Some orientation-reversing c_i's are nonprimitive*

Let $1 \leq r_1 < r_2 < ... < r_n$ be the odd powers of a primitive orientation-reversing class α represented by c_i's in the configuration. Each of these c_i's

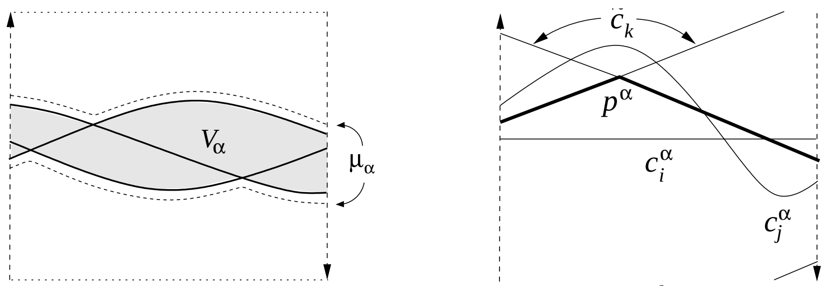

Figure 8

lifts to an immersed curve c_i^α in the Moebius band S^α. For each of these c_i's take a collection of measuring curves μ_{ix} as in case 4, each made of an arc that runs parallel to c_i all the way around and a small arc that crosses c_i at one point. So each μ_{ix} intersects every c_j minimally and every polygonal curve p that intersects these μ_{ix}'s minimally is made of arcs of orientation-preserving c_j's For each polygonal curve p homotopic to c_i that intersects these μ_{ix}'s minimally, take a measuring curve μ_p that runs quasiparallel to p crossing each edge of p once so, as in case 4, μ_p intersects every c_j minimally, but μ_p intersects p minimally only when p has one corner.

To deal with these short polygonal curves with one corner representing α^{r_k}, we need to add an extra measuring curve $\mu_{\alpha\pm}$ homotopic to α^2 and measuring curves $\mu_{\alpha k}$ homotopic to $\alpha^{r_{k+1}}$ for each $k < n$. One can show as in case 4 that the polygonal curves with one corner representing α^{r_k} cannot be contained in the region $V\alpha^{r_k}$ of S^α determined by the images of all the c_i^α's representing α^{r_k}. The minimal intersection of the curves in S^α implies that all the curves representing α^{r_k} are contained in the region determined by each curve representing $\alpha^{r_{k+1}}$, so $V\alpha^{r_k} \subset V\alpha^{r_{k+1}}$, each curve representing $\alpha^{r_{k+1}}$ intersects $V\alpha^{r_k}$ along one arc, and each line that intersects $V\alpha^{r_k}$ does so along one essential arc.

Let $\mu_{\alpha\pm}$ be a curve whose lifting to S^α runs parallel to the boundary of $V\alpha^{r_n}$, so $\mu_{\alpha\pm}$ is homotopic to α^2. Then $\mu_{\alpha\pm}$ intersects every c_j minimally, but any polygonal curve representing an odd power of α that intersects $\mu_{\alpha\pm}$ minimally must be contained in $V\alpha^{r_n}$. Now for each $k < n$, choose a curve c_i^α representing $\alpha^{r_{k+1}}$ which is closest to $V\alpha^{r_k}$ in the sense that the region determined by its image does not contain any other c_j^α representing $\alpha^{r_{k+1}}$. Let $\mu_{\alpha k}$ be a curve whose lifting to S^α runs parallel to the arc $c_i^\alpha \cap V\alpha_{r_k}$ and then runs around the boundary of $V\alpha_{r_k}$ enough times to complete a curve homotopic to $\alpha^{r_{k+1}}$.

Figure 9a shows a lifting of $\mu_{\alpha k}$ to $S^{\alpha^{r_{k+1}}}$. To prove that $\mu_{\alpha k}$ intersects each c_j minimally, it is enough to show that in the covering $S^{\alpha^{r_{k+1}}}$ the preimages of c_j intersect the region determined by the liftings of c_i and $\mu_{\alpha k}$ along arcs that cross that region. An arc of intersection a that didn't cross that region would look as in figure 9b, but this arc cannot belong to a line $\widetilde{c}_j$ because then $\widetilde{c}_j$ would intersect $V\alpha^{r_k}$ in at least two arcs, and it cannot belong to a curve c_j^α representing α^{r_k} or a smaller power of α because these curves are contained in $V\alpha^{r_k}$. So the arc a must belong to a curve c_j^α representing some larger power of α, and so $c_j^\alpha = a \cup a'$, where a' is an arc in $V\alpha^{r_k}$. So c_j^α lies in the region determined by c_i^α, but by the choice of c_i^α no curve representing $V\alpha^{r_{k+1}}$ or a larger power of α can be contained in this region.

Now if p is a polygonal curve with one corner representing α^{r_k} then its lifting to S^α is not contained in $V\alpha^{r_k}$, so it is not contained in the region determined by $\mu_{\alpha k}$, so p does not intersect $\mu_{\alpha k}$ minimally.

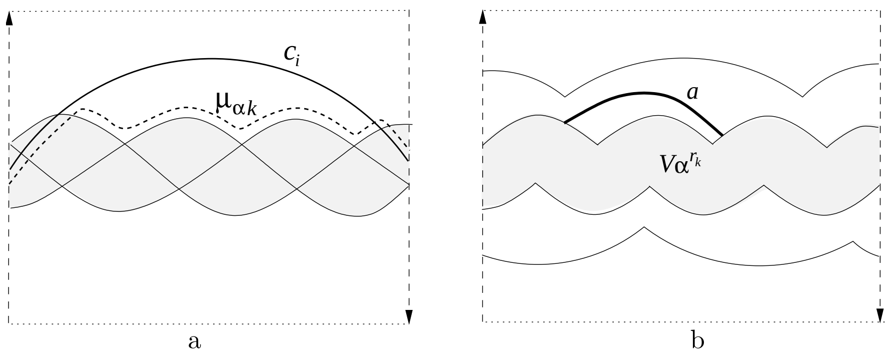

Figure 9

Choice of widths

The measuring collection for $\{c_i\}$ constructed above is made of curves homotopic to some c_i (μ_{i+}, μ_{i-}, $\mu_{i\times}$, μ_p and $\mu_{\alpha k}$) or the square of some primitive orientation-reversing class α ($\mu_{i\pm}$ and $\mu_{\alpha\pm}$). The choice of widths to prove the second part of the theorem is not obvious because the minimum number of self-intersections of an orientation-reversing c_i differs from the minimum number of intersections between c_i and a homotopic curve by 1. The condition that the length of each c_i in the configuration must be equal to the number of intersections along it gives a system of linear equations on the widths of the measuring curves that has a unique solution for the sums of widths of the measuring curves in each homotopy class:

(a) Make the sum of the widths of the measuring curves (μ_{i+}, μ_{i-} and μ_p's) in each orientation-preserving class equal to the number of c_i's in that class.

(b) If α is a primitive orientation-reversing class, and $1 \leq r_1 < r_2 < ...$ are the odd powers of α represented by some c_i 's, make the sum of the widths of the measuring curves (μ_{ix}, μ_p and $\mu_{\alpha k-1}$'s) representing α^{r_k}, $k > 1$, equal to the number of c_i's representing that class, but for the measuring curves representing α^{r_1} make the sum of their widths $\frac{1}{r_1}$ units less than the number of c_i's representing that class. Finally, make the width of each measuring curve ($\mu_{i\pm}$ and $\mu_{\alpha\pm}$) representing α^2 equal to $\frac{1}{2}$.

A problem arises when $r_1 = 1$ and only 1 curve c_i represents α^{r_1}, because then the sum of the widths of the measuring curves homotopic to α is 0, which means that these measuring curves cannot be used, and the rest of the measuring collection may not be good. This can be arranged by replacing the measuring curves representing α by suitable curves representing α^{r_2} as follows:

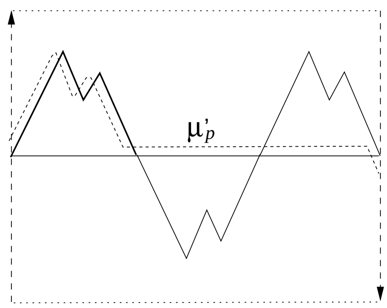

Figure 10

Trade each curve μ_{ix} made of an arc that goes once around c_i and a small arc that crosses c_i at one point, for a curve μ'_{ix} made of an arc that goes r_2 times around c_i and the small arc. And trade each μ_p representing α for a curve μ'_p obtained by replacing the small arc of μ_p that crosses c_i by an arc that goes $r_2 - 1$ times around c_i so it now represents α^{r_2} (figure 10 shows the lifting of μ'_p to $S^{\alpha^{r_2}}$). It is not hard to see that μ'_{ix} and μ'_p intersect each c_j minimally, but intersect nonminimally all the polygonal curves that intersect μ_{ix} or μ_p nonminimally. This proves the second part of the theorem. □

Figure 11 shows the lengths of the arcs in the configuration in figure 4 resulting from making μ_{i+} and μ_{1-} of width $\frac{1}{2}$ and μ_{2-}, μ_{2+} and μ_p of width $\frac{1}{3}$.

Remark Theorem 1.1 clearly holds for nonclosed surfaces, provided that the curves don't meet the boundary. It also works for minimal configurations of

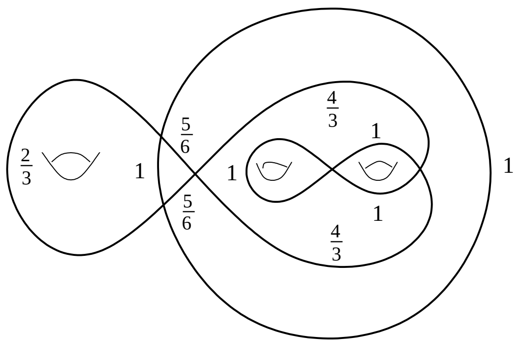

Figure 11

properly immersed curves and arcs in a surface with boundary, considering either minimal configurations with the endpoints of the arcs fixed or free to move along ∂S (one just needs to use measuring arcs analogous to the measuring curves).

Theorem 1.1 contrasts with the examples of Hass and Scott [4] of minimal configurations of primitive and nonhomotopic curves in a surface which are not configurations of geodesics for any metric of negative curvature on the surface. These configurations, however, can be realized by metrics of non-positive curvature.

Questions Which configurations of primitive curves in a surface are configurations of shortest geodesics for metrics of negative curvature? and for metrics of non positive curvature?

The second part of theorem 1.1 is only significant for configurations containing more than 1 curve. For configurations of 1 curve in general position one may ask if the lengths of all the arcs can be made equal (we know that the answer is yes if the curve is orientation-reversing, and no in general if the configuration is not in general position or contains more than one curve). One may also ask if every minimal configuration of curves in general position is contained in a configuration of shortest geodesics in which each arc has the same length. These questions are equivalent to the following:

Questions Do all minimal 1-curve configurations have close neighbors? Can every minimal configuration be extended to a configuration with close neighbors?

One may face strong restrictions when trying to extend a minimal configuration to contain other curves. For example, if one wants to extend the configuration

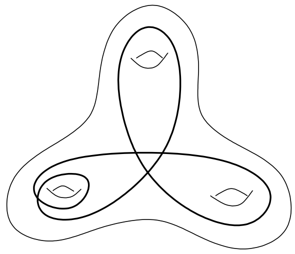

Figure 12

in figure 12 to one containing a curve in the homotopy class of figure 1, then the first curve must look as in figure 1b. Theorem 1.1 implies that some extension is always possible:

Corolary 1.5 *Every minimal configuration of curves in S can be extended to a minimal configuration containing curves in any given homotopy classes in S.*

Denote by $l_g(a)$ the minimum length in the free homotopy class of the curve a when S is given a Riemannian metric g. Denote by $a \bigcap b$ the minimum number of intersections between curves in the free homotopy classes of a and b, and by $a \bigcap a$ the minimal number of self-intersections in the homotopy class of a.

Corolary 1.6 *If $l_g(a) \leq k \cdot l_g(b)$ for every Riemannian metric g on S, then $a \bigcap c \leq k \cdot b \bigcap c$ for every curve c in S. In particular, $a \bigcap a \leq k \cdot a \bigcap b \leq k^2 \cdot b \bigcap b$.*

Proof Suppose that $a \bigcap c > k \cdot b \bigcap c$ for some curve c. We may assume that a, b and c are in general position and have minimal intersection and self-intersection. Apply the proof of theorem 1.1 to the configuration formed by a and b, but add to the resulting measuring collection a copy of the curve c with weight w. If

$$w > \frac{k \cdot b \bigcap b + (k-1) \cdot a \bigcap b - a \bigcap a}{a \bigcap c - k \cdot b \bigcap c}$$

then for the resulting metric g we have

$$l_g(a) = a \bigcap a + a \bigcap b + w \cdot a \bigcap c > k \cdot \left(b \bigcap a + b \bigcap b + w \cdot b \bigcap c \right) = k \cdot l_g(b)$$

contrary to the hypothesis that $l_g(a) \leq k \cdot l_g(b)$. □

Corollary 1.6 clearly holds if the curves a , b and c are replaced by any finite families of curves or arcs.

2 Cutting and pasting.

Let $\{a_i\}$ be a configuration of curves with transverse intersections in S. A *cut and paste* on $\{a_i\}$ is done by cutting these curves at some of their intersection points and glueing the resulting arcs in a different order to obtain a new collection of curves $\{b_j\}$. These curves have some 'corners' that can be rounded so the total number of intersections and the total length of the original configuration are reduced.

Lemma 2.1 *If a collection of curves $\{a_i\}$ in S can be cut and pasted to obtain the collection $\{b_j\}$, and $\{a_i\}$ can be homotoped to a collection $\{a'_i\}$ without removing any intersection points in the process, then $\{a'_i\}$ can be cut and pasted to obtain a collection homotopic to (the nontrivial) $\{b_j\}$.*

Proof We want to show that the instructions for cutting and pasting $\{a_i\}$ to get $\{b_j\}$ can be transmitted along the homotopy from $\{a_i\}$ to $\{a'_i\}$ so that the final result is homotopic to $\{b_j\}$. This is not obvious even though the intersection points of $\{a_i\}$ can be traced along the homotopy (they don't disappear), because the result of doing the "same" cut and paste before or after the homotopy may be different, as shown in figure 13.

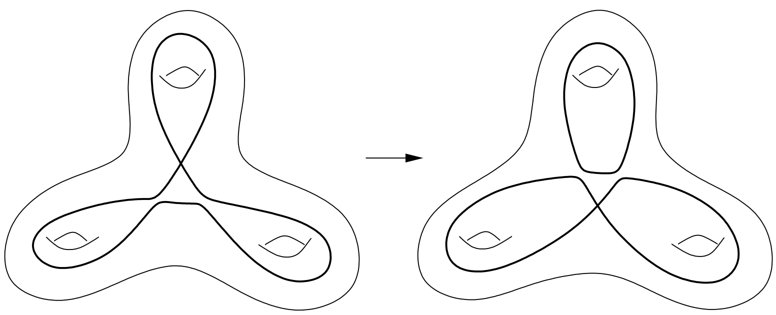

Figure 13

Any homotopy that doesn't remove intersection points can be done using 3 types of local moves in the configuration. The first two moves, adding a small loop and creating a small bigon, do not change the homotopy class of the resulting curves. Nevertheless, when doing these moves one can add cut and paste instructions at the new intersections to avoid increasing the number of intersections of the resulting curves (see figure 14a,b).

In the third move a local configuration of n arcs that intersect each other at different points collapses into one where all the arcs meet at a single point,

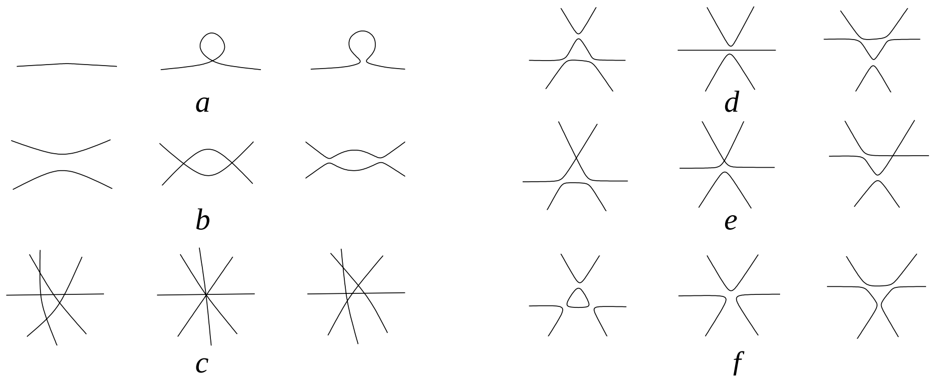

Figure 14

or viceversa: a configuration with a multiple intersection opens up (see figure 14c). Observe that to transmit some cut and paste instructions during these local moves one only needs that the endpoints of the arcs that were connected by the original cut and paste instructions get connected by the new instructions (any curve contained in the local configuration is trivial). As the endpoints of the arcs that meet at a single point can be connected at will by cutting and pasting at that point, then all cut and paste instructions can be transmitted when a local configuration collapses into a multiple intersection.

So the problem is to transmit the cut and paste instructions when a multiple intersection opens up. In the case $n = 3$ one can see how this can be done directly (figure 14d-f shows some cases). Observe that the new instructions may not be unique, but they can always be chosen to avoid creating new curves and to avoid increasing the number of intersections of the resulting curves. In the case $n > 3$, modify the homotopy so the multiple intersection opens up one arc at a time. If an arc a moves away from the multiple intersection point and the cut and paste instructions don't change, then the only connections that are affected are those involving the endpoints of a, which are connected to the endpoints of at most 2 other arcs of the local configuration. But we can change the cut and paste instructions at the intersections of these 3 arcs as in the case $n = 3$ to get the right connections for the endpoints of a, and then change the cutting and pasting instructions at the multiple intersection point as needed to get the right connections between all the other endpoints. Now repeat the argument until the multiple intersection opens up completely. □

In [3] Hass and Scott defined a 'curve flow' that takes any configuration of primitive curves in a surface to a configuration of shortest geodesics by a homotopy that does not increase the number of intersections at any moment. This result and the previous lemma imply the following version of the theorem of Freedman, Hass and Scott:

Proposition 2.2 *Any finite family of primitive, orientation-preserving curves in S can be cut and pasted to obtain a freely homotopic family of curves with minimal intersections and self-intersections.*

Proof By [3] there is a homotopy that takes the family $\{a_i\}$ to some minimal intersection family $\{a_i'\}$ without increasing the number of intersections, so running the homotopy backwards we get a homotopy that takes $\{a_i'\}$ to $\{a_i\}$ without removing any intersection points. Now lemma 2.1 shows how to transmit the "don't cut anything" instructions in $\{a_i'\}$ to cutting and pasting instructions in $\{a_i\}$ without increasing the number of intersections of the resulting curves. □

Figure 15 shows a nonminimal configuration of 2 curves and a cut and paste that transforms it into a minimal configuration.

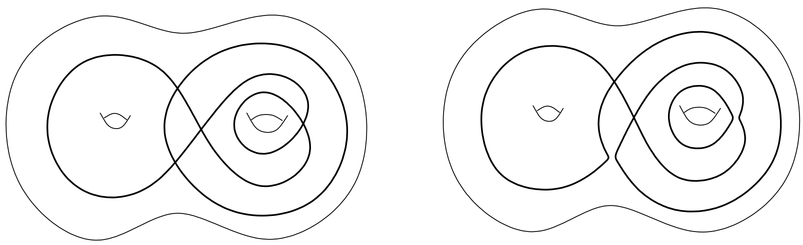

Figure 15

Corolary 2.3 *If a collection $\{a_i\}$ of curves with minimal intersection and self-intersection in S can be cut and pasted to obtain the collection $\{b_j\}$, then $l_g(\{a_i\}) > l_g(\{b_j\})$ for every Riemannian metric g on S.*

Proof Observe that the hypothesis that $\{a_i\}$ has minimal intersections is essential. If g is a Riemannian metric on S and $\{a_i'\}$ is a collection of shortest geodesics (for the metric g) homotopic to $\{a_i\}$, then by [3] there is a homotopy from $\{a_i\}$ to $\{a_i'\}$ that does not increase the number of intersections, so as $\{a_i\}$ already had minimal intersections the number of intersections must remain constant. So by lemma 2.1 the cutting and pasting instructions to get $\{b_j\}$ from $\{a_i\}$ can be transmitted to get a homotopic collection $\left\{b_j'\right\}$ from $\{a_i'\}$, so $l_g(\{a_i\}) = l_g(\{a_i'\}) > l_g(\left\{b_j'\right\}) = l_g(\{b_j\})$. □

Example The converse to corollary 2.3 is not true. Figure 16 shows 2 curves a and b on a surface such that a cannot be cut and pasted to obtain a curve homotopic to b but one can show that $l_g(a) > l_g(b)$ for every Riemannian metric g on S (so $a \bigcap c \geq b \bigcap c$ for every curve c).

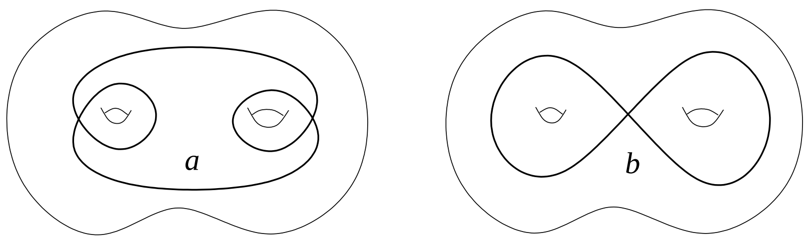

Figure 16

Question If $a \bigcap c \leq b \bigcap c$ for every curve c in S, is it true that $l_g(a) \leq l_g(b)$ for every Riemannian metric g on S?

References

[1] **M Freedman**, **J Hass**, **P Scott**, *Closed geodesics on surfaces*, Bull. London Math. Soc. 14 (1982) 385-391

[2] **J Hass**, **H Rubinstein**, *One-sided geodesics on surfaces*, Michigan Math. Journal 33 (1986) 155-168

[3] **J Hass**, **P Scott**, *Shortening curves on surfaces*, Topology, 33-1 (1994) 25-43

[4] **J Hass**, **P Scott**, *Configurations of curves and geodesics on surfaces*, Geometry and Topology Monographs 2: Proceedings of the Kirbyfest (1999) 201-213

[5] **M Shepard**, *Ph.D. Thesis*, U.C. Berkeley (1990)

Instituto de Matemáticas UNAM
Ciudad Universitaria, México D.F. 04510, México

Email: `max@math.unam.mx`

Received: 8 January 2001