PROCEEDINGS OF THE
AMERICAN MATHEMATICAL SOCIETY
Volume 134, Number 10, October 2006, Pages 3041–3047
S 0002-9939(06)08297-9
Article electronically published on April 10, 2006

THE LENGTH OF A SHORTEST CLOSED GEODESIC AND THE AREA OF A 2-DIMENSIONAL SPHERE

R. ROTMAN

(Communicated by Jon G. Wolfson)

ABSTRACT. Let M be a Riemannian manifold homeomorphic to S^2. The purpose of this paper is to establish the new inequality for the length of a shortest closed geodesic, $l(M)$, in terms of the area A of M. This result improves previously known inequalities by C.B. Croke (1988), by A. Nabutovsky and the author (2002) and by S. Sabourau (2004).

Let $l(M)$ denote the length of a shortest closed non-trivial geodesic on a closed Riemannian manifold M and let A be the area of M. In this paper we will prove the following theorem.

Theorem 0.1. *Let M be a manifold diffeomorphic to the 2-dimensional sphere. Then $l(M) \leq 4\sqrt{2}\sqrt{A}$.*

The first upper bounds for the length of a shortest closed geodesic on a 2-dimensional sphere were found by C.B. Croke (see [2]). In his paper Croke found estimates both in terms of the diameter and in terms of the area of a 2-dimensional sphere. Those results were later improved in [5] and in [9]. In particular, Croke proved that $l(M) \leq 31\sqrt{A}$, Sabourau proved that $l(M) \leq 12\sqrt{A}$, and A. Nabutovsky and the author proved that $l(M) \leq 8\sqrt{A}$, which was the best known estimate prior to this paper.

The estimate that we will obtain in this paper is not sharp. We are aware of only three sharp bounds for the length of a shortest closed geodesic in terms of the area. These results are due to K. Loewner, in the case when M is diffeomorphic to the 2-dimensional torus, to P. Pu in the case of M being diffeomorphic to RP^2 and to C. Bavard, in the case when M is diffeomorphic to the Klein bottle (see [3]).

It was suggested by E. Calabi and C.B. Croke that the sharp bound for the length of a shortest closed geodesic on a manifold diffeomorphic to S^2 should be $l(M) < (12)^{\frac{1}{4}}\sqrt{A}$. (It is not hard to see that the optimal constant is at least $12^{\frac{1}{4}}$; see [2].)

The existence of a closed geodesic on an arbitrary closed Riemannian manifold was proven by L. Lusternik and A. Fet (see [1]). Their proof uses Morse theory on the space ΛM of closed curves on M. One approach to estimating the length of a shortest closed geodesic can be interpreted as effectivization of this existence theorem. In particular, this approach is used in papers [2], [6].

Received by the editors February 24, 2005 and, in revised form, April 14, 2005.
2000 *Mathematics Subject Classification.* Primary 53C22; Secondary 58E10.
Key words and phrases. Geometric inequalities, closed geodesics.

The classical proof of the Lusternik and Fet theorem suggests that in order to estimate the length of a shortest closed geodesic it is enough to construct a non-contractible 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.

However, in this paper, as in [5], we will replace the space ΛM with the space Γ of parametrized 1-cycles made of two segments. In fact, this paper closely follows the proof of Theorem 2 of [5].

The space Γ was introduced by E. Calabi and J. Cao. It is defined as follows: $\Gamma = \{\Phi | \Phi = (\phi_1, \phi_2), \phi_i : [0,1] \longrightarrow S^2$ is a piecewise smooth path and $\phi_1(1) + \phi_2(1) - \phi_1(0) - \phi_2(0) = 0\}$; see ([4]). They defined the length functional on Γ by letting $L(\Phi) = L(\phi_1) + L(\phi_2)$. They further defined a distance on Γ that makes L into a continuous map.

The critical points of L on Γ are defined as in geometric measure theory: Let $\chi(M)$ denote the set of all smooth vector fields on M. Any $X \in \chi(M)$ generates a one-parameter group of diffeomorphisms h_t. The derivation of L at $\phi \in \Gamma$ in the direction of X is defined by $\delta L_\Phi(X) = \frac{d}{dt}(L(h_t \circ \Phi))|_{t=0}$. If $\delta L_\Phi(X) = 0$ for all $X \in \chi(M)$, then Φ is called a critical point of L on Γ and $L(\Phi)$ is called a critical value of L.

Following J. Pitts, E. Calabi and J. Cao observed that critical points of L in Γ, in the case when M is diffeomorphic to S^2, satisfy the property that both of the segments together form either a single closed geodesic, or a pair of closed geodesics, if we assume that both of the segments are smooth. Note, that without this assumption, the conclusion would be false.

Indeed, applying the first variational formula for the length functional, it is easy to see that any critical point of L on any Riemannian manifold is either a closed geodesic (possibly a constant geodesic), a pair of closed geodesics, or two geodesic loops emanating from the same point. In the last case the sum of four unit tangent vectors at the singular point coming out of this point must be zero. If the manifold is two-dimensional this implies that these two-geodesic loops form a self-intersecting closed geodesic.

Finally, Calabi and Cao note that if $f : S^1 \longrightarrow \Gamma$ is non-contractible, then there exists a closed geodesic on M of length $\leq \max_{t \in S^1} length(f(t))$ (see also Proposition 5 in [7] for a rigorous proof of this assertion).

So, the approach that we will use here is to construct a non-contractible $f : S^1 \longrightarrow \Gamma$ corresponding to the fundamental homology class of M and to estimate $\max_{t \in S^1} L(f(t))$. Using this approach, we have established in [5] that $l(M) \leq 4d$. We will use this result in this paper.

We will improve the technique used in [2] by Croke to obtain a bound for $l(M)$ in terms of the area of M. We will need the following definition from [2].

Definition 0.2. 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.

Lemma 2.2 in [2] asserts 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 stay in $\bar{\Omega}$, or to be more exact, they satisfy the following two conditions:

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

(2) γ_t is simple and convex to $\Omega_t = \Omega - \{x \in \gamma_s | 0 \leq s \leq t\}$.

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

We are now ready to prove Theorem 0.1.

Proof. The notation that will be encountered in this proof is borrowed from the proof of Theorem 2 in [5]. Let us begin by assuming that $l(M) > 3\sqrt{2A}$ and that the diameter d of M is greater than $\sqrt{2A}$. Otherwise, $l(M) \leq 4d \leq 4\sqrt{2A}$ (Theorem 1 in [5]).

As in [5], in the proof of this theorem we use ideas of Croke's proof of Theorem 4.2 in [2].

Step 1. Let $x, y \in M$ be two points that are a diameter apart, that is, $d(x, y) = d$. Let $\alpha(t)$ be a minimal geodesic joining x and y parametrized proportionally to its arclength. Then either

(1) For some $t \in (\frac{\sqrt{2A}}{2}, d - \frac{\sqrt{2A}}{2})$ there exist two geodesic loops γ_x and γ_y of length $\leq \sqrt{2A}$ based at $\alpha(t)$ and intersecting only at that point, 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 $\gamma_x \cup -\gamma_y$ is convex to the complement of $\Omega_x \cup \Omega_y$ in M; see Figure 1(a), (b)).

or (2) There exists a simple closed geodesic loop γ of length $\leq \sqrt{2A}$ not contractible in $M - \{x, y\}$ 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 $\alpha(\frac{\sqrt{2A}}{2})$ and is convex to Ω_y, or γ is based at $\alpha(d - \frac{\sqrt{2A}}{2})$ 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 2(a).

The following proof of the above statement is an improvement of the proof of the similar fact in [2].

Let t_0 be any number in the interval $(\frac{\sqrt{2A}}{2}, d - \frac{\sqrt{2A}}{2})$.

We will begin by showing that there exists a simple closed geodesic loop based at $\alpha(t_0)$ that divides M into two domains, one containing x and the other containing y. In order to prove this take the geodesic spheres $S(x, t)$ in M centered at x of radius t for all $t \in [t_0 - \frac{\sqrt{2A}}{2}, t_0 + \frac{\sqrt{2A}}{2}]$.

Note that for any generic t there exists a closed curve $\sigma \subset S(x, t)$ with no self-intersections that intersects α transversally at $\alpha(t)$, and this intersection with α is unique. Now, compare the inequality $\int_{t_0 - \frac{\sqrt{2A}}{2}}^{t_0 + \frac{\sqrt{2A}}{2}} length(S(x, t))dt < A$ and the equality $\int_{t_0 - \frac{\sqrt{2A}}{2}}^{t_0 + \frac{\sqrt{2A}}{2}} \sqrt{2A} - 2|t - t_0| dt = A$. They imply that there exists a generic $t_* \in [t_0 - \frac{\sqrt{2A}}{2}, t_0 + \frac{\sqrt{2A}}{2}]$ such that the length of $S(x, t)$ is less than $\sqrt{2A} - 2|t - t_0|$. The same inequality will hold for σ, which, by Jordan Separation theorem, divides M into two domains. Moreover, one of them will contain x and another will contain y. Next consider a loop based at $\alpha(t_0)$ that goes to $\alpha(t_*)$ along α, then along σ, and finally returns to $\alpha(t_0)$ along α. This curve is not contractible in $M \setminus \{x, y\}$ (see the proof of Lemma 3.2(1) in [2]), and it has length smaller than $\sqrt{2A}$.

This implies that the length of a shortest closed curve passing through $\alpha(t_0)$ and not contractible in $M \setminus \{x, y\}$ is also less than $\sqrt{2A}$. This shortest curve exists (though it does not have to be unique), and is in fact a geodesic loop based at $\gamma(t_0)$. If ρ_{t_0} is such a geodesic loop, it is a simple curve dividing two domains M_x and M_y, such that $x \in M_x$ and $y \in M_y$. The fact that this curve is the shortest among curves satisfying the above properties implies that it has no self intersections, which in turn implies that it divides M into two domains (see also the proof of Lemma 3.3 in [2]). Such a curve will also have the following property: the two tangents at its endpoints will lie on opposite sides of the tangent to α, otherwise could homotop this curve into $M - \alpha([0,1])$ and contract it there (see the proof of Lemma 3.3(2) in [2]).

Note that, being a geodesic loop, ρ_{t_0} is convex to one of these domains. Following [2] let S_x denote the subset of $(\frac{\sqrt{2A}}{2}, d - \frac{\sqrt{2A}}{2})$ formed by t_0 such that there exists such a ρ_{t_0} which is convex to M_x. One can similarly define S_y. Both S_x and S_y are closed subsets of $(\frac{\sqrt{2A}}{2}, d - \frac{\sqrt{2A}}{2})$. If both of these sets are non-empty, then their intersection is non-empty, and thus there exists $t_0 \in (\frac{\sqrt{2A}}{2}, d - \frac{\sqrt{2A}}{2})$ and two geodesic loops γ_x, γ_y based at $\alpha(t_0)$ of equal length that is smaller than $\sqrt{2A}$ as in case (1) above. Note that γ_x and γ_y will intersect only at their base point (see the proof of Lemma 3.3(3) in [2]). If one of these two sets, for example, S_x, is empty, then for any $t_0 \in (\frac{\sqrt{2A}}{2}, d - \frac{\sqrt{2A}}{2})$, ρ_{t_0} will be convex to M_y. It is easy to see that when $t_0 \longrightarrow \frac{\sqrt{2A}}{2}^+$ a subsequence of ρ_{t_0} converges to a geodesic loop γ of length $\leq \sqrt{2A}$ based at $\alpha(\frac{\sqrt{2A}}{2})$, as in case (2). (One needs to be a little careful about the case when the length of γ is exactly $\sqrt{2A}$. Indeed, if γ passes through x it consists of two minimizing geodesics between $\alpha(\frac{\sqrt{2A}}{2})$ and x such that the angle between them at x is equal to π. But this cannot happen because since α minimizes past $\alpha(\frac{\sqrt{2A}}{2})$, the segment of α between x and $\alpha(\frac{\sqrt{2A}}{2})$ is the unique minimizing geodesic between these points.)

Step 2. We are now ready to finish the proof of this theorem. Let us first consider case (1). We have subdivided the manifold into three domains: Ω_x, Ω_y and Ω, which denote the complement of $\Omega_x \cup \Omega_y$. We have assumed that $l(M) > 3\sqrt{2A} > \sqrt{2A}$. Therefore, γ_x is contractible in Ω_x to $p_x \in \Omega_x$ along the curves γ_{x_t}, γ_y is contractible in Ω_y to a point $p_y \in \Omega_y$ along the curves γ_{y_t} of length $\leq \sqrt{2A}$, and $\gamma_x \cup -\gamma_y$ is contractible to p in Ω along the curves γ_t of length $\leq 2\sqrt{2A}$, since this curve is convex to Ω.

We can now construct $f : S^1 \longrightarrow \Gamma$, so that S^1 passes through pairs of curves of length $\leq 2\sqrt{2A}$ as shown on Figure 1: We begin with pairs of constant curves $(p_x \cup p_y)$ (Figure 1(b)), which is homotopic to a pair of curves $(\gamma_{x_t} \cup -\gamma_{y_t})$ (Figure 1(c)). These maps are homotopic to the "figure 8" $(\gamma_x \cup -\gamma_y)$ (Figure 1(d)), which we can finally connect with p (Figure 1(f)) along the curves γ_t (Figure 1(e)). Note that $p \sim (p_x \cup p_y)$, and, thus, we have formed a loop in the space of the parametrized 1-cycles. Moreover, this loop is not contractible (the proof of Theorem 2 in [5] contains the proof of a similar statement).

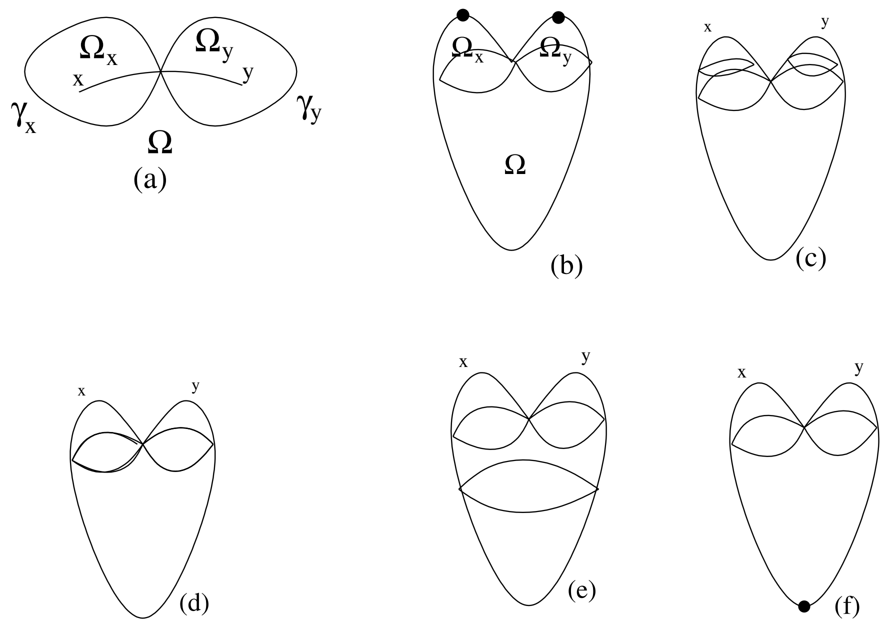

FIGURE 1.

One can calculate that the total length of curves in the homotopy is bounded from above by $2\sqrt{2A}$.

In case (2), assume without loss of generality that there exists a simple closed geodesic loop γ of length $\leq \sqrt{2A}$ that is not contractible in $M - \{x, y\}$ based at $\alpha(\frac{\sqrt{2A}}{2})$ and that is convex to Ω_y.

Following [2] we use Berger's lemma. It implies the existence of 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 $\alpha_1 = \alpha_2$.)

Let a, b, c denote the unique points of intersection of γ with $\alpha, \alpha_1, \alpha_2$, respectively. (Having more than one point of intersection would contradict the minimality of γ; see the proof of Theorem 4.2 in [2]). It follows that (1) the length of a geodesic segment joining x and a and denoted α_{xa} equals $\frac{\sqrt{2A}}{2}$ by assumption; (2) the length of a geodesic segment joining x and b and denoted α_{xb} is $\leq \sqrt{2A}$; (3) the length of a geodesic segment α_{xc} that connects x and c is $\leq \sqrt{2A}$.

Let us further denote the segments of γ that connect a and c, c and b and b and a as γ_{ac}, γ_{cb} and γ_{ba}, respectively.

Since $l(M) > 3\sqrt{2A}$ and since the closed curve $\beta_1 = \alpha_{xa} \cup -\gamma_{ba} \cup -\alpha_{xb}$ is convex to Ω_1, it is contractible in Ω_1 to a point p_1 along the curves β_{1_t}. Its length does not increase during this homotopy.

Similarly, $\beta_2 = \alpha_{xa} \cup \gamma_{ac} \cup -\alpha_{xc}$ is contractible in Ω_2 to a point p_2 along the curves β_{2_t}, $\beta_3 = -\alpha_{xc} \cup \alpha_{xb} \cup -\gamma_{cb}$ is contractible in Ω_3 to a point p_3 along the curves β_{3_t}, and γ is contractible in the complement of $\Omega_1 \cup \Omega_2 \cup \Omega_3$ along the curves γ_t without the length increase (see Figure 2(a)).

We are now ready to construct $f : S^1 \longrightarrow \Gamma$:

(1) $(p_1 \cup p_2) \sim (-\beta_1 \cup \beta_2)$ (see Figure 2(b)-(d)).

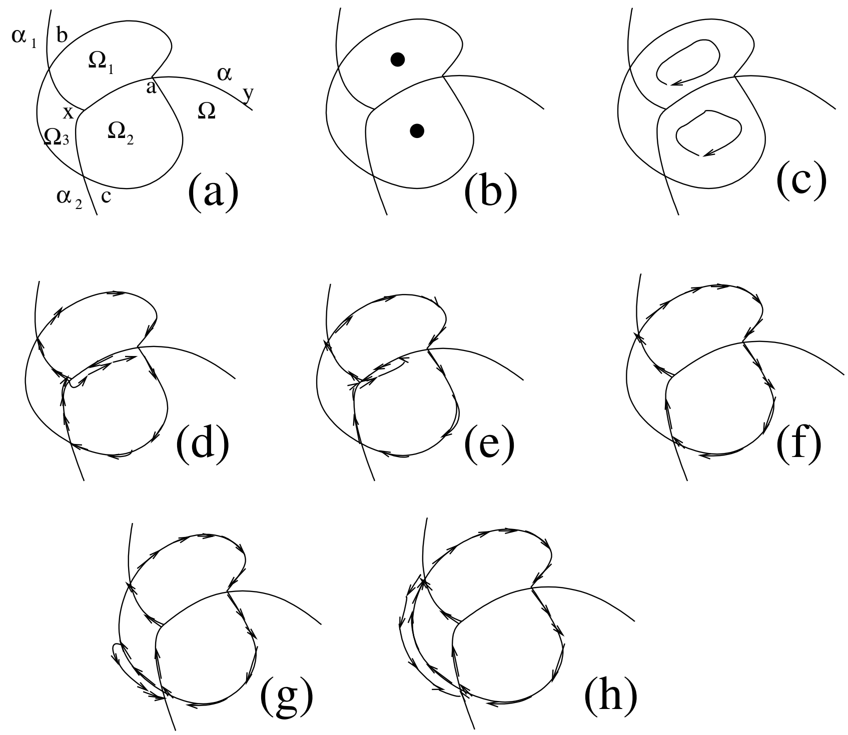

FIGURE 2.

(2) $(-\beta_1 \cup \beta_2) \sim (\alpha_{xb} \cup \gamma_{ba} \cup \gamma_{ac} \cup -\alpha_{xc}) = \beta_4$ (see Figure 2(d)-(f)).
(3) $\beta_4 \sim (\beta_3 \cup \gamma)$ (see Figure 2(f)-(h)).
(4) Finally, $(\beta_3 \cup \gamma) \sim (p_3 \cup p_4) \sim (p_1 \cup p_2)$.
Note that the length of curves during this homotopy is $\leq 4\sqrt{2A}$.

It remains to check that this homotopy corresponds to a non-contractible map of S^1 to the space Γ. This argument is identical to the similar argument in [5]. $\square$

Acknowledgements

The author thanks the referee for his helpful comments.

References

1. R. Bott, *Lectures on Morse theory, old and new*, Bull. Am. Math. Soc., **7** (1982), 331-358. MR0663786 (84m:58026a)
2. C.B. Croke, *Area and the length of the shortest closed geodesic*, J. Diff. Geom., **27** (1988), 1-21. MR0918453 (89a:53050)
3. C. B. Croke and M. Katz, *Universal volume bounds in Riemannian manifolds*, Surveys in differential geometry, **VIII** (2002), 109-137. MR2039987 (2005d:53061)
4. E. Calabi and J. Cao, *Simple closed geodesics on convex surfaces*, J. Diff Geom., **36** (1992), 517-549. MR1189495 (93h:53039)
5. A. Nabutovsky and R. Rotman, *The length of a shortest closed geodesic on a 2-dimensional sphere*, IMRN, **2002** (2002), no. 39, 2121-2129.
6. A. Nabutovsky and R. Rotman, *Upper bounds on the length of a shortest closed geodesic and quantitative Hurewicz theorem*, J. Eur. Math. Soc., **5** (2003), 203-244. MR2002213 (2004f:53043)

7. A. Nabutovsky and R. Rotman, *Volume, diameter and the minimal mass of a stationary 1-cycle*, GAFA, **14** (2004), 748-790. MR2084979 (2005g:53069)
8. J. Pitts, *Existence and regularity of minimal surfaces on Riemannian manifolds*, Ann. Math. Studies, **27** (1981), Princeton University Press. MR0626027 (83e:49079)
9. S. Sabourau, *Filling radius and short closed geodesics of the 2-sphere*, Bull. Soc. Math. France, **132** (2004), 105-136. MR2075918 (2005g:53065)

Department of Mathematics, Pennsylvania State University, University Park, Pennsylvania 16802 – and – Department of Mathematics, University of Toronto, Toronto, Ontario, Canada M5S 3G3

E-mail address: `rotman@math.psu.edu`

E-mail address: `rina@math.toronto.edu`