THE LENGTH OF THE SHORTEST CLOSED GEODESIC ON A 2-DIMENSIONAL SPHERE

A. Nabutovsky and R. Rotman

Abstract. Let M be a Riemannian manifold homeomorphic to S^2. We prove that $l(M)$, the length of the shortest closed geodesic on M, does not exceed $4d$, where d denotes the diameter of M. We also prove that $l(M) \leq 8\sqrt{A}$, where A is the area of M. These results improve previously known inequalities by C. Croke [2] and Maeda [4].

1. Introduction

The connection between the length of a shortest closed geodesic, and diameter, or volume of a compact Riemannian manifold has been of interest to many people. Despite that fact the only known curvature free results establishing that connection for a simply connected manifold are the results of Croke [2] and Maeda [4]. Croke proved that if a Riemannian manifold M is diffeomorphic to the 2-dimensional sphere then $l(M) \leq (25+4\sqrt{2})\sqrt{A}$ $(< 31\sqrt{A})$ and $l(M) \leq 9d$. The last inequality was later improved by Maeda who demonstrated that $l(M) \leq 5d$ if M is diffeomorphic to S^2. Nevertheless, even in this simplest case of a 2-dimensional sphere the question about the optimal constants remains open. Croke conjectured that $l(M) < (12)^{\frac{1}{4}}\sqrt{A}$. (It is easy to see that the optimal constant cannot be smaller than $12^{\frac{1}{4}}$, see [2]). We do not know a counterexample to the conjecture that $l(M) \leq 2d$.

In this paper we will improve the known estimates on the length of a shortest closed geodesic and obtain the following two theorems.

Theorem 1. *Let M be a manifold diffeomorphic to the 2-dimensional sphere. Then $l(M) \leq 4d$.*

Theorem 2. *Let M be a manifold diffeomorphic to the 2-dimensional sphere. Then $l(M) \leq 8\sqrt{A}$.*

In order to prove Theorem 2 we will use a combination of the technique used by Croke in [2] with the F.Almgren-J.Pitts technique from the geometric measure theory adapted by Calabi and Cao in [3] to study shortest closed geodesics on surfaces. In order to prove Theorem 1 we will combine the Almgren-Pitts technique with an obstruction theory trick.

The approach used by Croke in [2] can be described as the effectivization of the existence theorem of Lusternik and Fet, that states that on any compact Riemannian manifold there exists at least one closed geodesic. From their proof it becomes obvious that in order to estimate the length of a shortest closed geodesic on a manifold diffeomorphic to S^2 it is enough to find a 1-parameter family of curves of controled length starting and ending at a point curve in such a way that the induced map $F: S^2 \longrightarrow S^2$ is of nonzero degree, (or alternatively one can view this as a map $f: S^1 \longrightarrow \Lambda_L M$, where $\Lambda_L M$ is the space of closed continuous curves on M of length bounded from above by L and where $f: S^1 \longrightarrow \Lambda_L M$ is not contractible.) That approach is based on the fact that closed geodesics are critical points of the length functional or the energy functional $E: \Lambda_L M \longrightarrow R$, on the Morse theory and on the connection between homotopy groups of ΛM and M.

In the case when M is diffeomorphic to S^2 one can prove the existence of a closed geodesic in a different way. Namely, one considers the space $Z_1(M, Z)$ of one-dimensional cycles with integer coefficients on M (see [6] and its references for the definition and properties of spaces of cycles). The classical result of F. Almgren ([1]) implies that the fundamental group of this space is isomorphic to $H_2(M) = Z$. The length functional on the space of closed curves is closely related to the so-called mass functional on the space of one-dimensional cycles. (These two functionals coincide if we consider a closed curve with finitely many points of self-intersection as a one-dimensional cycle. In the case when the curve can backtrack over itself the values of these functionals can be different due to the cancellation of mass.) One can look for a critical value of the mass on the space of one-dimensional cycles looking at $\min_{[\gamma]\neq 0} \max_{t \in S^1}$ mass $(\gamma(t))$. Here the minimum is taken over all closed curves γ in $Z_1(M, Z)$ that are not null homotopic. Pitts proved that the minimax value is always attained at a one-dimensional cycle corresponding to a closed geodesic or the union of two closed geodesics (cf. Theorem 2.4 in Part II of [3]). The assumption that M is two-dimensional is quite important here. Without this assumption the minimal 1-cycle need not be a geodesic but could have a finite set of singular points.

Finally note that some generalizations of Theorems 1 and 2 in the case when M is an arbitrary closed Riemannian manifold (of an arbitrary dimension) can be found in our preprint [5].

2. Proofs of Theorems.

Calabi and Cao observed that one does not really need the language of geometric measure theory to study shortest closed geodesics on surfaces ([3]). They observed that one can consider the space of one-dimensional cycles on M that are "made" of either one or two closed curves on M instead of the (much larger) space of all one-dimensional cycles. One obtains a closed geodesic as (a connected component of) the cycle where the minimum over all non-contractible closed curves in this space of the maximal length of 1-cycle in the curve is attained.

More formally, Calabi and Cao consider the space $\Gamma = \{\Phi | \Phi = (\phi_1, \phi_2), \phi_i : [0,1] \longrightarrow M$ is a piecewise smooth path and $\{\phi_1(1), \phi_2(1)\} = \{\phi_1(0), \phi_2(0)\}\}$.

Calabi and Cao note that there is a natural map $\rho : \Gamma \longrightarrow Z_1(M, Z)$, which is just the addition of ϕ_1 and ϕ_2. The length functional on Γ is defined as follows: $L(\Phi) = L(\phi_1) + L(\phi_2)$. They further define a topology on Γ that makes ρ and L into continuous maps. (One can define such a topology, for example, as follows: Using the Nash embedding theorem we can assume that our Riemannian manifold M is isometrically embedded into an Euclidean space. Now one can define the distance between (ϕ_1, ϕ_2) and (ψ_1, ψ_2) as $\sup_{t\in[0,1], i\in\{1,2\}} |\phi_i(t) - \psi_i(t)| + \Sigma_{i=1}^2 \sqrt{\int_0^1 |\phi_i'(t) - \psi_i'(t)|^2 dt}$. Note however that this definition is somewhat different from the definition in [3].)

Following Pitts Calabi and Cao observed that one can always choose a non-contractible closed curve $\gamma \subset Z_1(M, Z)$, where $\min_{[\gamma]\neq 0} \max_{t\in S^1}$ mass $(\gamma(t))$ is attained (at some $t_0 \in S^1$), such that γ is the image under ρ of a closed curve $\tilde{\gamma} \subset \Gamma$ (Lemma A.1 in [3]), and that $\tilde{\gamma}(t_0)$ or a connected component of $\tilde{\gamma}(t_0)$ must be a non-trivial closed geodesic.

Therefore one can obtain an upper bound for the length of a shortest closed geodesic on M (when M is diffeomorphic to S^2) in the following way: One can construct a non-contractible $f : S^1 \longrightarrow \Gamma$ and find an upper bound for $\max\{L(f(t))|t \in S^1\}$. It follows from the above that this upper bound is an upper bound for the length of a shortest non-trivial closed geodesic. (A self-contained detailed and elementary proof of the last assertion can be found in our paper [5], where it is stated as Proposition 5.) This is precisely how we are going to derive the upper bounds in Theorems 1 and 2.

We need the following definition from the C. Croke paper [2].

Definition 1. *Let γ be a simple closed curve on M dividing M into two components. Then γ will be called convex to Ω, where Ω is one of those components if there exists an $\epsilon > 0$ such that for all $x, y \in \gamma$, satisfying $d(x,y) < \epsilon$ the minimizing geodesic τ from x to y lies in the closure of Ω.*

Croke had shown (see Lemma 2.2 in [2]) that if γ is convex to Ω and $\bar{\Omega}$ is compact then there exists a "large enough" N such that if we apply Birkhoff curve shortening process with N breaks to γ then all curves in the resulting homotopy γ_t satisfy:

(1) $\gamma_t \in \bar{\Omega}$

(2) γ_t is simple and convex to $\Omega_t = \Omega - \{x \in \gamma_s | 0 \leq s \leq t\}$. That is, if we apply Birkhoff curve shortening process to γ then curves in the homotopy stay in $\bar{\Omega}$.

As a corollary, if there are no geodesics in M of length $\leq$ length (γ), and $\gamma = \partial\Omega$ is convex to Ω then γ can be contracted inside Ω by a homotopy that passes through closed curves of length $\leq$ length (γ).

If $\gamma = \partial\Omega$ is piecewise geodesic and all angles between consecutive segments are $\leq \pi$ then, obviously, γ is convex to Ω.

Proof of Theorem 1:

Suppose $l(M) > 3d$. Let $g : S^2 \longrightarrow M$ be a diffeomorphism between the standard sphere S^2 and our manifold M. We are going to try to extend g to the 3-dimensional disc, which should not be possible, and as an obstruction we will obtain a new non-contractible map $F : S^2 \longrightarrow M$ over which we will have some control.

We will try to extend g as follows. Triangulate S^2 so that the induced triangulation on M is very fine and let δ be the diameter of 1-simplex of M. Triangulate D^3 as the cone over S^2. We will try to extend g skeleton by skeleton. To extend to 0-skeleton, pick a point p on M and let the new map take the center of the disc to that point. To extend to 1-skeleton we proceed as follows: Let edges that connect the center of the disc and the vertices be mapped to (any of) the shortest geodesic segments that connect the point p with the corresponding vertices on M. To extend to 2-skeleton, we need to be able to map simplices of the form $[\tilde{p}, \tilde{v}_1, \tilde{v}_2]$, where $\tilde{p}$ is the center of the disc, and $\tilde{v}_1, \tilde{v}_2$ are vertices of the triangulation of S^2. By our assumption, the closed curve that is formed by two geodesics $[p, v_1]$, $[p, v_2]$ that connect the point p with vertices $v_1 = g(\tilde{v}_1)$ and $v_2 = g(\tilde{v}_2)$ respectively and by the edge $[v_1, v_2]$ is contractible to a point along the curves of length $\leq 2d + \delta$. We let the image of $[\tilde{p}, \tilde{v}_1, \tilde{v}_2]$ be the surface generated by the above homotopy. Furthermore, we should not be able to extend our map to 3-skeleton. This means that there exists a non-contractible sphere: $F : \partial[\tilde{p}, \tilde{v}_1, \tilde{v}_2, \tilde{v}_3] \longrightarrow ([v_1, v_2, v_3] \cup [p, v_1, v_2] \cup [p, v_2, v_3] \cup [p, v_1, v_3])$, where $[v_1, v_2, v_3]$ is a 2-simplex in the triangulation of the sphere, and $[p, v_i, v_j]$ is a surface generated by the homotopy that connects the closed curve formed by the 1-simplex $[v_i, v_j]$ and two geodesic segments $[p, v_i], [p, v_j]$ to a point. Let γ_i be a minimal geodesic connecting p and v_i, $i = 1, 2, 3$. Let e_{12} be the edge connecting v_1 and v_2; e_{13} be the edge connecting v_1 and v_3 and e_{23} be the edge connecting v_2 and v_3. Let $\alpha_1 = \gamma_2 \cup -e_{12} \cup -\gamma_1$; $\alpha_2 = \gamma_1 \cup e_{13} \cup -\gamma_3$; $\alpha_3 = \gamma_3 \cup -e_{23} \cup -\gamma_2$; $\alpha_4 = e_{12} \cup e_{23} \cup -e_{13}$. Note that the length of $\alpha_i \leq 2d + \delta, i = 1, 2, 3$ and the length of $\alpha_4 \leq 3\delta$. Let us assume that δ is less than the convexity radius of the sphere, so that α_4 is contractible to v_1 without length increase inside $[v_1, v_2, v_3]$. Let us also assume that some homotopies contracting α_i to p_i for $i = 1, 2, 3$ without length increase inside $[p, v_1, v_2], [p, v_2, v_3], [p, v_1, v_3]$ are fixed. We are now going to construct $f : S^1 \longrightarrow \Gamma$. We will start with a pair of constant maps $\Phi = (\phi_1, \phi_2)$, where $\phi_1(t) = p_1$ and $\phi_2(t) = p_2$. Now, we see that $(p_1, p_2) \sim (\alpha_1, \alpha_2) \sim (\gamma_2 \cup -e_{12}, e_{13} \cup -\gamma_3) \sim (\alpha_3, \alpha_4) \sim (p_4, p_3) \sim (p_1, p_2)$, (see Figure 1). (Here and below "$\sim$" means "is homotopic". Each of these homotopies is a path in Γ. The union of these paths will be the desired circle in Γ.)

Figure 1 (b)-(d) shows that $(p_1, p_2) \sim (\alpha_1, \alpha_2)$ and the total length of curves in the homotopy is bounded from above by $4d + 2\delta$. Figure 1 (d)-(f) shows that

$(\alpha_1, \alpha_2) \sim (\gamma_2 \cup -e_{12}, e_{13} \cup -\gamma_3)$ and the total length of curves in the homotopy is bounded from above by $4d+2\delta$. Figure 1 (f)-(g) shows that $(\gamma_2 \cup -e_{12}, e_{13} \cup -\gamma_3) \sim (\alpha_3, \alpha_4)$ and the length of the curves in the homotopy is bounded from above by $2d+4\delta$, and finally, figure 1 (g)-(i) shows that $(\alpha_3, \alpha_4) \sim (p_3, p_4)$ and the length of curves in the homotopy is bounded above by $2d+4\delta$. One can also see that the length of curves in the homotopy is bounded from above by $4d+4\delta$. Finally, we let δ go to zero, to obtain $l(M) \leq 4d$.

However, it remains to check that we obtained a non-contractible closed curve in Γ in order to conclude that $l(M) \leq 4d$. This can be done as follows: We can consider this closed curve as a closed curve in the space of 1-cycles in M, and following [1] assign to it a two-cycle in M. This cycle in our case is homologous to the cycle made of four singular 2-simplices obtained by contracting the boundaries of $[p, v_1, v_2]$, $[p, v_1, v_3]$, $[p, v_2, v_3]$ and $[v_1, v_2, v_3]$ as above. Thus, the homology class of this cycle is equal to the image of the fundamental homology class of $S^2 = \partial[\tilde{p}, \tilde{v}_1, \tilde{v}_2, \tilde{v}_3]$ under the homomorphism induced by F. This class is non-zero since F is non-contractible. (Recall that we are talking about a map of S^2 into itself.) Using Almgren's isomorphism we see that our curve in the space of 1-cycles in M is not contractible. Therefore it is not contractible in Γ.

Further note that one can avoid using [1] here. Instead we can introduce a much easier "handmade" version of [1] which is, however, quite adequate for our purposes: Observe that Γ is a subspace of the space of maps of the disjoint union of two copies of $[0,1]$ into M (with an appropriate topology). Therefore to any map of S^1 into Γ we can assign a map from the quotient space X_1 of two cylinders $S^1 \times [0,1]$, where the four boundaries get identified in a possibly complex way, into M. It is however sufficient for us to consider those maps from S^1 to Γ for which the corresponding quotient space is a polyhedron. Similarly, to a map of D^2 into Γ we can assign a map from the quotient space X_2 of $D^2 \times ([0,1] \bigcup [0,1])$ with four boundaries identified in some way into M. Again, we will be only dealing with the case, when the quotient space is a polyhedron. Triangulating X_1 (or X_2) we obtain a singular 2-cycle (or, respectively, a singular 3-chain) that corresponds to a map of S^1 into Γ (respectively, D^2 into Γ). Observe that if we consider the singular 3-chain corresponding to a map of D^2 into Γ under this correspondence, then its boundary will be a singular 2-cycle corresponding to the restriction of this map to $S^1 = \partial D^2$. Therefore, a map of S^1 into Γ is contractible if and only if the corresponding 2-cycle represents 0 in $H_2(M)$. (A more detailed discussion of this easier version of [1] can be found in section 4 of our paper [5]. In particular, there we describe a slight modification of this construction that works even if X_1 (or X_2) is not assumed to be a polyhedron.)

Now observe that when we apply this construction to our map $f: S^1 \longrightarrow \Gamma$ then we obtain a singular 2-cycle homologous to the sum of the images under F of four triangles $[\tilde{p}, \tilde{v}_1, \tilde{v}_2], [\tilde{p}, \tilde{v}_1, \tilde{v}_3], [\tilde{p}, \tilde{v}_2, \tilde{v}_3]$ and $[\tilde{v}_1, \tilde{v}_2, \tilde{v}_3]$ (with signs corresponding to their orientations). The homology class of this chain is non-trivial since F is not contractible. Therefore f is not null homotopic. QED

Proof of Theorem 2: In the proof of theorem 2 we use ideas of the Croke proof of his upper bound $d(M) \leq 31\sqrt{A}$ (Theorem 4.2 in [2]). Let $x, y \in M$ be two points, such that $d(x,y) = d$. According to Croke there are four possibilities (see the proof of Theorem 4.2 in [2]):

(0) $l(M) \leq 2\sqrt{A}$;

(1) $d \leq 2\sqrt{A}$. In this case we can conclude that $l(M) \leq 8\sqrt{A}$ as a corollary of our Theorem 1;

(2) Let $\alpha(t)$ be a minimal geodesic joining x and y. For some $t \in (\sqrt{A}, d - \sqrt{A})$ there exist two geodesic loops γ_x and γ_y based at $\alpha(t)$ and intersecting only at that point of length $\leq 2\sqrt{A}$, such that γ_x separates M into two domains: Ω_x and its complement, where $x \in \Omega_x$ and γ_x is convex to Ω_x; γ_y separates M into two domains: Ω_y and its complement, where $y \in \Omega_y$ and γ_y is convex to Ω_y. Moreover $\Omega_x \cap \Omega_y = \emptyset$, and $\gamma_x \bigcup -\gamma_y$ is convex to the complement of $\Omega_x \bigcup \Omega_y$ in M, see Figure 2(a), (b)).

(3) Let $\alpha(t)$ be a minimal geodesic joining x and y. There exists a simple closed geodesic loop γ not contractible in $M - \{x, y\}$ of length $\leq 2\sqrt{A}$ separating M into two domains Ω_x and Ω_y such that $x \in \Omega_x$ and $y \in \Omega_y$ and either γ is based at $\alpha(\sqrt{A})$ and is convex to Ω_y, or γ is based at $\alpha(d - \sqrt{A})$ and is convex to Ω_x. Moreover, (a) γ intersects α only at its base point; and (b) the tangent vectors to γ at its beginning and end lie on opposite sides of the straight line tangent to α in the tangent plane of M at the base point of γ. (See Figure 4).

We will refer to the above fact as Croke's lemma. For the sake of completeness we are going to give a brief exposition of the proof of Croke's lemma at the end of the paper.

Let us first consider the case (2). Assume there is no closed geodesic of length $\leq 4\sqrt{A}$. Then γ_x is contractible in Ω_x to p_x and γ_y is contractible in Ω_y through curves of length $\leq 2\sqrt{A}$ and $\gamma_x \cup -\gamma_y$ is contractible to p through curves of length $\leq 4\sqrt{A}$. Moreover, $\gamma_x \bigcup -\gamma_y$ can be contracted to p in the complement of $\Omega_x \bigcup \Omega_y$ because it is convex to the complement of $\Omega_x \bigcup \Omega_y$. One can now constract $f : S^1 \longrightarrow \Gamma$, so that S^1 passes through pairs of curves of length $\leq 4\sqrt{A}$ as follows (see Fig. 2):

Suppose γ_x is contractible to p_x in Ω_x along the curves $\gamma_{x t}$, γ_y is contractible to p_y in Ω_y along the curves $\gamma_{y t}$ and finally, $\gamma_x \cup -\gamma_y$ is contractible to p in Ω along the curves γ_t. Then $(p_x, p_y) \sim (\gamma_{x t}, -\gamma_{y t}) \sim (\gamma_x, -\gamma_y) \sim \gamma_t \sim p \sim (p_x, p_y)$, (see Figure 2 (b)-(f)) and the total length of curves in the homotopy is bounded from above by $4\sqrt{A}$.

In the case (3), assume without loss of generality that there exists a simple closed loop at $\alpha(\sqrt{A})$ that is not contractible in $M - \{x, y\}$ and that is convex to Ω_y of length $\leq 2\sqrt{A}$. As in [2] we apply Berger's lemma which implies that there exist minimizing geodesics α_1, α_2 joining points x, y such that $\alpha_1'(0), \alpha_2'(0), \alpha'(0)$ do not lie in an open half plane. (It is possible that $\alpha_1 = \alpha_2$.) (Here it is crucial that the distance between x and y is equal to the diameter of M.)

Let a, b, c be unique points of intersection of γ with $\alpha, \alpha_1, \alpha_2$ respectively. Then the following can be said: (1) the length of a geodesic segment joining x and a equals $\sqrt{A}$ by assumption. We will denote this segment α_{xa}; (2) the length of a geodesic segment joining x and b and denoted α_{xb} is $\leq 2\sqrt{A}$; (3) the length of a geodesic segment α_{xc} that connects x and c is $\leq 2\sqrt{A}$. (The last two inequalities follow from the triangle inequality for the triangles xab and xac and the fact that the length of γ does not exceed $2\sqrt{A}$.) Furthermore, we will denote the segment of γ that connects a and c as γ_{ac}, the segment of γ that connects c and b as γ_{cb} and the segment that connects points b and a as γ_{ba}. Suppose $l(M) > 6\sqrt{A}$. Then the closed curve $\beta_1 = \alpha_{xa} \cup -\gamma_{ba} \cup -\alpha_{xb}$ separates M into two domains, where one of them is a subdomain of Ω_x, which we will denote Ω_1. Our curve is convex to Ω_1 and therefore is contractible in it to a point p_1 without length increase along the curves β_{1_t}. (In order to establish the convexity we need to check that all angles between consecutive geodesic segments along this curve do not exceed π. This fact is immediate from our constructions.) Similarly, the closed curve $\beta_2 = \alpha_{xa} \cup \gamma_{ac} \cup -\alpha_{xc}$ separates M into two domains, where one of the domains is subdomain of Ω_x and will be denoted Ω_2. β_2 is contractible in Ω_2 to a point p_2 without length increase along the curves β_{2t}; the closed curve $\beta_3 = -\alpha_{xc} \cup \alpha_{xb} \cup -\gamma_{cb}$ separates M into two domains, where one of them is also subdomain of Ω_x and will be denoted Ω_3; β_3 is contractible in Ω_3 to a point p_3 without length increase along the curves β_{3t} and, finally γ is contractible in the complement of $\Omega_1 \bigcup \Omega_2 \bigcup \Omega_3$ to a point p_4 without length increase along the curves γ_t, (see Figure 3 (a)). (Observe that the complement of $\Omega_1 \bigcup \Omega_2 \bigcup \Omega_3$ is Ω_y, and γ is convex to Ω_y by our assumptions.)

Finally, we will construct $f : S^1 \longrightarrow \Gamma$.

(1) $(p_1, p_2) \sim (-\beta_1, \beta_2)$ and the length of curves in this homotopy $L(\beta_{1t}) + L(\beta_{2t}) \leq \sqrt{A} + \sqrt{A} + 2\sqrt{A} + 2\sqrt{A} + 2\sqrt{A} = 8\sqrt{A}$, (see Figure 3 (b)-(d)).

(2) $(\beta_1, \beta_2) \sim (\alpha_{xb} \cup \gamma_{ba}, \gamma_{ac} \cup -\alpha_{xc}) = \beta_4$ and the length of curves in this homotopy is bounded from above by $8\sqrt{A}$, (see Figure 3 (d)-(f)).

(3) $\beta_4 \sim (\beta_3, \gamma)$ and the length of curves in this homotopy is bounded from above by $8\sqrt{A}$, (see Figure 3 (f)-(h)).

(4) Finally, $(\beta_3, \gamma) \sim (p_3, p_4) \sim (p_1, p_2)$, and the length of curves in this homotopy is bounded by $8\sqrt{A}$ as well. (This last step is not shown in the figure, but it is easy enough to see).

It remains to check that the homotopy class of f is a non-trivial element of $\pi_1(Z_1(M, Z))$. Again we can either use the Almgren isomorphism between $\pi_1(Z_1(M, Z))$ and $H_2(M, Z)$ ([1]) or a much simpler "hand-made" version of the correspondence between loops and discs in Γ on one side and 2-cycles and 3-chains in M on the other (described at the end of the proof of Theorem 1 above). In both cases everything immediately boils down to proving that the 2-cycle in M obtained as the sum of all singular simplices corresponding to homotopies used to construct f has a non-trivial homology class. This class is represented by a map of S^2 into M in the obvious way: In the case (2) we divide S^2 into three 2-cells that are mapped into $\Omega_x, \Omega_y, \Omega$ using the homotopies contracting the boundaries of these domains inside these domains introduced above (see Fig. 2 (c),(e)). In the case (3) we can decompose S^2 into four 2-cells (or simplices) and map them into $\Omega_1, \Omega_2, \Omega_3, \Omega$ using the homotopies contracting boundaries of these domains inside these domains introduced above.

The degree of this map is equal to one. This can be immediately seen from the fact that Ω_y is covered only by the image of the homotopy contracting γ_y in the case (2) or γ in the case (3) inside Ω_y. This homotopy clearly corresponds to the map of $(D^2, \partial D^2)$ into itself of degree one. So, taking the composite of our map of S^2 into M with the quotient map of M into S^2 that sends the boundary and the complement of Ω_y in M to a point we obtain a map of degree one. Therefore the homology class of interest for us is non-zero. QED.

Proof of Croke's lemma: Now let us sketch the proof of the result from [2] used in the proof of Theorem 2. Assume that $d > 2\sqrt{A}$ and $l(M) > 2\sqrt{A}$. Consider any of the shortest geodesics $\alpha(t), t \in [0, d]$ connecting x and y, parametrized proportionally to the arclength. Let t_0 be any number in the interval $(\sqrt{A}, d - \sqrt{A})$.

First, one proves that there exists a simple closed geodesic loop based at t_0 and dividing M into two domains one of which contains x and the other contains y. In order to see this consider the metric spheres $S(x,t)$ in M of radius t centered at x for all $t \in [t_0 - \sqrt{A}, t_0 + \sqrt{A}]$. For any generic t there exists a simple closed curve $\sigma \subset S(x,t)$ that intersects α transverally at $\alpha(t)$. It has only one point of intersection with α, namely $\alpha(t)$. Comparing the obvious inequality $A > \int_{t_0-\sqrt{A}}^{t_0+\sqrt{A}} length(S(x,t))$ dt and the equality $\int_{t_0-\sqrt{A}}^{t_0+\sqrt{A}} 2\sqrt{A} - 2|t - t_0| dt = 2A$ we see that there exists a generic $t \in [t_0 - \sqrt{A}, t_0 + \sqrt{A}]$ such that the length of $S(x,t)$ is less than $2\sqrt{A} - 2|t - t_0|$. Obviously, the same inequality is true for σ. σ divides M into two domains one of which contains x and another contains y. Now consider a loop based at $\alpha(t_0)$ that goes to $\alpha(t)$ along α then goes along σ and then returns to $\alpha(t_0)$ along α. The length of this closed curve is less then $2\sqrt{A}$ and this curve is not contractible in $M \setminus \{x,y\}$ (since σ is not contractible in $M \setminus \{x,y\}$. Therefore the length of the shortest closed curve passing through $\alpha(t_0)$ and not contractible in $M \setminus \{x,y\}$ also does not exceed $2\sqrt{A}$. It is easy to see that this shortest curve exists and is a geodesic loop based at $\gamma(t_0)$. (However this geodesic loop need not be unique.) Denote one of these geodesic loops by ρ_{t_0}. It is easy to prove that ρ_{t_0} is a simple curve dividing M into two domains one of which, M_x, contains x and another, M_y, contains y.

Now observe that since ρ_{t_0} is a geodesic at all points but $\alpha(t_0)$, it is convex to one of these domains. (This domain will have the property that the angle between two branches of ρ_{t_0} at $\alpha(t_0)$ measured inside this domain is less than π .) Following [2] denote by S_x the subset of $(\sqrt{A}, d-\sqrt{A})$ formed by t_0 such that ρ_{t_0} is convex to M_x . Similarly define S_y . S_x and S_y are closed subsets of $(\sqrt{A}, d-\sqrt{A})$. If both these sets are non-empty then their intersection is non-empty. Therefore in this case there exists $t_0 \in (\sqrt{A}, d-\sqrt{A})$ and two geodesic loops γ_x , γ_y based at $\alpha(t_0)$ of (equal) length less than $2\sqrt{A}$ as in (2) above. If one of these two sets, say S_x , is empty, then for any $t_0 \in (\sqrt{A}, d-\sqrt{A})$ ρ_{t_0} will be convex to M_y . It is easy to see that when $t_0 \longrightarrow \sqrt{A}^+$ a subsequence of ρ_{t_0} converges to a geodesic loop γ of length $\leq 2\sqrt{A}$ based at $\alpha(\sqrt{A})$, and we find ourselves in the situation of (3). (The only somewhat non-trivial fact that should be checked here is that even if the length of γ is equal to $2\sqrt{A}$, then γ cannot pass through x . Indeed, if γ passes through x in this case then it consists of two minimizing geodesics between $\alpha(\sqrt{A})$ and x such that the angle between them at x is equal to π . But this cannot happen because since α minimizes past $\alpha(\sqrt{A})$, the segment of α between x and $\alpha(\sqrt{A})$ is the unique minimizing geodesic between these points.)

Remark. After a previous version of the paper was finished we were informed by Chris Croke that similar results were recently also obtained by Stephane Sabourau in the preprint [7] that had not been yet submitted for publication.

Acknowledgements. The paper was partially written during the visit of the authors to IHES in Summer of 2001. The authors would like to thank IHES for its kind hospitality. Both authors gratefully acknowledge a partial support by NSERC Research Grants during their work on the present paper. Finally, we would like to thank the anonymous referee for his comments that helped to improve the exposition.

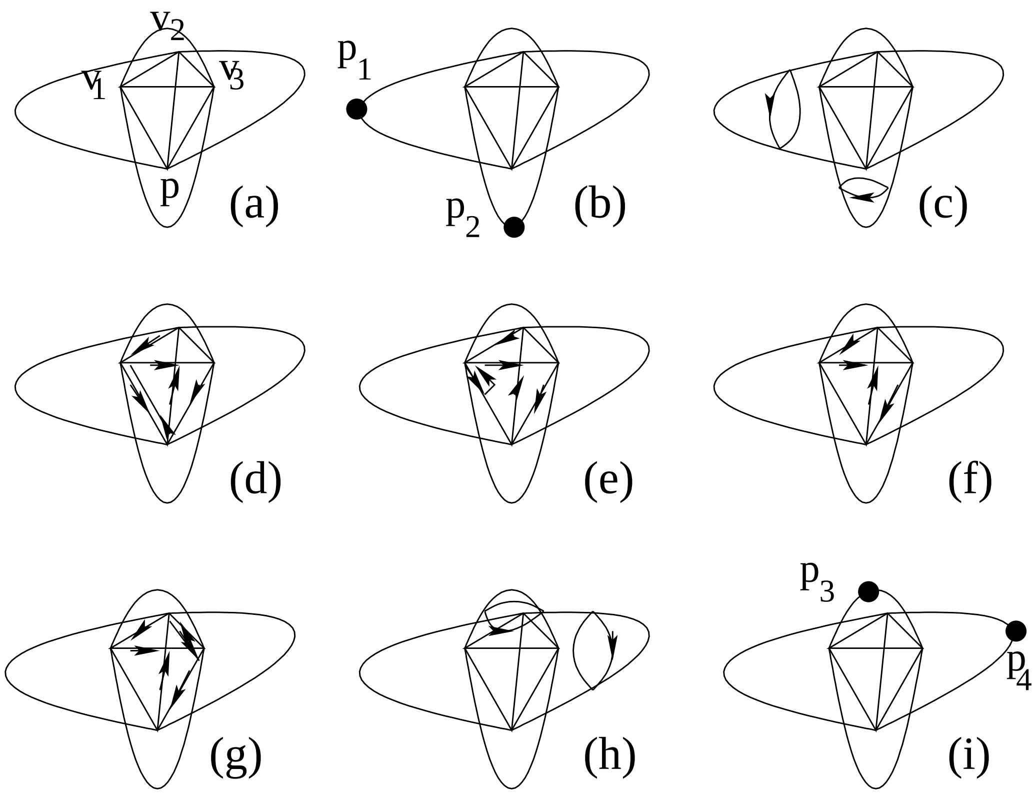

Figure 1

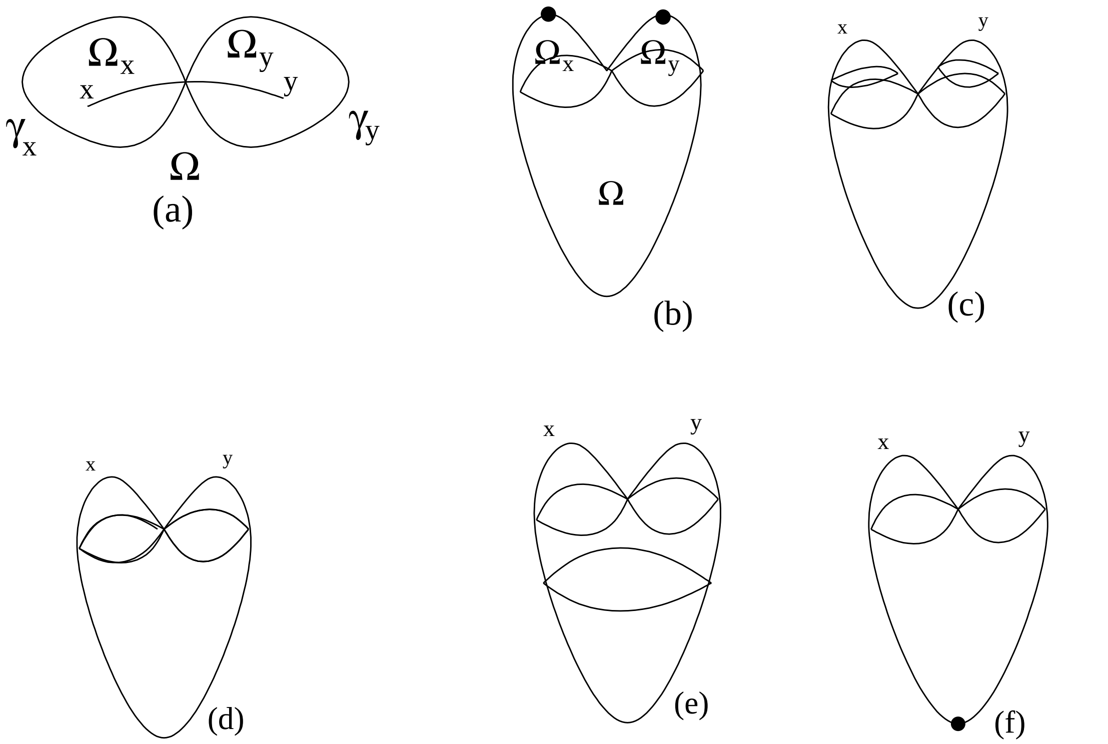

Figure 2

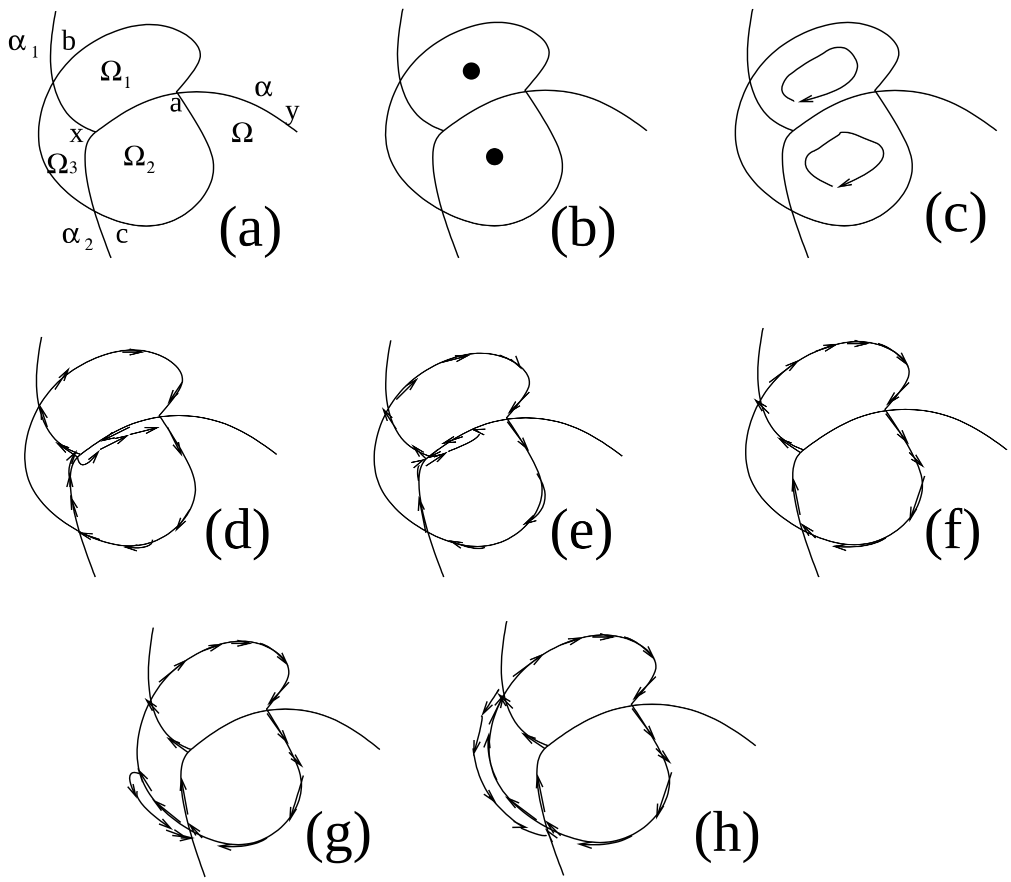

Figure 3

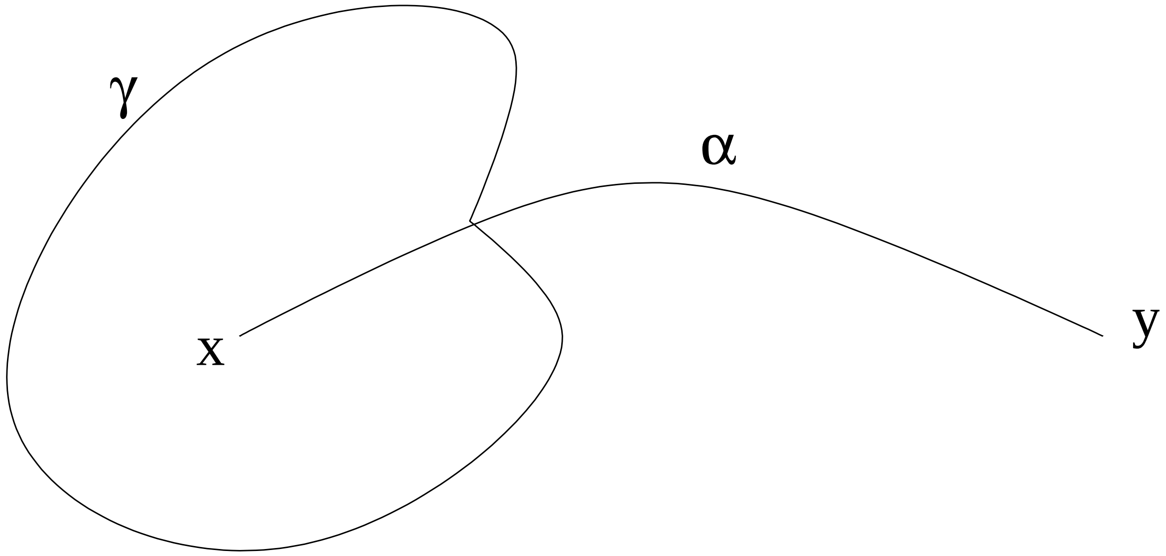

Figure 4

References

[1] F. Almgren, Jr., *The homotopy groups of the integral cycle groups*, Topology **1** (1962).

[2] C.B. Croke, *Area and the length of the shortest closed geodesic*, J. Diff. Geom. **27** (1988), 1-21.

[3] E. Calabi and J. Cao, *Simple closed geodesics on convex surfaces*, J. Diff Geometry **36** (1992), 517-549.

[4] M. Maeda, *The length of a closed geodesic on a compact surface*, Kyushu J. Math. **48** (1994, no.1), 9-18.

[5] A. Nabutovsky and R. Rotman, *Volume, diameter and the minimal mass of a stationary 1-cycle*, preprint, arXiv:math.DG/0201269 v2 , available at http://arXiv.org/abs/math.DG/0201269.

[6] J. Pitts, *Existence and regularity of minimal surfaces on Riemannian manifolds*, Ann. Math. Studies **27** (1981), Princeton University Press.

[7] S. Sabourau, *Filling radius and short closed geodesics of the sphere*, preprint.

Dept. of Mathematics, University of Toronto, Toronto, Ont. Canada M5S 3G3

E-mail address: alex@math.toronto.edu and rina@math.toronto.edu