Geometriae Dedicata (2005) 113: 243–254
DOI 10.1007/s10711-005-5177-1

The Minimal Length of a Closed Geodesic Net on a Riemannian Manifold with a Nontrivial Second Homology Group

ALEXANDER NABUTOVSKY and REGINA ROTMAN
Department of Mathematics, The Pennsylvania State University, University Park, PA 16802, U.S.A. e-mails: nabutov@math.psu.edu;rotman@math.psu.edu

(Received: 10 December 2004; accepted in final form: 8 April 2005)

Abstract. Let M^n be a closed Riemannian manifold with a nontrivial second homology group. In this paper we prove that there exists a geodesic net on M^n of length at most 3 diameter(M^n). Moreover, this geodesic net is either a closed geodesic, consists of two geodesic loops emanating from the same point, or consists of three geodesic segments between the same endpoints. Geodesic nets can be viewed as the critical points of the length functional on the space of graphs immersed into a Riemannian manifold. One can also consider other natural functionals on the same space, in particular, the maximal length of an edge. We prove that either there exists a closed geodesic of length $\leqslant 2\,\mathrm{diameter}(M^n)$, or there exists a critical point of this functional on the space of immersed θ-graphs such that the value of the functional does not exceed the diameter of M^n. If $n=2$, then this critical θ-graph is not only immersed but embedded.

Mathematics Subject Classifications (2000). 53C23, 49Q10.

Key words. closed geodesics, geodesic nets, Riemannian manifold, surfaces.

1. Main Results

Let M^n be a Riemannian manifold. Define (stationary) *geodesic nets* on M^n as finite graphs immersed in M^n so that each edge is an immersed geodesic segment and for each vertex $p \in M^n$ the following *stationarity condition* holds: the sum of unit vectors tangent to all edges (=geodesic segments) meeting at p and diverging from p equals to zero. The last condition ensures that the embedded graph is a stationary point for the length functional on the space of embedded graphs. More precisely, let v be a geodesic net on M^n, let X be a smooth vector field on M^n, let $\Phi_X(t)$ denote the corresponding 1-parametric family of diffeomorphisms of M^n, and $l_{X,v}(t) = \mathrm{length}(\Phi_X(t)(v))$. Then the formula for the first variation, of the length functional implies that $\mathrm{d}l_{X,v}/\mathrm{d}t(0) = 0$. Obviously, each closed geodesic on M^n can be regarded as a geodesic net. However, a geodesic loop corresponds to a geodesic net if and only if it is a closed geodesic. Our definition is closely related to the definition of geodesic nets given in a paper of Hass and Morgan [2]. The only difference is that while Hass and Morgan require graphs to be

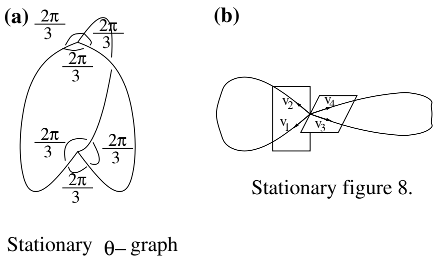

Figure 1. Geodesic nets.

embedded, not merely immersed, we allow self-intersections and intersections of edges at points different from the endpoints.

One of the simplest examples of geodesic nets that does not correspond to closed geodesics is a θ-graph shaped geodesic net. Namely, consider the graph with two vertices and three edges connecting them. Embbed it into M^n so that each edge becomes a geodesic segment. Denote the images of vertices in M^n by p_1 and p_2. The stationarity condition is equivalent to the following condition: for each $i=1,2$ the geodesic segments meeting at p_i lie in a two-dimensional plane, and all three angles between them are equal to $2\pi/3$ (see Figure 1(a)).

Hass and Morgan conjectured that if M^n is a closed convex surface in $\mathbf{R}^3$ then there exists a θ-graph shaped geodesic net on M^n. In [2] they proved this conjecture for all convex surfaces in R^3 sufficiently close to the standard sphere.

Another possible shape of a geodesic net is figure eight. Namely, consider two geodesic loops emanating from the same point $p \in M$ (see Figure 1(b)). The stationarity condition is equivalent to the following two conditions: (I) The angles formed by four vectors at p tangent to branches of these two loops are equal; (II) The bissectors of these two angles lie on the same straight line in the tangent space $T_p M$ and have opposite directions. It is easy to see that for $n=2$, where n is the dimension of M^n the geodesic net will be just a closed geodesic that has a self-intersection at p but if dim $M>2$ this need not necessarily be the case.

In [4] the authors studied the minimal length of a nontrivial closed geodesic on a Riemannian manifold diffeomorphic to S^2. Improving the constant in the previous results of Croke [1] and Maeda [3] the authors proved that when M^n is diffeomorphic to S^2 there exists a nontrivial closed geodesic on M^n of length not exceeding 4 diameter(M). (The same result was also independently proven by Sabourau [8]). Other curvature-free upper bounds for the length $l(M^2)$ of a shortest closed geodesic on a manifold diffeomorphic to S^2 can be found in [6,7]. The paper [6] contains upper bounds for $l(M^2)$ in terms of radii of metric balls covering M^2. In [7] one of the authors proved that $l(M^2) \leqslant 4\sqrt{2}\text{Area}(M^2)$ improving the constant in previous upper bounds for l in terms of the area obtained in papers [1,4,8]. In [5] the authors among other results proved that if M^n is a closed Riemannian manifold with a nontrivial second homology group then there exists either a closed geodesic or a figure eight shaped geodesic net in M^n of length not exceeding 4 diameter(M^n).

Our first result improves the constant 4 in this theorem (at the expense of introducing another possible type of the geodesic net).

THEOREM 1. *Let M^n be a closed manifold with a nontrivial second homology group. Then either there exists a closed geodesic of length* $\leqslant 3\text{diameter}(M^n)$ *on M^n, or there exists a geodesic net of length not exceeding* $3\text{diameter}(M^n)$ *represented by an immersed figure eight or a θ-graph. If $n=2$, and there is no nontrivial closed geodesic of length* $\leqslant 3\text{diameter}(M^2)$, *then there exists a geodesic net of length* $\leqslant 3\text{diameter}(M^2)$ *represented by an embedded θ-graph.*

Along with the length functional one can consider other natural functionals on the spaces of immersed graphs, for example, the functional F defined as the maximal length of an edge of the graph. We call an immersed θ-graph Γ *critical* if (a) No two edges of Γ form a closed geodesic; and (b) For each smooth vector field X on M^n the first variation of F in the direction of X at Γ is nonnegative. Note that, in principle, F can be not differentiable in the direction of X at Γ. Therefore condition (b) should be stated more rigorously. A rigorous way to state condition (b) is that for every smooth vector field X on

$$M^n \liminf_{t\to 0^+} \frac{F(\Phi_X^t(\Gamma))-F(\Gamma)}{t} \geqslant 0.$$

Here Φ_X^t, $(t\in\mathbf{R})$, denotes the 1-parametric flow of diffeomorphisms of M^n generated by X. In fact, it is not difficult to see that for all X, Γ there exists the right-derivative $\phi'_{X,\Gamma}(0^+)$, where $\phi_{X,\Gamma}(t)$ is defined as $F(\Phi_X^t(\Gamma))$. Now condition (b) is equivalent to the condition that for all smooth vector fields $X \phi'_{X,\Gamma}(0^+)\geqslant 0$. The following proposition provides an explicit characterization of critical θ-graphs:

PROPOSITION 2. *Assume that no two edges of an immersed θ-graph Γ in M^n form a closed geodesic. Then Γ is a critical θ-grpah if and only if the following three conditions hold*:

(1) *Each edge of Γ is a geodesic between its endpoints.*
(2) *For each of the two vertices of Γ vectors tangent to the three geodesic segments meeting at this vertex lie in a 2-dimensional plane in the corresponding tangent space. However, they do not belong to any half-plane of this 2-plane.*
(3) *All three edges of Γ have the same length.*
(4) *Each digon formed by two edges of Γ has equal angles at its two vertices.*

THEOREM 3. *Let M^n be a closed Riemannian manifold with a nontrivial second homology group. Then either there exists a nontrivial closed geodesic on M^n of length not exceeding* 2 diameter(M^n) *or there exists a critical θ-graph with edges of length*

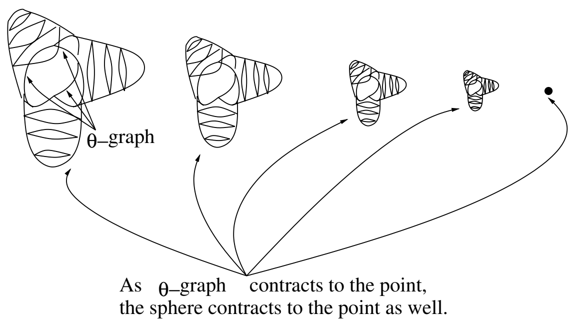

Figure 2. Illustration of the proof of Theorems 1 and 3.

$\leqslant$ diameter(M^n). *If* $n=2$, *and there is no nontrivial closed geodesic of length* $\leqslant$ 2 diameter(M^2) *on* M^2, *then there exists an embedded critical* θ*-graph with edges of length* $\leqslant$ diameter(M^2).

Note that it is a well-known fact that if M^n is not simply connected, then there exists a (noncontractible) closed geodesic on M^n of length $\leqslant 2\,\mathrm{diameter}(M^n)$. Therefore, we need to prove Theorems 1 and 3 only in the case, when M^n is simply-connected. Now the proofs of Theorems 1 and 3 boil down to the following idea: assume that there are no 'short' closed geodesics and 'short' θ-graphs in the manifold. Consider a cell subdivision of S^2 into three 2-cells, three 1-cells and two 0-cells by an embedded θ-graph. Consider a noncontractible map of the subdivided sphere into M^n such that each of 2-cells is mapped by contracting the image of the digon in its boundary without increasing the length using the Birkhoff curve shortening process (BCSP). Then we use an appropriate curve shortening process to contract the image of the 1-skeleton to a point and the rest of the sphere will follow by application of the BCSP. That is, the homotopy between θ-graph and a point will, in the lack of geodesics, extend to a homotopy of the map of the whole sphere, thus providing us with a contradiction, (see Figure 2). Of course, here we use the continuity of the BCSP (with respect to the initial closed curve) in the absence of 'short' closed geodesics.

Note that Theorems 1 and 3 are new even in the case when M is a convex surface in $\mathbf{R}^3$.

2. Proof of Proposition 2

Consider an immersed graph in M^n and a smooth vector field X. Assume that all edges of the graph are immersed as geodesic segments. The formula for the first variation of the arclength implies that the first variation of the length of an edge $e=[v_i v_j]$, ($v_i v_j \in M^n$ are the endpoints of e), is determined by the formula $-\langle X(v_i), e'(v_i)\rangle - \langle X(v_j), e'(v_j)\rangle$, where $e'(v_i), e'(v_j)$ denote the unit tangent vectors to e at v_i and v_j that are directed inwards. In particular, it depends only on the values of the vector field at v_i and v_j. Therefore, if we would like to investigate the first variation in the direction of a vector field of lengths of edges of an immersed θ-graph such that all its edges are realized by geodesics, then only the values of this vector field at the vertices matter.

Now we are going to prove the necessity of conditions (1)–(4). The necessity of condition (1) is obvious. The necessity of condition (2) is easy to prove: Assume that three vectors tangent to geodesic segments meeting at a vertex do not lie in any 2-plane. Then they lie in an open half-space of the three-dimensional subspace spanned by these three vectors. A small perturbation of the vertex in the direction of the vector normal to the boundary of this half-space and pointing inwards decreases the lengths of all three geodesic segments. Similarly, these three vectors cannot lie in a half-plane in this 2-plane since otherwise a small perturbation in the direction of the normal pointing inwards will be length-decreasing.

Also, if no two of the three segments form a closed geodesic then all three edges have the same length. Indeed, there is always a small perturbation of the θ-graph that lowers the lengths of any two chosen edges (but this perturbation can increase the length of the third edge). The same argument implies that no small perturbation of the θ-graph can lower the length of one edge keeping the lengths of other two edges constant. Otherwise, we will be able to make the lengths of all three edges smaller proceeding as follows: Let F_1 be a vector field such that the first variation of length of edge 1 and the first variation of length of edge 2 with respect to F_1 is 0, but the first variation of the length of edge 3 with respect to F_1 is negative. Let F_2 be a vector field such that the first variation of length of edge 1 and the first variation of length of edge 2 with respect to F_2 is negative, but the first variation of the length of edge 3 with respect to F_2 is, possibly, positive. Then for all sufficiently small positive ε the first variations of lengths of all three edges with respect to $F_1 + \varepsilon F_2$ will be negative.

Denote the vertices of the θ-graph by $v_1, v_2 \in M^n$. Now our strategy will be to look for $X \in TM^n_{v_1}$, $Y \in TM^n_{v_2}$ such that the first variation of lengths of two edges in the direction of a vector field extending X, Y is zero, and the first variation of the length of the third edge is negative. (As we previously noticed the first variation of the length of a geodesic with respect to a vector field depends only on its values at the endpoints. So, only X and Y matter.) Enumerate all three edges of the θ-graph by 1, 2, 3. Choose Y as a parallel translation of X along the geodesic corresponding to the edge 1. So, the variation of the length of edge 1 is zero for any choice of X. The variations of the length of two other edges now become functions of X. These functions are odd. Therefore, varying X we can find X_0 such that the variation of the length of edge 2 is zero. If the variation of the edge 3 is negative, we arrive at a contradiction. If it is positive, we replace X_0 by $-X_0$ and arrive at a contradiction. Therefore, it must be zero, and we obtain a constraint on six unit vectors tangent to three edges of the graph at v_1 and v_2. Now, we can obtain similar constraints for the remaining five ways to number the three edges of the graph by 1,2,3. An elementary analysis of these constraints shows that they are equivalent to equiangularity of each of the three digons formed by the edges of the graph.

By analyzing all possibilities for choosing X, Y it is not difficult to see that conditions 1–4 are also sufficient conditions for an immersed θ-graph being the local minimum of the considered functional.

3. Length-Decreasing Processes for θ-graphs

Choose one of two functionals considered in this paper: the total length or the maximal length of an edge. Denote by T_x the set of immersed θ-graphs in M^n, where the value of this functional does not exceed x. We would like to construct a flow on T_x that decreases the value of the given functional. We are going to use the following

ASSUMPTION A. *If the considered functional is the total length functional then we assume that there are no stationary θ-graphs, stationary double loops and closed geodesics of a positive length $\leqslant x$ in M^n. If the considered functional is the maximal length of an edge, then we are going to assume that there are no critical θ-graphs with edges of nonzero length $\leqslant x$ and nontrivial closed geodesics of length $\leqslant 2x$ in M^n.*

Formally speaking, we will be looking for a deformation $D : \bar{T}_x \times [0, 1] \to \bar{T}_x$, where $\bar{T}_x$ denotes the closure of T_x in the space of continuous maps of the θ-graph to M^n, such that for every Γ $D(\Gamma, 1) = \Gamma$, $\text{length}(D(\Gamma, 1)) = 0$, and the considered functional decreases along every trajectory $D(\Gamma, t), t \in [0, 1]$, of D. (The closure of T_x will also contain closed curves and points that can be regarded as degenerate θ-graphs.) The next lemma establishes the existence of such D if assumption A holds.

LEMMA 4. *If assumption A holds, then there exists a deformation retraction of the closure of T_x to M^n (regarded as the subset of the closure of T_x made of θ-graphs that degenerate to a point) such that the considered functional decreases along every trajectory of the deformation.*

Sketch of the proof. Here we are going to present only the basic ideas of the construction of the deformation retraction, since a detailed description can be found in [5]. (There we dealt with the case of the total length functional but defined on spaces of arbitrary immersed graphs. Some technical difficulties that arise in the general situation do not arise for θ-graphs. The same proof works for the functional defined as the maximal length of one of three edges of a θ-graph.)

The first step is to find a deformation retraction of the closure of T_x to a finite-dimensional space such that the length of every edge does not increase during the deformation. This space will be the subset of $\bar{T}_x$ formed by all elements such that each edge is mapped to a piecewise geodesic with geodesic segments of length $\leqslant \text{inj}(M^n)/4$, where $\text{inj}(M^n)$ denotes the injectivity radius of the ambient manifold. (We will be denoting this space by G_x.) Indeed, it is well known how to construct a length nonincreasing homotopy between a curve of length $\leqslant x$ and a piecewise geodesic connecting its endpoints and made of $[4x/\text{inj}(M^n)] + 1$ geodesic segments

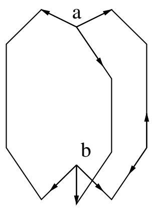

Figure 3. Length decreasing process for a θ-graph.

of length $< \mathrm{inj}\,(M^n)/4$ so that the homotopy will depend continuously on the initial curve. (Such a homotopy is the first step of the BCSP; cf. [1,5]). The space G_x can be regarded as a subset of $(M^n)^N$ for a sufficiently large $N = N(M^n, x)$.

If assumption A holds, then for every point $g \in G_x$ there exists a vector tangent to G_x such that the first variation of the considered functional in the direction of this vector is negative. Let us call this vector a *vector of a steep descent*. (See Figure 3 for the illustration of how to construct this vector for the total length functional.) The same will be true for every point p of G_x sufficiently close to g, if we consider the parallel translation of the choosen vector of a steep descent to p along the minimal geodesic from g to p in G_x. (This geodesic can be regarded as a finite collection of geodesics on M^n.) Then choosing an appropriate locally finite partition of unity we can construct a vector field on G_x such that the first variation of the functional in the direction of this field is negative, and G_x deforms to G_0 in a finite time. (This field will be called a *gradient-like vector field* below.)

Note, that for our purposes we do not need to know an explicit vector of a steep descent for the considered functionals on G_x. Yet, it is not difficult to construct such a vector. Every element of G_x has two *triple points*, where three edges of the graph meet, and a finite number of *double points*, which are, by definition, all endpoints of the geodesic segments that are not triple points. To define the vector of a steep descent for G_x we need to define its components at all double and triple points. (Each of these components is a tangent vector to M_n at the considered point.) For both functionals we define the component of the vector of steep descent at each double point as the sum of unit tangent vectors to two geodesic segments meeting at this point, and directed from it. For the total length functional we define the three unit tangent vectors to geodesic segments meeting at this point and directed from this point. In the case of the 'maximal length of an edge' functional we define the components of the vector of a steep descent at two triple points using the proof of Proposition 2: if one of conditions (2)–(4) in the text of Proposition 2 provides us with an explicit way to define the components of the vector of a steep descent at both triple points. If conditions (2)–(4) hold, then both these components are equal to zero.

There will be two technical difficulties worth mentioning here. First, we need to prove that the constructed deformation reaches G_0 in finite time. The difficulty here is that we constructed a gradient-like vector field on a (noncompact) set $G_x \backslash G_0$, so the finiteness of time does not immediately follow from a compactness argument. For the total length functional we checked the finiteness of time in [5], Sections 3.5 and A.8. It is very easy to verify that the same proof with obvious modifications works for the maximal edge length functional.

The second difficulty arises from the fact that even if the total length decreases, the length of individual geodesic segments can increase. We definitely do not want the distance between two points, that need to be connected by the minimizing geodesic, to become as large as the injectivity radius of M^n. Therefore we proceed as follows. We deform G_x using the constructed gradient-like flow for a 'safe' time t_*. This time is so small, that we can be sure that the length of the individual geodesic segments will stay $\leqslant \mathrm{inj}(M)/2$. Then we 'forget' that there edges between the triple points are already piecewise-geodesic, and replace them by broken geodesics with segments of length $\geqslant \mathrm{inj}\,(M^n)/4$ exactly as it was done at the beginning of the construction. Then we continue to deform G_x following the trajectories of the gradient-like field for the time t_*. Then we stop and replace all three edges by piecewise geodesics with shorter segments, etc. (cf. [5], Section 3.6). There is, however, one 'residual' inconvenience: since the length of individual geodesic segments can become $> \mathrm{inj}(M^n)/4$ we need to consider a space $\tilde{G}_x$ defined almost in the same way as G_x with the only one difference: the geodesic segments are allowed to have length $\leqslant \mathrm{inj}(M^n)/2$ instead of $\leqslant \mathrm{inj}(M^n)/4$. The space G_x will be deformed to G_0 inside a larger space $\tilde{G}_x$.

4. Proof of Theorems 1 and 3

As we mentioned before, it is sufficient to consider only the case, when M^n is simply-connected. (Otherwise, there exists a noncontractible closed geodesic of length $\leqslant$ 2 diameter(M^n.) Let δ be a positive number, which can be chosen as small as one wishes. Let d denote the diameter of M. In order to prove Theorem 1 we need to assume that $l(M^n) > 3d + \delta$ and in order to prove Theorem 3 assume that $l(M^n) > 2d + \delta$. Let f: $S^2 \rightarrow M^n$ be a noncontractable map. If $n = 2$ and M^2 is diffeomorphic to S^2, assume that f is a diffeomorphism. Let S^2 be triangulated into very fine simplices, so that the image of every 1-simplex has length $< \delta$. Let D^3 be triangulated as a cone over S^2. We will begin by extending the map f to the 2-skeleton of D^3. The procedure will be inductive to skeleta of D^3. We will begin with the 0-skeleton. To extend to 0-skeleton we will assign to the center of the disc p an arbitrary point $\tilde{p} \in M^n$. Next to extend to 1-skeleton we will assign to 1-simplex $[p, v_{i_1}]$ a minimal geodesic segment connecting $\tilde{p}$ and a point $\tilde{v}_{i_1} = f(v_{i_1})$ and denoted $[\tilde{p}, \tilde{v}_{i_1}]$. This segment has length of at most d. Finally, to extend to the 2-skeleton of D^3 consider an arbitrary 2-simplex $\sigma_i^2 = [p, v_{i_1}, v_{i_2}]$. Consider its boundary $\partial\sigma_i^2 = [v_{i_1}, v_{i_2}] - [p, v_{i_2}] + [p, v_{i_1}]$. It is mapped by f to a closed curve of length at most $2d+\delta$. By our assumption it can be contracted to a point without increasing the length using the BCSP (cf. [1] for a detailed description of the BSCP). We will define f on σ_i^2 using this BSCP homotopy regarded as a map of the 2-disc into M^n.

It is impossible to extend f to the 3-skeleton of D^3, since this would imply that the map is contractible. Therefore, there exists a 3-simplex $\sigma_i^3=[p, v_{i_1}, v_{i_2}, v_{i_3}]$ such that $f|_{\partial\sigma_i^3}$ is not contractible. Thus, we obtain a noncontractible sphere with three large 'tentacles' and one small one.

At the next stage we will try to deform the map of 1-skeleton of this sphere to a point.

We will begin by contracting the boundary of $[\tilde{v}_{i_1}, \tilde{v}_{i_2}, \tilde{v}_{i_3}]$ to a point. That is, we would like to get rid of the small 2-simplex and to replace the 1-skeleton of the considered 3-simplex in M^n by a θ-graph that lies very close to the orginial 1-skeleton. This step would have been entirely obvious, if not for the fact that in the case, when $n=2$, we would like to obtain a θ-graph with edges that intersect only at the endpoints. Withough any loss of generality we can assume that the distances between $\tilde{v}_{i_1}, \tilde{v}_{i_2}$ and $\tilde{v}_{i_3}$ do not exceed the convexity radius of M^n. First, we introduce a homotopy that transforms the sides of the triangle $v_{i_1}v_{i_2}v_{i_3}$ to the unique minimizing geodesics. Now there exists an obvious length decreasing homotopy h_t that contracts $\partial[\tilde{v}_{i_1}, \tilde{v}_{i_2}, \tilde{v}_{i_3}]$ to $\tilde{v}_{i_1}$ over the above small simplex: points on the segment $[v_{i_2}v_{i_3}]$ move with constant speeds to v_{i_1} along the corresponding minimizing geodesics connecting them with v_{i_1}. This homotopy can be extended to the remaining three one-dimensional simplices of the 1-skeleton of $\partial\sigma_i^3$ in the following way: let $\alpha_j=[\tilde{p}, \tilde{v}_{i_j}]$. We will let $h_t(\alpha_j)=\alpha_j+[\tilde{v}_{i_j}, h_t(\tilde{v}_{i_j})]$, where $[\tilde{v}_{i_j}, h_t(\tilde{v}_{i_j})]$ denotes the trajectory of a point $\tilde{v}_{i_j}$ under the homotopy h_t from the starting time to time$=t$. It is obvious, that the length of the image of each edge of the θ-graph after this homotopy does not exceed $d+\delta$. Also, note that we can choose to contract the small simplex to v_{i_2} or to v_{i_3} instead of v_{i_1}. This fact is important only, when $n=2$ (and we wish to obtain a θ-graph with edges that intersect only at their endpoints). Recall that in this case f is a diffeomorphism, and that two minimizing geodesics emanating from the same point can intersect only at their endpoints. Having this fact in mind, we are going to analyze, whether or not three edges of the θ-graph obtained at the end of this stage can intersect at points different from their common endpoints. First, the minimizing geodesic $[pv_{i_1}]$ can intersect the minimizing geodesics $[pv_{i_2}], [pv_{i_3}], [v_{i_1}v_{i_2}]$ and $[v_{i_1}v_{i_3}]$ only at p and v_{i_1}. Therefore $[pv_{i_1}]$ intersects $h_1([pv_{i_2}])=[pv_{i_2}]+[v_{i_2}v_{i_1}]$ only at p and v_{i_1}. The same is true for the intersection of $[pv_{i_1}]$ and $h_1([pv_{i_3}])$. Yet we did not exclude a possibility that $h_1([pv_{i_2}])$ intersects $h_1([pv_{i_3}])$ at a point different from p and v_{i_1} since $[pv_{i_2}]$ can interest $[v_{i_1}v_{i_3}]$, and $[pv_{i_3}]$ can intersect $[v_{i_1}v_{i_2}]$. However, these are the only possibilities of the undesirable intersections. Further, note that two minimizing geodesics on a Riemannian manifold (with possibly different endpoints) can have at most one point of intersection. Therefore, if, say, $[pv_{i_2}]$ intersects $[v_{i_1}v_{i_3}]$, then the point of intersection is unique, and $[pv_{i_2}]$ divides the triangle $v_{i_1}v_{i_2}v_{i_3}$ on the sphere into two halves. Now we see that in this case $[pv_{i_1}]$ cannot intersect $[v_{i_2}v_{i_3}]$ since otherwise it must also cross $[pv_{i_2}]$, which is impossible. Similarly, $[pv_{i_3}]$ cannot cross $[v_{i_1}v_{i_2}]$. Our argument demonstrates that there is at most one $l \in \{1, 2, 3\}$ such that $[pv_{i_l}]$ crosses the opposite side of the triangle $v_{i_1}v_{i_2}v_{i_3}$. Now we see that if $h_1([pv_2])$ intersects $h_1([pv_3])$ at a point different from p and v_{i_1}, then such an l exists and is equal to either 2 or 3. In this case we can choose to contract the boundary of the triangle $v_{i_1}v_{i_2}v_{i_3}$ to v_{i_1} instead of v_{i_1} (and then to extend the homotopy to the 1-skeleton of the considered 3-simplex as above). Denote this new homotopy by h. Now our analysis shows that $h([pv_{i_j}])$ and $h([pv_{i_k}])$ intersect only at their common endpoints p and v_{i_l}, if $j \neq k$.

At the second step we extend this homotopy to the 2-skeleton of $f(\partial\sigma_i^3)$. We do that by letting $h_t([\tilde{v}_{i_0}, \dots, \hat{\tilde{v}}_{i_j}, \dots, \tilde{v}_{i_3}])$ be the surface generated by the BCSP homotopy that connects its boundary to a point.

Now let S^2 denote $h_1(f(\partial\sigma_i^3))$. Its 1-skeleton is a θ-graph that consists of edges $h_1(\alpha_i), i = 1, 2, 3$. We will next use the deformation retraction from Lemma 4 to prove the Theorems. Indeed, under our assumptions this deformation will contract the θ-graph to a point. But, the above homotopies extend to all of S^2, since at each point we can contract the three individual digons of θ-graph to a point. Those homotopies depend continuously on the curve, unless there is a short closed geodesic. Thus, we have contracted f to a point deriving a contradiction. This contradiction proves Theorems 1 and 3 with the exception of the assertion that if $n = 2$, then one can ensure that the stationary (or critical) θ-graph is embedded.

Here is the sketch of the proof of this assertion. Note, that since all edges of a stationary (or critical) θ-graph are geodesics then all intersections at inner points (if such intersections exist at all) are transversal. Analyzing the proof of Lemma 4 we see that the stationary (or critical) θ-graph produced by our proof of Theorems 1 and 3 was, in fact, obtained from a θ-graph with three nonselfintersecting edges that intersect only at their endpoints by applying to it a flow that decreases the functional of interest (that is, the total length or the maximal length of an edge). (Recall that we obtained this θ-graph with edges that intersect only at their endpoints after we contracted a small triangle $v_{i_1}v_{i_2}v_{i_3}$ to a point as described above in our proof.) If there exist transversal intersections in the stationary (or critical) θ-graph, then they must appear in pairs bifurcating from a tangent intersection as

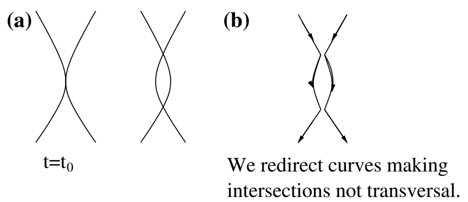

Figure 4. A modification of the length-decreasing homotopy.

shown on Figure 4(a) below. But for each appearance of a pair of such points at $t = t_0$ we can alter the homotopy of the θ-graph determined by the flow at times very close to t_0 and greater than t_0 to avoid transveral intersections so that the length of every edge at every moment will not exceed its value before the alteration. The idea of this alteration is to swap the segments between the intersection points as shown Figure 4(b). Then we just restart the flow constructed in the proof of Lemma 4. We can similarly treat self-intersections of geodesic segments. It is not difficult to see that this simple trick ensures that we can get stuck only at an embedded stationary (or critical) θ-graph. More formally, we will need the finiteness of the set of points of intersection of the segments formed during the flow. This finiteness is trivial, if the Riemannian metric on the surface is analytic. But the general case (of a C^∞-smooth Riemannian metric) easily follows from the analytic case using an analytic approximation of the Riemannian metric.

Remark. Note that we did not use the full strength of the conclusion of Lemma 4 in the proof of Theorems 1, 3. We needed only the connectedness of T_x (instead of the fact that there is a deformation retraction of T_x on $T_0 = M^n$). The proof of Lemma 4 easily implies that the connectedness of T_x follows from a weaker assumption: instead of demanding that there are no stationary (correspondingly, critical) θ-graphs as in assumption A, one can require only that there are no stationary (correspondingly, critical) θ-graphs providing a local minimum of the considered functional. (The idea is that when one descends from a point of T_x and gets stuck at a critical point in G_x which is not a local minimum, one can exit from this critical point in the direction, where the functional decreases, and then continue the descent as in the proof of Lemma 3 until one gets stuck at a critical point again, etc.) Therefore the conclusion of Theorem 1 (correspondingly, Theorem 3) can be strengthened as follows: Instead of the existence of a stationary (correspondingly, critical) θ-graph one can claim the existence of a stationary θ-graph providing a local minimum for the total length functional (correspondingly, a critical θ-graph providing a local minimum for the functional defined as the maximal length of an edge).

Acknowledgements

The research of A. Nabutovsky had been partially supported by NSF grant DMS-0405954. This work had been partially done during the visit of the authors to MSRI in April–May, 2004. The authors would like to thank MSRI for its warm hospitality. We are grateful to the referee for his comments that helped to improve the exposition.

References

1. Croke, C.B.: Area and the length of the shortest closed geodesic, *J. Differential Geom.* **27** (1988), 1–21.

2. Hass, J. and Morgan, F.: Geodesic nets on the 2-sphere, *Proc. Amer. Math. Soc.* **124** (1996), 3843–3850.
3. Maeda, M.: The length of a closed geodesic on a compact surface, *Kyushu J. Math.* **48**(1) (1994), 9–18.
4. Nabutovsky, A. and Rotman, R.: The length of the shortest closed geodesic ion a 2-dimensional sphere, *Internat. Math. Res. Notices* **23** (2002), 1211–1222.
5. Nabutovsky, A. and Rotman, R.: Volume, diameter and the minimal mass of a stationary 1-cycle, *Geom. Funct. Anal.* **14**(4) (2004), 748–790.
6. Rotman, R.: The length of a shortest closed geodesic on a 2-dimensional sphere and coverings by metric balls, *Geom. Dedicata* (to appear).
7. Rotman, R.: The length of a shortest closed geodesic and the area of a 2-dimensional sphere, to appear in *Proc. Amer. Math. Soc.*
8. Sabourau, S.: Filling radius and short closed geodesic on the 2-sphere, *Bull. Soc. Math. France* **132**(1) (2004), 105–136.