ONE-SIDED CLOSED GEODESICS ON SURFACES

Joel Hass and J. H. Rubinstein

Let M^2 be a closed Riemannian 2-manifold and let α be a non-trivial element of $\pi_1(M)$. Among the set of all smooth loops in M which are freely homotopic to a curve representing α, there is a shortest member $f: S^1 \to M$, which is a smooth closed geodesic. Both f and the image of f will not be unique, in general. If α is orientation-preserving, then it was shown in [2] that f has the least possible number of self-intersections, unless f factors through a covering. In particular, if α is represented by an embedded loop, then f is either an embedding or a double cover of an embedded one-sided curve.

If α is orientation-reversing, then any loop which is freely homotopic to a curve representing α is one-sided. The features of one-sided loops differ significantly from two-sided curves, in particular those properties associated with coverings. Thus covers of one-sided shortest geodesics are not necessarily shortest, unlike the two-sided case.

A specific example of the difficulties encountered in the one-sided situation is seen by starting with a flat Möbius band M^2 and putting a bump in it as in Figure 1.

The bump is formed by multiplying the metric by a rotationally symmetric function on the shaded disk in Figure 1. A large enough bump will force the shortest geodesic representing a generator α of $\pi_1(M^2)$ to go around the bump. It is now clear that a shortest geodesic representing α^2 will not double cover a shortest loop representing α. This contrasts with Lemma 1.3 of [2]. Note that there are at least two distinct shortest geodesics representing α, by the symmetry of the construction. One goes above and one below the bump.

Nonetheless, we will show in this paper that shortest one-sided geodesics still minimize the number of double points in their intersection sets.

DEFINITION. We say that a loop $f: S^1 \to M$ represents $\alpha \in \pi_1(M,x)$ if f is freely homotopic to a loop at x in the homotopy class (rel x) of α. $f \sim \alpha$ will be used to denote that f represents α.

DEFINITION. $f: \mathbf{R} \to M$ is length-minimizing (or shortest) if f is shortest on any compact arc $I \subset \mathbf{R}$, in the homotopy class relative to ∂I of f restricted to I.

DEFINITION. Two maps $f: \mathbf{R} \to M$ and $g: \mathbf{R} \to M$ are homotopic by a homotopy with compact support if there is a homotopy $H: \mathbf{R} \times I \to M$ with $H(s,0) = f(s)$, $H(s,1) = g(s)$ and if there is a $K > 0$ such that $|s| > K$ implies $H(s,0) = H(s,t)$ for all $0 \le t \le 1$. Equivalently, the homotopy only moves a compact arc of $\mathbf{R}$.

Received August 27, 1984. Revision received February 15, 1985.

The second author would like to thank the N.S.F. for support while visiting the Institute for Advanced Study, Princeton.

Michigan Math. J. 33 (1986).

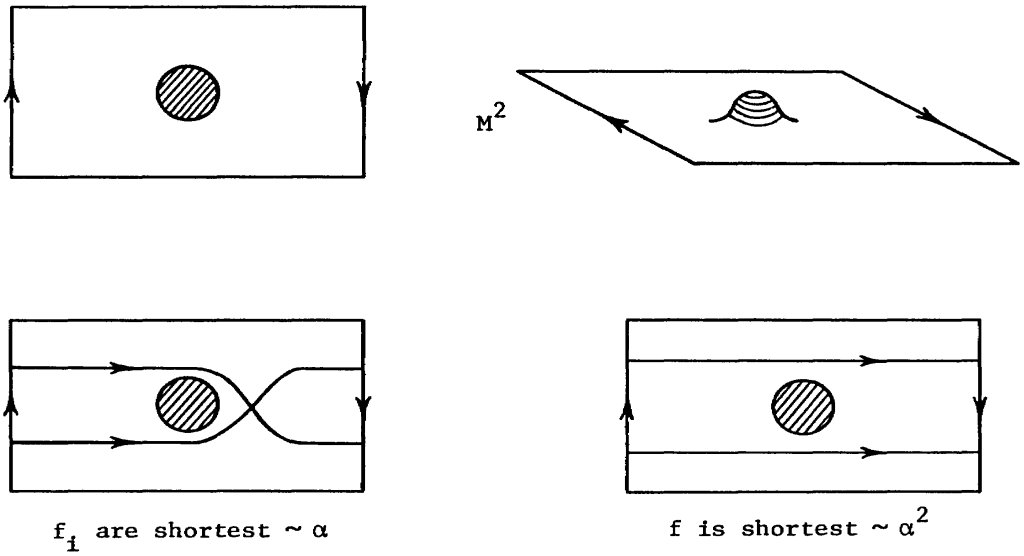

Figure 1

DEFINITION. Let $f, g: S^1 \to M$ be general position maps and let f (resp. g) represent α (resp. β) in $\pi_1(M)$. Define $D(f, g)$ as the number of double points $\# f(S^1) \cap g(S^1)$. Define $D(f)$ to be the number of double points of f, that is, $D(f) = \#\{x \in S^1 : f^{-1}(f(x))$ has 2 points$\}$. Define $D(\alpha, \beta)$ as $\inf\{D(f, g): f \sim \alpha$ and $g \sim \beta\}$ and $D(\alpha)$ to be $\inf\{D(f): f \sim \alpha\}$.

Note that shortest loops are geodesics and so are always transverse and self-transverse, unless they factor through coverings. A similar procedure to [2] could have been adopted to count intersections and self-intersections of multiplicity greater than 2, for shortest loops not in general position. However this is more complicated in the one-sided case, due to difficulties with coverings of a Möbius band (cf. §2). So we have restricted attention to general position maps for simplicity. The general case follows readily from this one.

Analogous results about intersections and self-intersections of least area incompressible two-sided surfaces in 3-manifolds are obtained in [3]. In an Appendix we consider the following question. Suppose M is a closed $\mathbf{R}P^2$-irreducible Riemannian 3-manifold and F is a closed surface not S^2 or $\mathbf{R}P^2$. Let $f: F \to M$ be a least area incompressible map which is homotopic to a one-sided embedding g. Is f an embedding? We show by the techniques of dealing with one-sided curves that this reduces to the case where f is a homotopy equivalence, that is, M is a twisted line bundle over $g(F)$. However we do not know how to complete this case.

We would like to thank D. McCullough for suggesting a simplified proof of Lemma 2.6, and also the referee for helpful comments.

1. Preliminaries. We recall here some results established in [2] and [4].

LEMMA 1.1. *Let $f: S^1 \to M$ be a shortest two-sided geodesic. Let $p_1: \tilde{S}^1 \to S^1$ and $p_2: \tilde{M} \to M$ be coverings and let $\tilde{f}: \tilde{S}^1 \to \tilde{M}$ be a lift of $f \cdot p_1$. Then $\tilde{f}$ is length-minimizing.*

Proof. If $\tilde{S}^1$ is a circle, this is established in Lemma 1.4 of [2]. If $\tilde{S}^1$ is a line, one can project the compact arc I homeomorphically into some large finite cover of S^1. This construction is carried out for the two-dimensional case in [4] and the argument is identical here. □

LEMMA 1.2. *Let f, g be two-sided shortest maps from $S^1 \to M$, where $f \sim \alpha$, $g \sim \beta$ and $\alpha, \beta \in \pi_1(M)$. If f and g are in general position then $D(f, g) = D(\alpha, \beta)$ and $D(f) = D(\alpha)$.*

NOTE. A similar result is true if f or g is a shortest arc or length-minimizing line. For length-minimizing lines, $D(\alpha)$ and $D(\alpha, \beta)$ are defined to be the infimum of double points in the compactly supported homotopy classes α and α, β (respectively). For arcs, the homotopies are relative to their boundaries. For the proof, see Theorems 3.2 and 3.3; see also §4 in [2].

We next state a Proposition which shows just how different the one-sided and two-sided cases are. If $\alpha \in \pi_1(M)$ is orientation-preserving, then we know that a shortest loop representing α^n, for $n > 1$, always factors through a covering (cf. Lemma 1.4 of [2]). This happens sometimes for shortest one-sided curves (e.g., in the case that M has a hyperbolic metric). However, we have the following:

PROPOSITION 1.3. *Let M be a Riemannian surface, $M \neq \mathbf{R}P^2$, let $\alpha \in \pi_1(M)$ be orientation-reversing, and assume that there are shortest geodesics $f_1, f_2 \colon S^1 \to M$, both representing α with distinct images C_1, C_2. Then a shortest loop h representing α^n $(n > 1)$ never factors through a cover of a curve representing α.*

Proof. Suppose h factors through an n-fold covering and has image denoted by C. Then at least one of the two shortest geodesics representing α, say f_1, crosses C transversely in an odd number of points (by Z_2 intersection theory). We can then form a new loop h' by traversing C_1 $n-1$ times and C once, by "cutting and pasting" at some chosen crossing point of C_1 and C (cf. Figure 2). But then h' is homotopic to h and has length less than or equal to that of h. Rounding the corner of h' at the cut-and-paste point decreases length, and this contradicts h being shortest.

REMARK. By passage to a k-fold covering space $\tilde{M}$ of M, we can see that h cannot factor through a cover of a loop representing α^k, where $1 \leq k < n$. In fact, since a shortest curve representing α^k does not cover a loop representing α

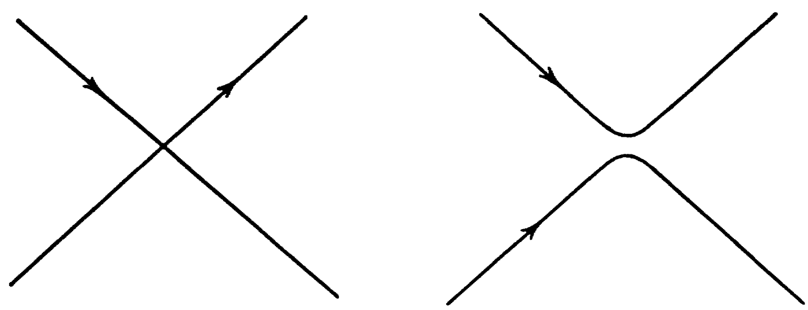

Figure 2

k times, it follows that there are at least two curves in $\tilde{M}$ with distinct images representing the generator α^k of $\pi_1(\tilde{M})$. Hence Proposition 1.3 can be applied in $\tilde{M}$.

2. The Möbius band. Since our methods will involve covering spaces, the Möbius band turns out to be the key space to understand. Let B denote a Möbius band with some Riemannian metric and let α be a generator of $\pi_1(B)$. We assume the metric is chosen so that shortest loops representing powers of α always exist. In applications, we will be considering compact Möbius bands B with ∂B geodesic and open Möbius bands covering closed Riemannian 2-manifolds. In both cases such shortest curves can be found.

LEMMA 2.1. *Let C be a shortest loop representing α. Then C is embedded. If C' is another such loop then C and C' either coincide or intersect transversely at a single point.*

Proof. Suppose C is not embedded. Then C crosses itself transversely at some point P. (C cannot be multiply-covered since $C \sim \alpha$). Perform a cut and paste at P as in Figure 2. This yields two new loops, one of which is one-sided. Repeating, we eventually arrive at an embedded one-sided curve with length less than that of C. Such a loop represents α and this contradicts the shortest length property of C.

If C and C' are distinct shortest loops, they must cross at least once, by Z_2 intersection theory. If they cross in at least two points, pick an arc λ in C with $\lambda \cap C' = \partial\lambda$. Let μ, μ' be the arcs of C' with $\partial\mu = \partial\mu' = \partial\lambda$. Then $\lambda \cup \mu$ and $\lambda \cup \mu'$ are closed curves in B, one of which is one-sided, say $\lambda \cup \mu$, and one two-sided, say $\lambda \cup \mu'$. Then $C_1' = C' - \mu' + \lambda$ and $C_1 = C - \lambda + \mu'$ are one-sided. Let $l(C_0)$ denote the length of curve C_0. We see that $l(C_1) + l(C_1') = 2l(C)$, and each of C_1, C_1' has a corner. By rounding the corner of, say, C_1 and C_1', we obtain a loop C_1'' which is shorter than C. If C_1'' is not embedded, we repeat the argument in the first paragraph to obtain an even shorter curve which is embedded. So we get a contradiction, since any such a curve is one-sided and so represents α. □

NOTE. This result applies also to $\mathbf{R}P^2$.

LEMMA 2.2. *Let C be a shortest geodesic representing α and let $g: S^1 \to B$ be a shortest loop representing α^2, with $g(S^1) = C_0$. Then either $C = C_0$ or C and C_0 are disjoint.*

Proof. Lemma 2.1 shows C is embedded, and Theorem 2.1 of [2] establishes that C_0 is embedded and g is either an embedding or a double cover. Suppose the latter is true. If C is distinct from C_0, then Proposition 1.3 gives a contradiction to g a covering. So $C = C_0$ in this case.

If g is an embedding, then C_0 is two-sided and bounds a Möbius band B_0 in B. If $C \cap C_0 \neq \varnothing$ then C crosses C_0 transversely, and we can find an arc λ of C with $\partial\lambda = \lambda \cap B$. Let μ, μ' be the arcs of C_0 with $\partial\lambda = \partial\mu = \partial\mu'$. Then $\lambda \cup \mu$ and $\lambda \cup \mu'$ are both embedded, two-sided loops, as they are disjoint from int B_0 and so one is contractible, say $\lambda \cup \mu$. But then the exchange argument of Lemma 2.1 applies and the result is proved. □

We next consider how shortest arcs intersect C, where C is a shortest loop $\sim \alpha$ and B is a compact Möbius band, with $\partial B = C_0$ a shortest geodesic $\sim \alpha^2$. Note that C lies in B.

LEMMA 2.3. *Let* $A\colon (I, \partial I) \to (B, C_0)$ *be an arc of shortest length in its homotopy class* rel ∂I. *Then either A intersects C transversely in a single point or A has image in* C_0.

Proof. Suppose first that A is homotopic rel ∂I to be an arc A' running along C_0. A' may run several times around C_0, but it is still shortest rel its boundary, by Lemma 1.1. By passing to a suitable covering space $\bar{B}$ of B, we may assume that a lift $\bar{A}'$ of A' is embedded. Also the covering $\bar{C}_0$ of C_0 lying in $\bar{B}$ is shortest, by Lemma 1.1, as is also the lift $\bar{A}$ of A with $\partial\bar{A} = \partial\bar{A}'$. But then

$$l(\bar{C}_0) = l(\bar{C}_0 - \bar{A} + \bar{A}'),$$

and rounding the corner of $\bar{C}_0 - \bar{A} + \bar{A}'$ gives a contradiction.

If A cannot be homotoped into C_0, then A transversely crosses C in at least one point, as $B - C$ retracts to C_0. Suppose there are two or more intersection points. Let $\tilde{B}$ denote the universal cover of B, let $\tilde{A}$ be a lift of A to $\tilde{B}$ and let $\tilde{C}$ be the line in $\tilde{B}$ covering C. Then $\tilde{A}$ meets $\tilde{C}$ in the same number of points as A intersects C. Therefore there is an arc in $\tilde{A}$ with both endpoints on $\tilde{C}$. Let μ be such an arc with the property that the arc λ of $\tilde{C}$ with $\partial\lambda = \partial\mu$ is as short as possible. If $l(\lambda) \le l(C)$ then we could make an exchange argument and get a contradiction, since in this case both λ and μ are shortest arcs (rel ∂), and they are homotopic. So we can assume $l(\lambda) > l(C)$.

In $\tilde{B}$ let x generate the covering transformation group. Clearly λ and $x\lambda$ overlap as in Figure 3. Since $\tilde{A}$ crosses $x\lambda \cup x\mu$ at least twice, either there is an arc in $\tilde{C}$ with ends on $\tilde{A}$ which is shorter than $x\lambda$ or else $\tilde{A}$ intersects $x\mu$. The former possibility is ruled out by the choice of μ. Similarly, $x\tilde{A}$ must cross μ and so $\tilde{A} \cap x\tilde{A}$ contains at least two points. But $\tilde{A}$ and $x\tilde{A}$ are shortest arcs and so another exchange argument gives a contradiction, completing the proof. □

REMARK 2.4. The same proof applies in an open Möbius band B to show that a length-minimizing line intersects a shortest loop which represents a generator of $\pi_1(B)$ in at most one point.

LEMMA 2.5. *Let B be a Riemannian Möbius band. Let* $f, g\colon S^1 \to B$ *be shortest geodesics in general position representing* α^k, α^m *respectively, where* α *generates* $\pi_1(B)$ *and* $1 \le k \le m$, k, m *odd. Then* $D(f) = k-1$ *and* $D(f, g) = k$.

REMARKS. The assumption that f, g are in general position rules out the possibilities that f covers its image and that the images of f and g coincide. The

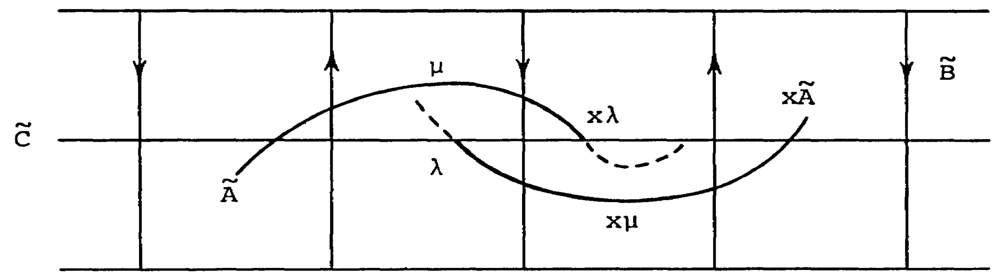

Figure 3

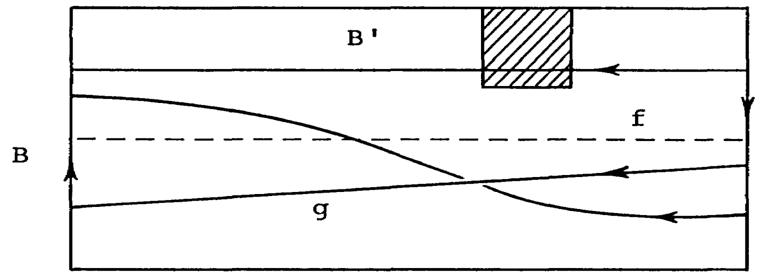

Figure 4

following lemma will show that these are the least possible numbers of double points.

Proof. To calculate $D(f)$, we look in the k-fold cover B_k of B. If $\tilde{f}$ is a lift of f to B_k, then $\tilde{f}$ is a shortest loop representing a generator of $\pi_1(B_k)$ and so is embedded by Lemma 2.1. Clearly $D(f)$ is the number of intersections of all the other lifts of f to B_k with $\operatorname{Im}\tilde{f}$. But Lemma 2.1 implies that each of these $k-1$ curves crosses $\operatorname{Im}\tilde{f}$ once, so that $D(f)=k-1$ as claimed. Note that no two of the lifts of f can coincide, since f is in general position.

We next check that $D(f, g)=1$ if $k=1$. A disk (shaded in Figure 4), which misses $f(S^1)$ and meets $g(S^1)$ in a small embedded arc, can be removed from B to form a new Möbius band B'. Then $g(S^1)\cap B'$ is a shortest arc, rel ∂, and we can apply Lemma 2.3 (with minor changes to the proof, since $\partial B'$ is no longer a shortest geodesic) to conclude that $g(S^1)$ and $f(S^1)$ intersect at most once. Hence $D(f, g)=1$ as desired.

Let C be a shortest curve representing α and consider now $D(f, g)$ for f shortest $\sim\alpha^k$ and g shortest $\sim\alpha^m$, where $k, m>1$. It follows that $f(S^1)$ and $g(S^1)$ intersect C at one point each, by the previous case. Note that $f(S^1)=C$ is ruled out since f is in general position, and similarly for g.

We now look in B_{km}, the km-fold cover of B. Pick lifts $\hat{f}$ and $\hat{g}$ of the m-fold and k-fold covers of f and g (respectively) to B_{km}, and let $\hat{C}$ be a km-fold cover of C in B_{km}. Note that $\hat{f}$ and $\hat{g}$ are embeddings, since they cover shortest loops $\tilde{f}: S^1\to B_k$ and $\tilde{g}: S^1\to B_m$ which are embeddings by Lemma 2.1 ($\tilde{f}$ is a lift of f to the k-fold cover of B and $\tilde{g}$ is defined analogously).

In Figure 5, the loops $\hat{f}$, $\hat{g}$ and $\hat{C}$ are depicted in B_{km} for the case where $k=3$, $m=5$.

Note that $\operatorname{Im}\hat{f}$ intersects $\hat{C}$ in m points and $\operatorname{Im}\hat{g}$ and $\hat{C}$ cross in k points, since $f(S^1)\cap C$ and $g(S^1)\cap C$ both contain one point. Since $m>1$, an arc γ of $\operatorname{Im}\hat{f}$ can be chosen with ends on $\hat{C}$ and interior disjoint from $\hat{C}$. As γ is homotopic into $\hat{C}$, there is a disk in B_{km} with boundary consisting of γ and an arc of $\hat{C}$. By an innermost disk argument, we can then find a disk D in B_{km} which satisfies $\partial D=\lambda\cup\mu$, where λ is an arc of $\operatorname{Im}\hat{f}$, μ is an arc of $\hat{C}$, and $\operatorname{int} D$ is disjoint from $\operatorname{Im}\hat{f}$ and $\hat{C}$. Suppose that $\operatorname{Im}\hat{g}$ crosses μ. Then $\operatorname{Im}\hat{g}$ must intersect λ as $\hat{g}$ meets μ in at most one point, since the points of $\operatorname{Im}\hat{g}\cap\hat{C}$ are spaced at distance $m\cdot l(C)$ along $\hat{C}$, a distance $\geq k\cdot l(C)$, which is the distance between the endpoints of μ.

On the other hand, suppose there is an arc ν of $\operatorname{Im}\hat{g}\cap D$ with $\partial\nu\subset\lambda$. Clearly ν lies between two successive intersections of $\operatorname{Im}\hat{g}$ and $\hat{C}$. But $\hat{g}$ is shortest between

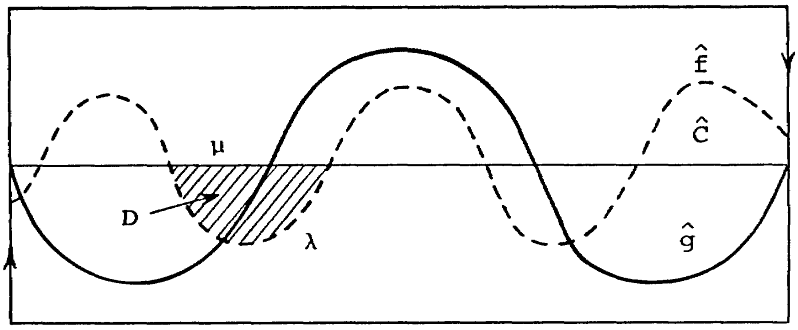

Figure 5

such points, and so ν is shortest rel endpoints. Similarly, if τ is the subarc of λ with $\partial\tau=\partial\nu$, then τ is shortest rel ends also. An exchange argument between ν and τ then gives a contradiction.

We conclude that $\operatorname{Im}\hat{g}$ crosses D in at most one arc, which must have one endpoint on each of λ,μ. Let x generate the covering transformation group for $B_{km}\to B$. Applying the same argument using the disks $x^{ik}D$, for $1\le i\le m-1$, we find that $D(\hat{C},\hat{g})=D(\hat{f},\hat{g})=k$. Notice that $\hat{C}=\bigcup_{1\le i\le m-1}x^{ik}\mu$ and so all the intersections of $\operatorname{Im}\hat{g}$ and $\hat{C}$ occur in the disks $x^{ik}D$. There are k lifts of an m-fold cover of f in the total pre-image of f in B_{km}, and m lifts of a k-fold cover of g to B_{km} in the full pre-image of g. The entire number of crossings of all the lifts of f with the total pre-image of g is thus k^2m. Dividing by km, we see that f and g intersect in k points as claimed. □

We now show these values are the best possible.

LEMMA 2.6. *Let B be a Möbius band and let α be a generator of $\pi_1(B)$. Then $D(\alpha^k)=k-1$ and $D(\alpha^k,\alpha^m)=k$, if $1\le k\le m$ are odd.*

Proof. It is clear, from the proof of Lemma 2.5, that $D(\alpha^k)=k-1$ and that $D(\alpha^k,\alpha^m)\le k$. We will show that $D(\alpha^k,\alpha^m)\ge k$, completing the argument.

Let f,g be general position curves representing α^k,α^m respectively. Choose an arc from ∂B to a point on either $\operatorname{Im} f$ or $\operatorname{Im} g$, but meeting $\operatorname{Im} f\cup\operatorname{Im} g$ only at this endpoint. Let $B'=\operatorname{cl}(B-D)$, where D is a small regular neighborhood of this arc. Then one of f or g produces a proper arc f' or g' in B'.

Let $\tilde{B}'$ be the universal cover of B'. The arc f' or g' lifts to a proper arc $\tilde{f}'$ or $\tilde{g}'$ in $\tilde{B}'$ which has one endpoint on each component of $\partial\tilde{B}'$. (Otherwise $\tilde{f}'$ or $\tilde{g}'$ is homotopic rel ends into $\partial\tilde{B}'$ and this projects to a similar homotopy for f' or g', contradicting k,m odd.) Also the loop f or g in B' lifts to a total of k or m lines in $\tilde{B}'$. Hence we see that $\tilde{f}'$ or $\tilde{g}'$ establishes the desired conclusion that $D(\alpha^k,\alpha^m)\ge k$. □

3. The general case. We first prove an embedding result for shortest loops.

THEOREM 3.1. *Let M^2 be a closed Riemannian 2-manifold and let $\alpha\in\pi_1(M)$ be represented by an embedded one-sided loop. Then any shortest curve $f: S^1\to M$ representing α is an embedding. Any two such shortest loops intersect in a single point or have identical images.*

Proof. Let $g: S^1 \to M$ be a shortest loop which represents α^1. It follows from Theorem 2.1 of [2] that C_0, the image of g, is embedded, and that g is an embedding or a double covering of C_0. In the latter case, the argument in Lemma 2.2 shows that f must have image C_0 and the theorem is proved. □

So we can assume that C_0 is an embedded two-sided loop in M and hence bounds a Möbius band B_0, by classical surface theory. Let M_α be the covering of M corresponding to the subgroup of $\pi_1(M)$ generated by α. Then M_α is a Möbius band. Let $\tilde{B}_0$ be a component of the pre-image of B_0 in M_α which projects homeomorphically onto B_0. Note that $\partial\tilde{B}_0 = \tilde{C}_0$ is a lift of C_0 and so is a shortest loop representing α^2. By Lemmas 2.1 and 2.2, f lifts to an embedding $\tilde{f}: S^1 \to M_\alpha$ which is a shortest loop representing α and has image disjoint from $\tilde{C}_0$. But $\operatorname{Im}\tilde{f}$ is contained in the smaller Möbius band $\tilde{B}_0$ in M_α, and $\tilde{B}_0$ projects one-to-one onto B_0. It now follows that f is an embedding in M with image in B_0. The second part of the theorem follows by Lemma 2.1. □

We next look at intersections of shortest one-sided loops with shortest two-sided curves.

THEOREM 3.2. *Let f, g be shortest loops representing α, β respectively in a closed Riemannian 2-manifold M. If f is one-sided, g is two-sided and f, g are in general position, then $D(f, g) = D(\alpha, \beta)$.*

Proof. Let M_α be the cover corresponding to the subgroup of $\pi_1(M)$ generated by α. A lift $\tilde{f}$ of f to M_α is an embedding which is shortest $\sim\alpha$ in the Möbius band M_α. The pre-image of g in M_α is a collection of length-minimizing lines and (possibly) shortest two-sided loops, by Lemma 1.1. As in [2], we refer to these lines and loops as components of the pre-image of g. A line component meets $\operatorname{Im}\tilde{f}$ at most once by Remark 2.4. A loop component either is disjoint from $\operatorname{Im}\tilde{f}$ or coincides with $\operatorname{Im}\tilde{f}$, by Lemma 2.2 above and Lemma 1.4 of [2].

Any homotopy of f or g lifts to a proper homotopy in M_α and so cannot decrease the number of intersection points of $\operatorname{Im}\tilde{f}$ with the pre-image of g, which is equal to $D(f, g)$. Thus $D(f, g) = D(\alpha, \beta)$ and the theorem is proved. □

We now show that a pair of shortest one-sided loops minimizes intersection. This includes the case of self-intersections of a single one-sided curve.

THEOREM 3.3. *Let M be a closed Riemannian 2-manifold. Let α, β be distinct orientation-reversing elements of $\pi_1(M)$. Let f, g be shortest loops in general position representing α, β respectively. Then $D(f) = D(\alpha)$ and $D(f, g) = D(\alpha, \beta)$.*

Proof. If M is a projective plane then $D(f) = D(\alpha)$ follows by Lemma 2.1. Assume that M is not a Klein bottle. We first show that $D(f) = D(\alpha)$.

Let $f_0: S^1 \to M$ be a shortest loop $\sim\alpha^2$ and let M_α be the cover of M corresponding to the subgroup of $\pi_1(M)$ generated by α. Let $\mathcal{U}$ be the universal cover of M. If $\tilde{f}: S^1 \to M_\alpha$ is a lift of f, then $\tilde{f}$ is an embedding with image denoted by C, by Lemma 2.1. Hence the pre-image of f in $\mathcal{U}$ is a collection of embedded lines (which we will call the components of the pre-image of f). Let $\tilde{f}_0: S^1 \to M_\alpha$ be a lift of f_0. By Lemma 2.2, the image of $\tilde{f}_0$ is an embedded curve C_0 and either

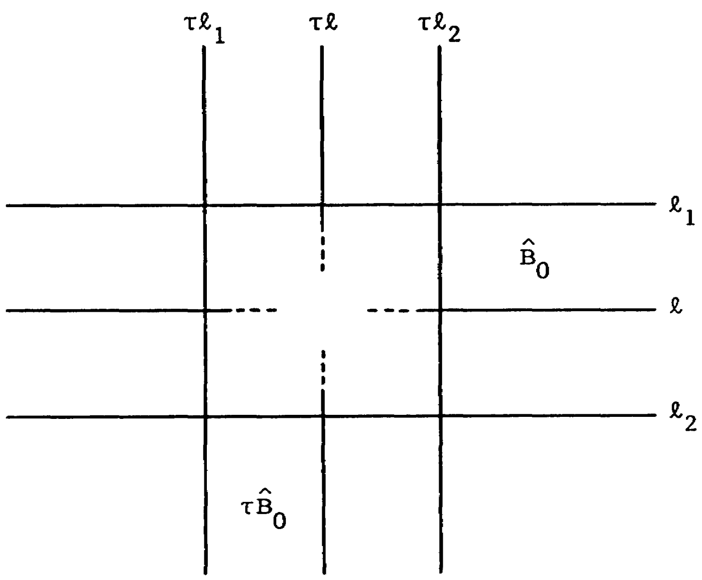

Figure 6

$C_0 = C$ or C_0 bounds a Möbius band B_0 containing C. If $C = C_0$, both parts of the theorem follow by Theorem 3.2, so we can suppose that $C_0 \cap C = \varnothing$.

In $\mathcal{U}$, the pre-image of f_0 is a collection of embedded length-minimizing lines (by Lemma 1.1) and the pre-image $\hat{B}_0$ of B_0 is bounded by two of these lines, say l_1 and l_2. Clearly $\hat{B}_0$ contains exactly one component, say l, of the pre-image of f. We call $\hat{B}_0$ a strip and will now consider how $\pi_1(M)$, acting as covering transformations on $\mathcal{U}$, moves such a strip.

By Lemma 3.1 of [2], two components of the pre-image of f_0 can meet in at most one point. If l_3 is such a component, then l_3 projects to a (length-minimizing) line or loop in M_α, which meets C_0 in an even number of points, since $C_0 = \partial B_0$. Lifting back to $\mathcal{U}$, we conclude that l_3 meets l_1 if and only if it meets l_2. Therefore a non-trivial intersection of $\hat{B}_0$ with some distinct translate $\tau\hat{B}_0$ can only be as in Figure 6. Note also that since l projects to C and τl_1 maps to a two-sided shortest line or loop in M_α, by Remark 2.4 and Theorem 3.2, $l \cap \tau l_1$ is a single point as shown in Figure 6. What is not apparent is the nature of the intersection between l and τl. We will see that these two lines intersect in exactly one point.

The strip $\hat{B}_0$ is stabilized by α and α^2. α^2 is an orientation-preserving map which translates points a distance $l(C_0)$ along l, l_1 and also l_2. Suppose $\sigma\hat{B}_0$ is a strip crossing $\hat{B}_0$, where σ can be assumed to be orientation-preserving. Then we claim that $\alpha^2\sigma\hat{B}_0$ is disjoint from $\hat{B}_0$, as in Figure 7. For if this is not true, then each pair of l_1, σl_1, and $\alpha^2\sigma l_1$ will intersect non-trivially.

However if M has a hyperbolic metric, this cannot happen as α^2 is simply a hyperbolic isometry translating points along the unique geodesic in $\mathcal{U}$ which is invariant under α^2 (and α). The picture is shown in Figure 8, in the hyperbolic case. We denote M equipped with a hyperbolic metric by M', and the corresponding geodesics by f', f_0', l', l_1', and l_2'. Now however f' and f_0' coincide, as do their lifts l', l_1', and l_2'. There is a homotopy of f_0 to f_0' in M', which moves any point on $\operatorname{Im} f_0$ a distance smaller than K, where K is some constant and distance is mea-

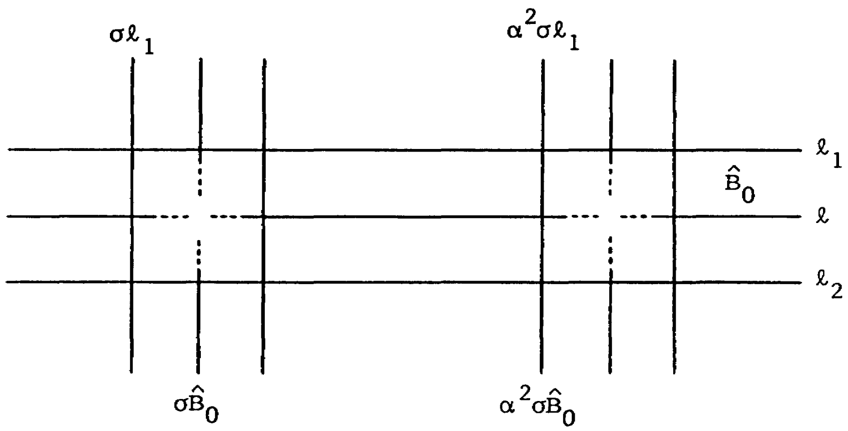

Figure 7

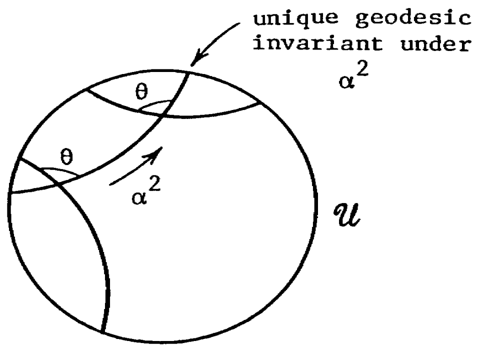

Figure 8

sured in the hyperbolic metric of M'. This homotopy lifts to $\mathcal{U}'$, the universal cover, giving a homotopy of l_1 to l_1' moving no point more than distance K, and similarly moving each translate of l_1 to the corresponding translate of l_1'. Suppose that σl_1 and $\alpha^2\sigma l_1$ intersect, but that $\sigma l_1'$ and $\alpha^2\sigma l_1'$ do not. Outside a large compact disk in $\mathcal{U}'$, which is just hyperbolic space, the geodesics $\sigma l_1'$ and $\alpha^2\sigma l_1'$ are never within distance $2K$ from one another. Thus it follows that there is a homotopy supported in this disk which makes σl_1 and $\alpha^2 l_1$ disjoint. Since these lines are length-minimizing and thus minimize the number of intersections in their compactly supported homotopy class, as in Lemma 3.1 of [2], this implies that σl_1 and $\alpha^2\sigma l_1$ are disjoint. More generally, $\alpha^{2n}\sigma\hat{B}_0$ is disjoint from $\sigma\hat{B}_0$ by the same method.

Consider now the intersection of the strips $\hat{B}_0$ and $\tau\hat{B}_0$ in $\mathcal{U}$, as in Figure 6. We will show that l and τl intersect at exactly one point, the minimal number possible in their proper homotopy class.

Suppose $l \cap \tau l$ has 3 points. Let $A = l \cap \tau\hat{B}_0$ and let $E = \tau l \cap \hat{B}_0$, so that A and E are arcs crossing 3 times. By the previous argument, since $\tau\hat{B}_0$ is disjoint from $\alpha^2\tau\hat{B}_0$ and α^2 translates points along l by a distance $l(C_0) = 2l(C)$, it follows that $l(A) < 2l(C)$ and similarly $l(E) < 2l(C)$. Clearly there are subarcs of A with end-

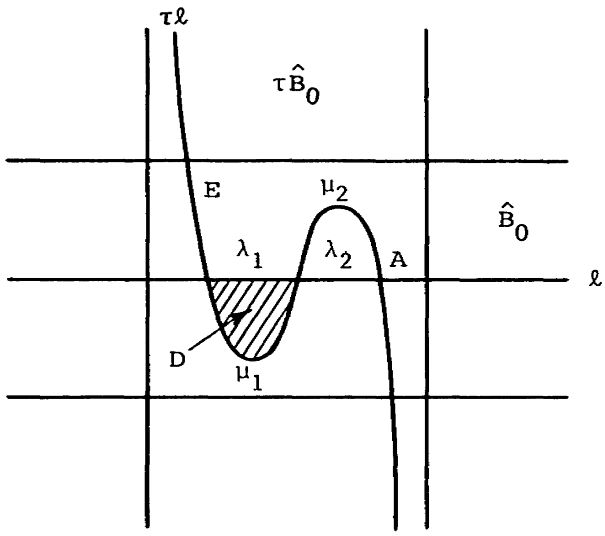

Figure 9

points on E which give rise to 2-gons in $\mathcal{U}$. Among all 2-gons in $\mathcal{U}$ between translates of l, we can pick one which is innermost (i.e., contains no smaller such 2-gon), and without loss of generality we can assume it is bounded by arcs λ_1 and μ_1 contained in A and E respectively. The situation is as in Figure 9. Note that translates of l may cross this 2-gon.

If $l(\lambda_1)<l(C)$ and $l(\mu_1)<l(C)$ then an exchange between projections of λ_1 and μ_1 gives a contradiction, and similarly for λ_2, μ_2 in place of λ_1, μ_1. So we can assume $l(\lambda_1)>l(C)$, $l(\mu_1)<l(C)$, $l(\lambda_2)<l(C)$, and $l(\mu_2)>l(C)$ without loss of generality, since $l(\lambda_1)+l(\lambda_2)\leq l(A)<2l(C)$, and similarly for μ_1, μ_2.

We now translate l and τl by applying α, as in Figure 9, and suppose points move to the left along l by a distance $l(C)$, without loss of generality. (Otherwise replace α by α^{-1}.) Since $l(\lambda_2)<l(C)$ and $l(\lambda_1)>l(C)$, it follows that $\alpha\lambda_2\subset\lambda_1$. Let D denote the disk bounded by $\lambda_1\cup\mu_1$. $\alpha\mu_2$ cannot lie in D, as D is an innermost 2-gon, but $\alpha\mu_2$ certainly lies on the same side of l as does D, because α is orientation-reversing. Thus the picture is as in Figure 10. As $\delta\subset\mu_1$, we have $l(\delta)<l(C)$. Thus $l(\gamma)>l(C)$ or else we can do an exchange and get a contradiction. Thus we must have that $l(\rho)<l(C)$, since $l(\rho)+l(\gamma)\leq l(\alpha E)<2l(C)$. But $l(\eta)<l(C)$ also, as $l(\eta)+l(\mu_2)\leq l(E)<2l(C)$. Thus there is an exchange between projections of ρ and η, giving a contradiction. A similar argument shows that there cannot be more than 3 points of intersection between l and τl.

We will compute $D(f)$ by looking in the cover M_α. Clearly $D(f)$ is the number of transverse intersection points of C with all the components of the pre-image of f in M_α, excepting C itself. We will show that each component intersects C in at most one point. This will imply that any proper homotopy in M_α cannot decrease $D(f)$, and so $D(f)=D(\alpha)$.

Let C' be a component of the pre-image of f in M_α. If C' is a line which meets C in at least two points, then there is an arc in C' with endpoints on C which is homotopic into C rel ends (M_α retracts to C). So there is a lift l' of C' to $\mathcal{U}$ which crosses l in two points or more, contrary to the preceding argument. Thus $C'\cap C$ has at most one point.

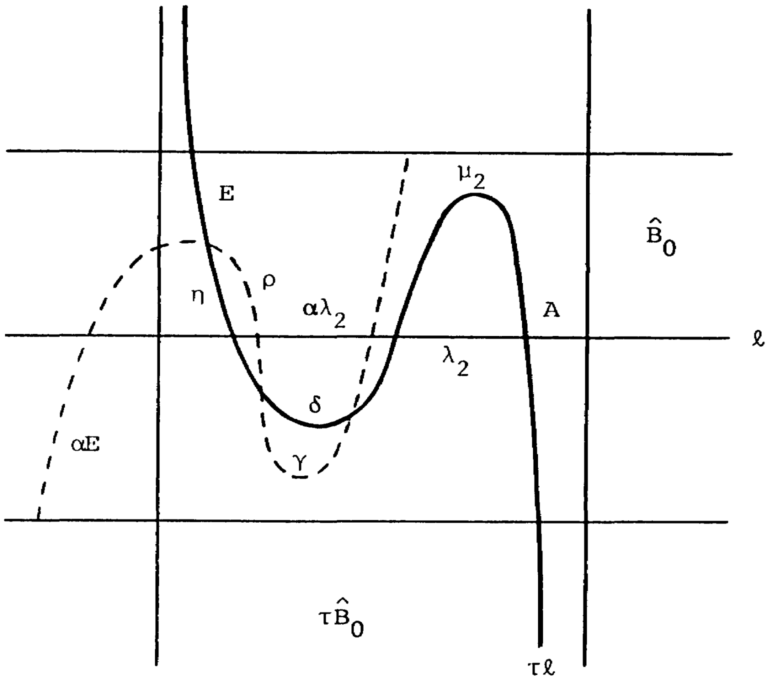

Figure 10

If C' is a loop, then C' is homotopic to some multiple C^n of the loop in C in M_α. We claim that $n=1$ (with correct orientations). Let $p: M_\alpha \to M$ be the covering projection and let $p_1, p_2: S^1 \to S^1$ be m- and n-fold coverings respectively. Assume that C' is the image of $\tilde{f}': S^1 \to M_\alpha$, which is a lift of $f \cdot p_1$ to M_α. If we apply p to the homotopy between $\tilde{f}'$ and $\tilde{f} \cdot p_2$, a homotopy in M is obtained between $f \cdot p_1$ and $f \cdot p_2$. This gives an equation $\gamma^{-1}\alpha^m\gamma = \alpha^n$ in $\pi_1(M)$, since $f \sim \alpha$. As M is not a Klein bottle, any 2-generator subgroup of $\pi_1(M)$ is free. Hence γ commutes with α and $m=n$. Let l, l' be the line in $\mathcal{U}$ over C, C'. Then $l' = \gamma l$ and so l, l' are both stabilized by the cyclic subgroup generated by α, because $\alpha\gamma = \gamma\alpha$. We conclude that $m=n=1$ and C' is homotopic to C. (Note that if, e.g., $\alpha = \gamma^3$, then there are at least 3 loops in the pre-image of f in M_α, since f is in general position.) But then C' is also shortest $\sim \alpha$, and so $C \cap C'$ has just one point, by Lemma 2.1. This completes the proof that $D(f) = D(\alpha)$.

To show that $D(f, g) = D(\alpha, \beta)$, we first note that any line component in the pre-image of g in $\mathcal{U}$ intersects l in at most one point, by essentially the same argument as used for translates τl meeting l. Also $D(f, g)$ is the number of crossings of C with the full pre-image of g in M_α. Exactly as above, it follows that any line component of the pre-image of g intersects C in at most one point.

If C' is a loop component, we need to show that the number of intersections of C' and C cannot be reduced by any homotopy of both f and g in M. As previously, the homotopy between C' and some multiple C^n of C projects to an equation $\gamma^{-1}\beta^m\gamma = \alpha^n$ in $\pi_1(M)$. Hence $\gamma^{-1}\beta\gamma = \delta^q$ and $\alpha = \delta^r$ for some $\delta \in \pi_1(M)$, since $\gamma^{-1}\beta\gamma$ and α must belong to a cyclic subgroup of $\pi_1(M)$. Note that $mq = nr$.

If M_δ denotes the covering of M corresponding to the subgroup generated by δ, then there is an r-fold covering $\bar{p}: M_\alpha \to M_\delta$. Let $\tilde{g}: S^1 \to M_\alpha$ be the lift of $g \cdot p_1$ to M_α with image C', where $p_1: S^1 \to S^1$ is an m-fold cover. Also let $\bar{f}, \bar{g}: S^1 \to M_\delta$ be the lifts of f, g with images $\bar{C}, \bar{C}'$ respectively, where $\bar{C}=\bar{p}(C)$ and $\bar{C}'=\bar{p}(C')$. Since $\bar{f}\sim\delta^r$ and $\bar{g}\sim\delta^q$, it follows that $\bar{f}$ and $\bar{g}$ are shortest geodesics which intersect in $\min\{q,r\}$ points, by Lemma 2.5. Let M_{qr} denote the qr-fold covering of M_δ and let $\hat{C}, \hat{C}'$ be components of the pre-images of $\bar{C}, \bar{C}'$ (respectively) in M_{qr}. Then $\hat{C}\cap\hat{C}'$ also contains $\min\{q,r\}$ points, by the proof of Lemma 2.5. Projecting to M_α, $\hat{C}$ is a q-fold cover of C and $\hat{C}'$ is a (q,r)-fold cover of C', where (q,r) is the g.c.d. of $\{q,r\}$. Hence C intersects C' in $\min\{q,r\}/(q,r)$ points. So if $\#\hat{C}\cap\hat{C}'$ is decreased by some homotopy in M of f and g, we see that $\#C\cap C'$ is also reduced, contrary to the argument in Lemma 2.6. This finishes the proof that $D(f,g)=D(\alpha,\beta)$.

Assume finally that M is a Klein bottle. Then $\pi_1(M)=\langle x,y\mid x^{-1}yx=y^{-1}\rangle$. An arbitrary element of $\pi_1(M)$ can be expressed as $x^m y^n$ and is orientation-reversing if and only if m is odd. In this case, $x^m y^n$ is conjugate to either x^m or $(xy)^m$. So a shortest one-sided geodesic represents either x^m or $(xy)^m$.

A shortest loop C_0 in M which represents $x^2=(xy)^2$ is embedded by Theorem 2.1 of [2], as there is a two-sided simple loop representing x^2, or C_0 is a double cover of an embedded one-sided loop C_0'. In the former case C_0 separates M into 2 Möbius bands which have center-lines representing x and xy. So by Theorem 3.2, any one-sided length-minimizing geodesic in M is disjoint from C_0, since it is homotopic to a multiple of one of these center-lines. To analyze self-intersections and intersections of one-sided shortest loops, it now suffices to work in a Möbius band, and so Lemmas 2.5 and 2.6 complete the argument. In the case that C_0 covers C_0', $M-C_0'$ is a single Möbius band and this case follows from Lemmas 2.5 and 2.6 also. □

Appendix. Let M^3 be a closed $\mathbf{R}P^2$-irreducible Riemannian 3-manifold, that is, there are no two-sided embeddings of $\mathbf{R}P^2$ in M and any embedded S^2 bounds a 3-ball. Let F be a closed surface different from S^2 and $\mathbf{R}P^2$. Suppose $f: F\to M$ is a least area incompressible map which is homotopic to a one-sided embedding f', that is, $f_*: \pi_1(F)\to\pi_1(M)$ is one-to-one and f has smallest area in its homotopy class. It is reasonable to expect that f is an embedding, by analogy with the two-sided case (cf. Theorem 5.1 of [3]).

Let $g: F_0\to M$ be a two-sided embedding onto the boundary of a regular neighborhood of $f'(F)$, where $p: F_0\to F$ is the double covering with the property that a loop C in F lifts to F_0 whenever a curve homotopic to $f'(C)$ has odd intersection number with $f'(F)$, and g is homotopic to $f'p$. By Theorem 5.1 of [3], since fp is homotopic to the two-sided embedding g, a least area map g^* representing fp is an embedding or a two-to-one map. In the latter case, $g^*=f^*p$, where $f^*: F\to M$ is an embedding. Also $fp: F_0\to M$ is least area in its homotopy class, since $\text{Area}(fp)=2\,\text{Area}(f)\le 2\,\text{Area}(f^*)=\text{Area}(g^*)$. By Theorem 5.1 of [3], we conclude that f must be an embedding in this case and so it suffices to assume that g^* is an embedding.

By Theorem VII.9 of [5], the covering $\tilde{M}$ of M corresponding to the subgroup $f_*(\pi_1(F))$ of $\pi_1(M)$ is an open twisted line bundle over a non-orientable surface homeomorphic to F. The maps f and g^* lift to maps $\tilde{f}$ and $\tilde{g}^*$ from F and F_0 respectively to $\tilde{M}$. Clearly $\tilde{f}$ and $\tilde{g}^*$ are both least area and $\tilde{f}$ is a homotopy equivalence. *Suppose one could show, in the special case that a one-sided least area map is a homotopy equivalence, then $\tilde{f}$ is an embedding.*

If $\tilde{f}(F)$ met $\tilde{g}^*(F_0)$, then exactly as in Lemma 4.1 of [3] there would be a product region between these embedded surfaces. So an exchange argument would reduce the area of one of the surfaces, which gives a contradiction. We conclude that $\tilde{f}$ and $\tilde{g}^*$ have disjoint images and so clearly $\tilde{f}(F)$ lies in the compact region of $\tilde{M}$ bounded by $\tilde{g}^*(F_0)$. In fact since g^* is homotopic to g, $g^*(F_0)$ bounds a twisted line bundle in M which lifts to the compact region with boundary $\tilde{g}^*(F_0)$ in $\tilde{M}$ (for suitable choice of $\tilde{g}^*$). Hence $\tilde{f}$ projects one-to-one to the embedding f, and it follows that a least area incompressible map which is homotopic to a one-sided embedding must be an embedding.

REFERENCES

1. W. Ballmann, *Doppelpunktfreie geschlossene Geodätische auf kompakten Flächen,* Math. Z. 161 (1978), 41–46.
2. M. Freedman, J. Hass, and P. Scott, *Closed geodesics on surfaces,* Bull. London Math. Soc. 14 (1982), 385–391.
3. ———, *Least area incompressible surfaces in 3-manifolds,* Invent. Math. 71 (1983), 609–642.
4. J. Hass, *Complete stable minimal surfaces that are not totally geodesic,* Pacific J. Math., to appear.
5. W. Jaco, *Lectures on 3-manifold topology,* CBMS Regional Conf. Series Number 43, American Math. Soc., Providence, R.I., 1980.

Department of Mathematics
University of Michigan
Ann Arbor, Michigan 48109

and

Department of Mathematics
University of Melbourne
Parkville, Victoria 3052
Australia