Linear bounds for lengths of geodesic segments on Riemannian 2-spheres

Alexander Nabutovsky and Regina Rotman

August 31, 2012

Abstract

In this paper we will show that for any pair of points $p, q \in M$, where M is a Riemannian manifold diffeomorphic to the 2-dimensional sphere, there always exist at least k geodesics connecting them of length at most $22kd$, where d is the diameter of M.

Introduction.

The paper is a continuation of our recent work [NR 3] in which we proved that on any manifold M diffeomorphic to S^2, there always exist at least k geodesic loops at p of length bounded above by $20kd$, where d is the diameter and p is an arbitrary point of M. It had been noticed in [NR3], that the proof in [NR3] cannot be directly generalized to the case, when one is interested in geodesics connecting p with $q \neq p$. Indeed, although our proof of the general case presented below uses many ideas from the proof of the case $p = q$ given in [NR3], it is significantly longer and more complicated than the proof for the case of geodesic loops.

The proof uses the existence theorem of J. P. Serre as a starting point. In 1951 he proved that given an arbitrary pair of points $p, q \in M^n$, where M^n is a closed Riemannian manifold M^n of dimension n, there exist infinitely many distinct geodesics connecting p and q, (see [Se]).

In fact, we will use the proof of this theorem, given by A. Schwartz in [Sch]. In particular, A Schwartz demonstrated that there exists a constant $c(M^n)$ depending on the Riemannian metric on M^n such that for every positive integer k there exist k distinct geodesics connecting p and q of length $\leq c(M^n)k$, but his proof gives little information about the dependence of $c(M^n)$ on the geometry of M^n.

In a series of papers, in particular in [NR 1], [NR 2], [NR 3], we have tried to answer the following question: Is there a function $f(n,k)$, such that for any pair of points $p,q \in M^n$ there exist at least k geodesics connecting these points of length at most $f(n,k)d$? The most general answer to this question was given in our paper [NR 2]. In this paper we have demonstrated that the statment is true if one takes $f(n,k) = 4nk^2d$.

The next natural question is whether there is a similar bound that is linear in k.

Theorem 0.1 *Let M be a smooth Riemannian manifold of diameter d diffeomorphic to S^2. Then for each pair of points $p,q \in M$ there exist at least k geodesics connecting the points p and q of length at most $(22k-22)d + dist(p,q) \leq (22k-21)d < 22kd$.*

Note that Theorem 0.1 implies the same upper bound for all closed Riemannian surfaces, and not only those surfaces that are diffeomorphic to S^2. Indeed, if a surface is diffeomorphic to RP^2, we can consider its double covering, and apply Theorem 1.1 to construct many short geodesics between a lifting of p and both liftings of q, which will project into distinct short geodesics between p and q. If a surface has an infinite fundamental group, then one has a better upper bound kd; see [NR2], Theorem 1.4.

1 Serre's theorem and meridianal sweep-outs.

Let us begin by defining meridianal sweep-outs of Riemannian 2-spheres.

Definition 1.1 *We define a meridianal sweep-out of M by curves of length $\leq L$ as a map $f : S^2 \longrightarrow M$ of non-zero degree such that the image of every meridian of S^2 under f is at most L. We will refer to the images of meridians of S^2 under f as meridians of M.*

We claim that upper bounds for the "k-th shortest" geodesic connecting an arbitrary pair of points p,q in a 2-sphere M can be obtained in terms of the maximal length L of the curves in a meridianal sweep-out of M. More precisely, assume that one of the poles of S^2 is mapped into one of the points p,q, say to p, and the other pole is mapped into some point $y \in S^2$. The already mentioned proof of Serre's theorem given by A. Schwartz implies that the length of the first k geodesics between p and q do not exceed $2(k-1)L + dist(p,q) \leq 2(k-1)L + d$. Here is a brief exposition of the proof of this fact (the details can be found, for example, in [NR1].) One can consider the 2-dimensional real homology class h of the space of loops based at p on S^2, $\Omega_p M$, represented by the following map of the 2-torus: The image of (ϕ, ψ) is the loop based at p that starts as the image under f of the meridian of S^2 corresponding to the angle ϕ, followed by the image under f of the meridian corresponding to ψ and travelled in the opposite direction towards p. All Betti numbers of $\Omega_p S^2$ are equal to one; a dual to h real cohomology class c in $H^2(\Omega_p S^2, R)$ has non-trivial cup powers c^i for all i, which are generators of $H^{2i}(\Omega_p S^2, R)$. On the other hand, generators of $H_{2i}(\Omega_p M, R)$ can be chosen as Pontryagin powers of h and represented by maps of tori T^{2i}, where the image of each point consists of $2i$ images of varying meridians of S^2 under f alternatively travelled from p and back to p. The lengths of all these loops are bounded by $2iL$. (Today all these facts easily follow from the rational homotopy theory; cf. ch. 16 of [FHT].) Attaching to all loops in each of these $2i$-cycles a fixed minimal geodesic between p and q we obtain corresponding real cycles in the space of paths connecting p and q, $\Omega_{pq} M$. All paths in these cycles have lengths $\leq 2iL + dist(p, q)$. Now we can use the Morse theory yielding $k-1$ non-trivial paths corresponding to the first $k-1$ positive values of i (in addition to the minimal geodesic). The length of the path that is a critical point corresponding to $H^{2i}(\Omega_{pq} M, R)$ does not exceed $2iL + dist(p, q)$. If two of these $k-1$ critical paths coincide (a degenerate situation), then according to the classical theory developed by Lyusternik and Shnirelman there will be not just one critical point but a whole critical level with uncountably many critical points.

Therefore, our first intention would be to construct a meridianal sweep-out of M by curves of length $\leq$ *const* d (for a value of *const* that is slightly larger thanh 11) that maps one of the poles of S^2 into p. More precisely, we will be looking for a meridianal sweep-out of M by curves of length $\leq 11d + o(1)$. Here and below we use the notation $o(1)$ for positive terms that can be made arbitrarily small by choosing appropriate values of parameters in our constructions. If such a sweep-out exist, then for each k there exist k distinct geodesics connecting p and q of length $\leq 22(k-1)d + dist(p, q)$. (Indeed, there exist $k-1$ geodesics of length $\leq 22(k-2) + dist(p, q) + o(1) < 22(k-1)d + dist(p, q)$ and either the kth geodesic of length $\leq 22(k-1)d + dist(p, q)$ or an infinite sequence of geodesics connecting p and q with lengths that form a strictly decreasing sequence converging to $22(k-1)d + dist(p, q)$. Passing to the limit of an appropriate subsequence we obtain a geodesic of length equal to $22(k-1) + dist(p, q)$ which is longer than the first $k-1$ geodesics.) However, one should be aware that, in general, it is impossible to sweep-out a 2-dimensional sphere by curves of length at most L, where L depends on the diameter of a sphere only. (Y. Liokumovich has constructed a family of Riemannian 2-spheres of diameter ≤ 1, such that the maximal length of curves in an optimal sweep-out becomes infinitely large, (see [L]). His paper is based on ideas of S. Frankel and M. Katz in [FK].) Therefore our plan cannot be carried over in some cases of more "rugged" Riemannian metrics on S^2. Yet we devise a method of demonstrating that in this cases there still will be many short geodesics between p and q because of the ruggedness of the metric. This was the general scheme of similar proofs in [NR1] and [NR2], but our new method for the rugged situation based on ideas from [Cr] and [NR3] leads to the linear in k estimate as opposed to the earlier quadratic bounds.

2 Structure of the proof.

Theorem 0.1 will be proved in two steps. Initially, we will prove it in the case when M is a real analytic Riemannian manifold. Then we will show that the analytic case implies the general (smooth) case. (This last step of the proof is straightforward and will be explained at the end of the paper.)

Thus, througout the most of the paper we will be assuming that M is real analytic. We will attempt to construct a meridianal sweep-out of M by curves of length $\leq 11\ d + o(1)$, where d denotes the diameter of M. If such a sweep-out exists, then the conclusion of Theorem 0.1 is true as a by-product of Schwarz's proof of Serre's theorem (as it was explained in the previous section). As it had been already mentioned, no matter what is the value of *const*, there are some Riemannian metric on S^2, for which there is no meridional sweep-out by curves of length $\leq$ *const* d. Yet we will proof that if there is no a meridianal sweep-out of M by curves of length $\leq 11d + o(1)$, then there exists a closed domain $C \subset M$ that is homeomorphic to a closed annulus, such that $p, q \in C$, both connected components of the boundary ∂C of C are short, and the boundary of C is convex to C. (Following C. Croke ([Cr]), we say that (a connected component of) the boundary of a closed domain $C \subset M$ is convex to C, if the minimal geodesic in M between each pair of sufficiently close points on (this connected component of) ∂C is contained in C. We use the term "closed domain" to denote the closure of an (open and connected) domain. All closed domains that appear in this paper are homeomorphic either to a closed 2-disc or a closed 2-annulus. We say that the boundary ∂C of a closed domain C in M is convex to C at a point $z \in \partial C$ if there exists an open subarc D of ∂C such that $z \in D$, for each two points of D there exists a unique minimizing geodesic in M that connects them, and this minimizing geodesic is contained in C.)

More specifically, we will show that each of the two boundary components is a simple periodic geodesic, or a simple geodesic loop, or a geodesic digon, or a geodesic triangle (see fig. 1). Moreover, the angles at all non-smooth points of the boundary measured inside C are less than π. (Clearly these conditions imply the convexity to C.) Finally, one can also ensure that the length of both connected components of the boundary of C will be less than $3d + o(1)$. Note that it is possible for the points p and q to lie on the boundary of C. Also, if one of two connected components of C will be a geodesic digon or a geodesic triangle, then the second component will be a non-trivial periodic geodesic.

Figure 1 (a) depicts C in which the two boundary components are periodic geodesics. On figure 1 (b) one of the boundary componenets is a geodesic digon, and the second one is a periodic geodesic. On figure 1 (c) one of the boundary components is a geodesic loop, while on figure 1 (d) both of the boundary components are geodesic loops.

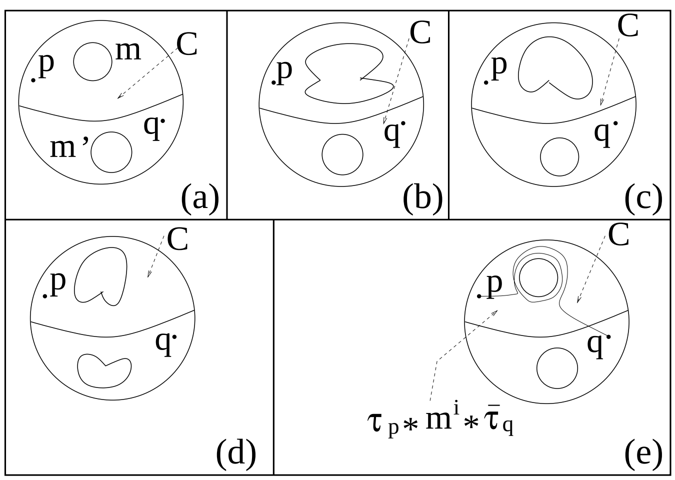

Figure 1:

The proof of this assertion takes a major part of this article. Once this assertion is established, one will need only to verify that the existence of an annulus C with the described properties also implies the existence of infinitely many "short" geodesics between p and q. This can be achieved as follows. Let us denote the shortest component of ∂C by m, and the other component by m', (see fig. 1). Connect p inside C with a point $x \in m$ that is the closest (in the inner metric of C) point to p in m by a path τ_p of length $\leq \frac{length(m')}{2} + d$. Also connect q with a closest in C point $y \in m$ by a path τ_q of length $\leq \frac{length(m')}{2} + d$. Now consider paths $\tau_p * m^i * \bar{\tau}_q$ formed by travelling along τ_p, then along m i times, then from x to y along m in the direction of m (we include this arc of m between x and y in m^i), and finally moving from y to q along $\bar{\tau}_1$, (see Fig. 1 (e)). (Here i can be any positive integer number. Also, here and below we are using the following notations. The symbol $*$ between symbols denoting two paths is used to denote the join of the two paths, and a bar above a symbol denoting a path means that the path is travelled in the opposite direction.)

Apply the Birkhoff curve shortening process with the fixed endpoints to these curves. The process will be described below. It turns out that the process terminates at different geodesics between p and q for different values of i (Lemma 3.8), because these curves have different winding numbers (in the disc bounded by m' that contains C) with respect to any point inside the disc bounded by m that is disjoint with C. These geodesics are sufficiently short to satisfy the upper bound for their lengths postulated in Theorem 0.1. It is important here that these curves stay in the closed annulus during the application of the Birkhoff curve shortening process. We are going to explain the reason why this so below following a description of the Birkhoff curve shortening process.

Now we would like to interupt our exposition of the proof of Theorem 0.1 in order to introduce several versions of the Birkhoff curve shortening process. (One of these versions had been already mentioned above.) We will begin by defining the *Birkhoff curve shortening Process for Free Loops* that will be denoted as BPFL (cf. [Cr] for a good and detailed description of this Birkhoff curve-shortening process and its properties for free loops). It is constructed as follows. We begin by parametrizing a curve γ_0 by its arclength. Next we subdivide it into N intervals of equal length for some large N, so that the distance between each consecutive pair is smaller than the injectivity radius of the manifold divided by some sufficiently large constant. We then join these consecutive points by the unique minimal geodesics. We thus obtain a broken geodesic curve $\gamma_{1/2}$. It is not difficult to see that there exists a length non-decreasing homotopy between γ_0 and $\gamma_{1/2}$ (cf. [Cr]. The idea is that at each moment of this homotopy and for each $i = 1, \ldots, N$ we first follow the ith segment of the original curve γ_0, and then go to the endpoint of this segment along the minimizing geodesic.) Now connect the midpoints of adjacent geodesic segments of $\gamma_{1/2}$ (including the first and the last segments) by minimal geodesics. These minimal geodesics will form a new curve γ_1. Note that $\gamma_{1/2}$ and γ_1 can be connected by a length non-increasing homotopy using the same idea that was previously used to connect γ_0 and $\gamma_{\frac{1}{2}}$. Then process is repeated. We can inductively define γ_{i+1} by parametrizing γ_i by the arclength, dividing it into N segments, and then connecting the midpoints of these segments by minimizing geodesics. Since N is sufficiently large, one can connect all pairs of curves γ_{i-1} and γ_i by length non-increasing homotopies "filling" the "triangles that we are "cutting away", (see the details on p. 4-5 in [Cr]). The process always converges to a possibly trivial periodic geodesic.

Anorther version of this process that is important for this paper is the *Birkhoff curve shortening Process for Based Loops* that we will refer to it as BPBL. In this case we do not connect the midpoints of the first and the last geodesic segments. Instead we connect these midpoints with x, so that all curves will be loops based at x.

Birkhoff curve shortening Process for Based Loops ends at a point or at a non-trivial geodesic loop based at x.

Finally, the third version of the Birkhoff curve shortening process, that we will call the *Birkhoff curve shortening Process for Segments* , denoted BPS. In this case, the original curve is a segment that connects a pair of points x and y. In this version, we define the process that shortens the length of a segment connecting the points x and y, while keeping the endpoints x and y fixed. We connect the midpoints of all segments. The midpoint of the first segment is connected with x, and the last segment with y. The process converges to a (not necessarily minimal) geodesic segment.

Note that process is length non-increasing, and the distance between points that we need to connect by geodesics does not exceed $length(\gamma)/N$. We call this ratio *the rate* of the Birkhoff curve-shortening process. As mentioned before, we have to choose N sufficiently large to ensure that the rate does not exceed the injectivity radius of M, $inj(M)$, but sometimes one needs to choose the rate to be small enough to satisfy some additional conditions.

We would like to observe that if a closed curve bounding a disc D is convex to D, and the rate of BPFL is sufficiently small, then all curves obtained from ∂D during BPFL will remain in the closure of D. (This fact played a prominent role in [Cr].) The same will be true if we regard ∂D as a based loop and apply BPBL to it. The same will be true when one considers a BPFL, or BPBL, or BPS with a sufficiently small rate for any curve contained in D. The last assertion will still be true, when one applies the BPFL or a BPBL with a sufficiently small rate to a closed curve (or a based loop), or BPS to a segment in a closed annulus C, such that both connected components of of C are convex to C. The reason why all these assertions are true is very simple. A curve can move out of a closed disc or annulus during a BPFL, or BPBL, or BPS only if the minimal geodesic segment connecting two very close points on this curve intersects the boundary of the disc (or the annulus), and at least partially passes through the complement of the closed disc (or the annulus). This means that there exist two very close points on the boundary of the disc (or the annulus) such that the minimizing geodesic connecting them in M is outside of the disc (or the annulus). This contradicts the assumption of the convexity of the boundary, and, therefore, is impossible. This explains the mentioned above fact that curves winding around an interior component in a closed annulus with a convex boundary (to the annulus) stay in the closed annulus during BPS and, therefore, do not change their winding number with repect to any point inside the interior disc of the annulus.

Let us now return to the description of our proof. We are going to attempt a construction of a meridional sweep-out of M by curves of length $\leq 11d + o(1)$, such that one of the poles of S^2 is mapped into p. We would like to demonstrate that our attempt can fail only if there is an annulus C containing p and q with the properties stated above. We already saw that this would be sufficient to prove Theorem 0.1.

We are attempting to construct a desired sweep-out using an obstruction to the following extension procedure. We begin with a diffeomorphism $F : S^2 \longrightarrow M$ and try to extend it to the standard 3-ball, which is, obviously, impossible. Assume that S^2 is triangulated into small simplices in such a way that the diameters of all simplices in the induced triangulation of M are smaller than some small parameter ε, which in its turn is much smaller than the injectivity radius of M (and can be made arbitrarily small). Let D^3 be triangulated as the cone over the triangulaion of S^2 with the new vertex at the center of D^3. The extension procedure will be inductive to the skeleta of D^3. Let us begin with the 0-skeleton, that contains the 0-skeleton of the chosen triangulation of S^2 that is being mapped to M by F as well as one additional point, namely the center of D^3 that will be denoted as $\tilde{p}$. Let us define $F(\tilde{p}) = p$. Next, let's consider the 1-skeleton of D^3. The new 1-simplices are of the form $[\tilde{p}, \tilde{v}_i]$, where $\tilde{v}_i$ are the vertices of the triangulation of S^2. We will map $[\tilde{p}, \tilde{v}_i]$ to a minimal geodesic that connects p with $v_i = F(\tilde{v}_i)$. Next, we would like to extend F to the 2-skeleton of D^3. Let us consider a 2-simplex of the form $\sigma_{ij}^2 = [\tilde{p}, \tilde{v}_i, \tilde{v}_j]$. Suppose, that there is a singular disk D_{ij} in M, that has $F(\partial\sigma_{ij}^2)$ as its boundary, and sweeped out by curves of length at most L, where the sweep out is generated by a continuous 1-parametric family of segments that continuosly connects the point p with the points on a 1-simplex $[v_i, v_j]$, (see fig. 2 (a)). Then we will map σ_{ij}^2 to the disc generated by this sweep-out. Since F is a diffeomorphism, we should not be able to extend this map to the 3-skeleton. Thus there exists a 3-simplex $\sigma_{ijk}^3 = [\tilde{p}, \tilde{v}_i, \tilde{v}_j, \tilde{v}_k]$, such that the degree of the map $F|_{\partial\sigma_{ijk}^3} : \partial\sigma_{ijk}^3 \longrightarrow M$is different from 0. Observe that this results in a meridianal sweep out of M by curves of length at most $L + \varepsilon$. Here is a proof of this simple observation:

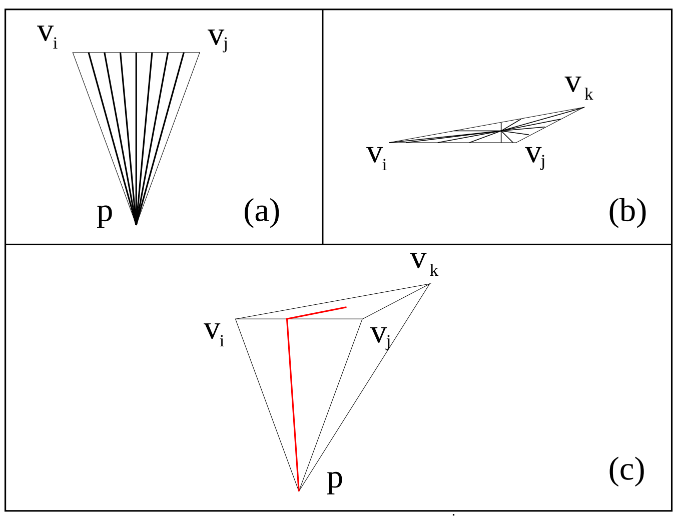

Figure 2:

Let z_0 be a point in the center of the triangle $[v_i v_j v_k]$, (see fig. 2 (b)). Continuously connect z_0 with $\partial[v_i, v_j, v_k]$ by minimal geodesics of length $\leq \varepsilon$. Extend each segment of the sweep out of each of the faces $[p, v_i, v_j]$, $[p, v_i, v_k]$ and $[p, v_j, v_k]$ by a short segment of length $\leq \varepsilon$ ending at z_0. Thus we obtain a sweep out of M of length at most $L + \varepsilon$, (see fig. 2 (c)).

Thus, we would be done if $L \leq 11d + o(1)$. So, the crucial step in our construction arises when we try to extend to the 2-skeleton. Namely, given $F(\partial\sigma_{ij}^2)$ can we always find such a disc D_{ij}? A positive answer will imply the conclusion of Theorem 0.1.

We will attempt to construct D_{ij} as follows. Without any loss of generality we can assume that $\delta_{ij} = F([v_i, v_j])$ intersects the cut locus of p, which, in the case of a real analytic metric, is homeomorphic to a tree, in a finite number of points. Without any loss of generality we can also assume that these points of intersection are inside the edges of the cut locus of M, and not at the verticies.

Let us attempt to continuously connect δ_{ij} with p by minimizing geodesics. Either we will succeed, which will result in D_{ij}, or there will be points x, such that there are exactly two minimizing geodesics connecting x with p. (All such points x of discontinuity are points in the intersection of δ_{ij} with the cut locus of p, and their set is finite.) We will denote one two minimizing geodesics connecting p with x as σ_1 and the second one as σ_2. Assume that we will find a way to connect any such pair of minimizing geodesics σ_1 and σ_2 between p and a point x on the cut-locus of x by a path homotopy that passes through curves of length $\leq L = 11d + o(1)$. (Recall that a homotopy between two paths connecting the same pair of points is called a path homotopy if it fixes the endpoints.) Then the above argument would imply the conclusion of Theorem 0.1.

Let us connect q with x by some minimal geodesic τ. Choose a minimal geodesic that connects p and q, and denote it as γ, (see fig. 3). On this figure you will see two geodesic triangles with the sides σ_i, γ, τ, where $i = 1, 2$. For each of those triangles we will attempt to construct a path homotopy between σ_i and $\gamma * \tau$ over paths of small length (that is, of length $\leq 11d + o(1)$). If such pair of homotopies can be constructed, then we can combine them into a short homotopy between σ_1 and σ_2, and we are done.

This brings us to the key technical result of this paper, namely Lemma 3.6, that will be proven in the next section. It says that one of the following three possibilities is always realized:

1.There is a path homotopy between σ_i and $\gamma * \tau$ passing through curves of length $\leq 11d + o(1)$ as we wanted;

2. There is a closed domain C homeomorphic to an annulus, containing p and q with boundary components of length $\leq 3d + o(1)$ convex to C, as described above. As it had been already mentioned, an appearance of one such annulus implies the existence of the desired countable collection of short geodesics between p and q. The last assertin is formally stated as Lemma 3.8 in the next section.

3. The path $\sigma_i * \tau$ can be transformed using BPS into a geodesic $\tilde{\gamma}$ (of length $\leq 2d$) that is different from γ. This case is dealt with in Lemma 3.7 that

implies that in this case one of the first two possibilities is still realized.

Note that the proof of Lemma 3.6 is long and uses technical lemmae 3.1-3.5.

Summarizing, we see that either for some point x on the intersection of the cut-locus of p with the image of some edge $[v_j v_k]$ of the triangulation of S^2 under F and some $i \in \{1, 2\}$ there exists a closed annulus C with short boundary components convex to C, and we will immediately obtain the desired short geodesics between p and q as a consequence of Lemma 3.8, or we will obtain a meridional sweep-out of M by curves of length $\leq 11d + o(1)$, and the conclusion of Theorem 0.1 follows from the proof of Serre's theorem given by Schwartz. The mentioned Lemmae 3.6 and 3.7 stated and proven below are the key steps required to complete the proof of Theorem 0.1.

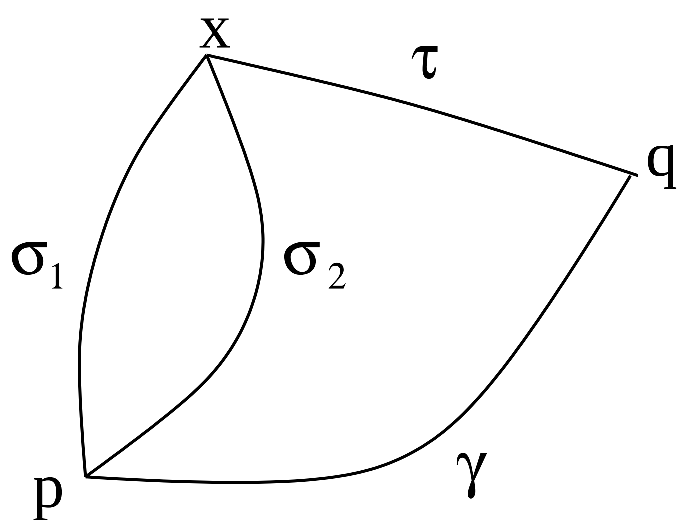

Figure 3:

3 Main Lemmas.

In this section we will prove several technical lemmae that are required in order to show either the existence of a "short" meridianal sweep-out, or of a convex annulus C defined in Section 2.

Lemma 3.1 *Let T be a geodesic triangle with the vertices p, q, x and segments σ_i connecting p and x, γ connecting p and q, and τ connecting q with*

x. Let the length of γ be l_1, and the length of τ be l_2.Then if there is a path homotopy between $\sigma_i * \bar{\tau}$ and γ, such that the length of curves in the homotopy is at most l_h, there is a path homotopy between σ_i and $\gamma * \tau$ passing through the curves of length at most $l_h + l_2$. Likewise, if there exists a path homotopy between $\bar{\gamma} * \sigma_i$ and τ passing through the curves of length at most l_h, then there exists a path homotopy between σ_i and $\gamma * \tau$ passing through the curves of length at most $l_1 + l_h$.

Proof.

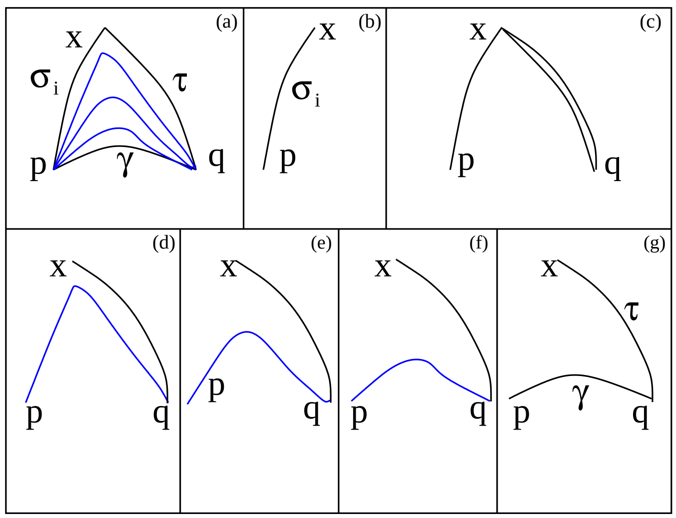

Figure 4:

Let H_t be a path homotopy between $\sigma_i * \bar{\tau}$ and the curve γ. This homotopy is schematically depicted in blue on a figure 4 (a). We will construct the required short path homotopy between σ_i and $\gamma * \tau$ as follows. We begin with the curve σ_i (fig. 4 (b)). This curve is obviously path homotopic to the curve $\sigma_i * \bar{\tau} * \tau$ (see fig. 4 (c)). The homotopy consists of following along $\bar{\tau}$ and back for longer and longer interval of time. Next we will homotope $\sigma_i * \bar{\tau}$ to γ, obtaining $\gamma * \tau$ at the end (fig. 4 (g)). The intermediate curves will be of the form $H_t * \tau$. They are depicted in fig. 4 (d)-(f), and their length is at most $l_h + l_2$. The second statement can be proven in a similar way. $\square$

Lemma 3.2 *Consider a digon made out of segments σ_1 and σ_2 that connect points p and q, and have lengths l_1 and l_2. Let $x \in \sigma_i$, where $i = 1$ or 2. Note* *that a join of σ_1 and $\bar{\sigma}_2$ is a closed curve. In particular, we can consider it as a loop based at the point x. Assume that there exists a path homotopy, contracting this loop to x over the loops of length at most l_s based at x. Then there exists a path homotopy between σ_1 and σ_2 through curves of length at most $l_i + l_s$. If $x \in \sigma_1 \bigcap \sigma_2$, then there exists a path homotopy between σ_1 and σ_2 that passes through curves of length $\leq \min\{l_1, l_2\} + l_s$.*

Proof. Without any loss of generality, assume $x \in \sigma_1$.

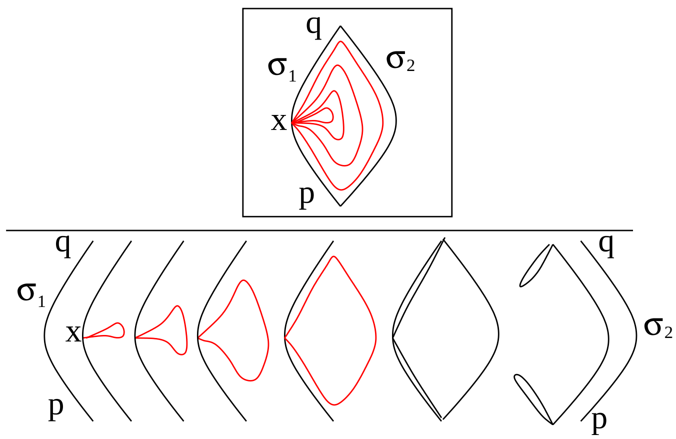

Figure 5:

Let $h_t, t \in [0, 1]$, be a path homotopy of $\sigma_1 * \bar{\sigma}_2$ to x. It is schematically depicted on fig. 5 (a). Let us begin with σ_1. Its homotopy with σ_2 is depicted on fig. 5 (b). First, we begin by going along σ_1 up to the point x and growing a loop out of x, by reversing the homotopy h_t, then following along σ_1. At the end of this process we will end up with the curve σ_1 that we follow up to the point x, then returning back to the point p and continuing along the curve σ_2 and then along σ_1 fro q to x, and then completing the curve by following σ_1 starting from x and ending at q. Both portions of σ_1 can be contracted along themselves to p and q respectively. That completes the proof. □

Definition 3.3 (Monotone homotopy) *Let M be a manifold diffeomorphic to S^2. We will say that a homotopy $\alpha_t : S^1 \times I \longrightarrow M$ between a simple (non self-intersecting) closed curve α_0 and a point y is monotone if the domain bounded by α_s, $s \in I$ containing y contains also all domains bounded by α_t that contain y for all $t > s$.*

Lemma 3.4 *Let α be a simple closed curve of length l in M that is convex to a domain $\Omega \in M$. Suppose that a BPFL with α as the initial curve converges to a point y in Ω. Then for every $x \in \alpha$ there exists a fixed point homotopy over the curves of length at most $2d + 2l$ that converges to the point x.*

Proof.

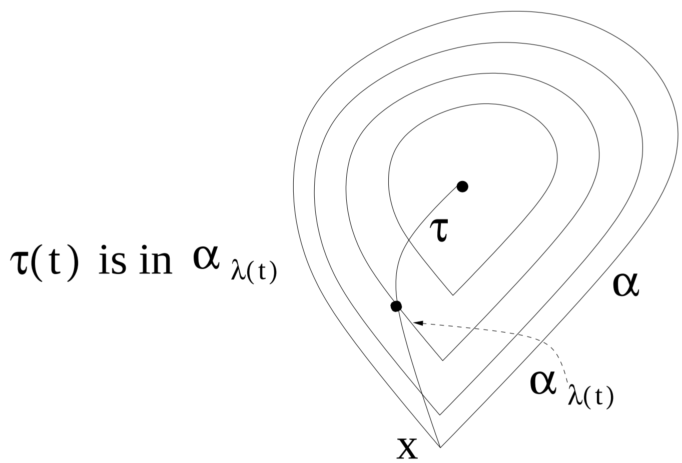

Figure 6: Transforming a free loop homotopy into a based loop homotopy

Let α_t be the homotopy generated by the BPFL that connects α with y. One can show that α_t is monotone and passes through curves that are convex with respect to domains containing y, (see Lemma 2.2 in [Cr]). Let us first assume that the curves α_t are disjoint. This property will be called *strict monotonicity*. Let D be the domain, such that $y \in D$, and $\partial D = \alpha$. We claim that the distance $dist_D(x, y) \leq d + \frac{l}{2}$. Indeed, let $w \in \partial D$ be a point that is closest to y in ∂D. It can be connected to y by a minimal geodesic in Ω that is contained inside D. x can then be connected to w by and arc of α of length at most $\frac{l}{2}$. So, the distance between x and y in D does not exceed the sum of the length of this arc and d.

Let $\tau : [0, 1] \longrightarrow D$ be a minimal geodesic in D that connects x and y. We can see that the following condition is implied by the strict monotonicity and convexity of α_t: For each $t \in [0, 1]$ there exists the unique $\lambda(t)$ such that $\tau(t) \in \alpha_{\lambda(t)}$, which continuously depends on t. Let us consider the loop $\beta_t = \tau_t * \alpha_{\lambda(t)} * \bar{\tau}_t$, where τ_t denotes an arc of τ obtained by restricting τ to the interval $[0, t]$. β_t is a loop based at x of length at most $2d + 2l$, as the length of τ_t does not exceed $d + \frac{l}{2}$.

Now let us assume that α_t is monotonous but not necesserily strictly monotonous. Proceed as in the previous paragraph and consider $A(t) = \{s : \tau(t) \in \alpha_s\}$. Convexity of α_s implies that $A(t)$ is either a point $\lambda(t)$ or a closed interval $[\lambda_1(t), \lambda_2(t)]$. Note also that the sets $A(t)$ are disjoint for different values of t (see the proof of Lemmae 6 and 7 in [M] for a proof of the two last assertions). Thus, in this case a typical loop in the path homotopy will be the following: $\tau_t * \alpha_s * \bar{\tau}_t$, where $s \in [\lambda_1(t), \lambda_2(t)]$, (see fig. 6). (For a given value of t the value of s increases from $\lambda_1(t)$ to $\lambda_2(t)$.)

□

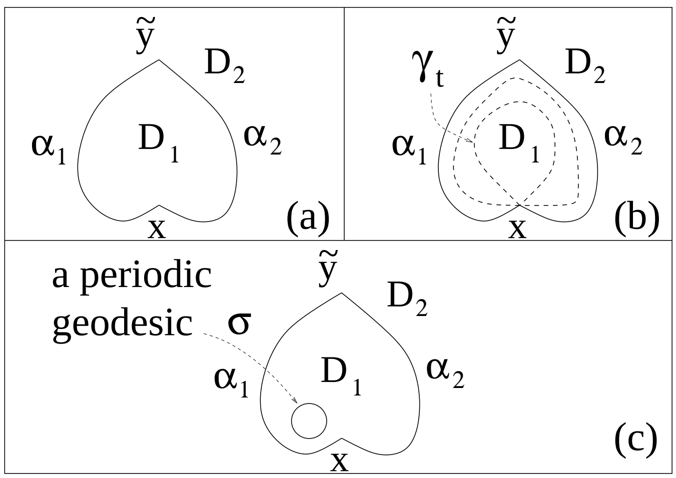

Figure 7:

Lemma 3.5 *(1) Let α_1, α_2 of length l_1, l_2 be two non-intersecting geodesics connecting a pair of points $x, \tilde{y}$. Suppose D_1, D_2 are two domains in M that have $\gamma = \alpha_1 * \bar{\alpha}_2$ as their common boundary. Assume that γ is not convex to neither D_1 nor D_2. More precisely, assume that γ is convex to D_1 at $\tilde{y}$ (and, therefore, not at x; see fig. 7 (a)). Then, if the BPBL with a sufficiently small rate that fixes the point x and has $\alpha_1 * \bar{\alpha}_2$ as its initial curve leads to a self intersection, then either there exists a periodic geodesic σ of length at most $l_1 + l_2 + o(1)$ inside D_1, (see fig. 7 (c)), or there exists a path homotopy between α_1 and α_2 that passes throught the curves of lengths at most $2(l_1 + l_2) + 2d + o(1)$.*

(2) More generally, we can reach the same conclusion if α_1, α_2 are any two curves joining x with $\tilde{y}$ such that $\alpha_1 * \bar{\alpha}_2$ is simple and convex with respect to D_1 at each point except for x.

Proof. Let γ_t denote the curves during the BPBL with fixed x and $\gamma_0 = \gamma$, (see fig. 7 (b)). Also, we would like to define a 1-parametric family of (embedded) discs D_t^1, such that $\partial D_t^1 = \gamma_t$, and $D_{t_2}^1 \subset D_{t_1}^1 \subset D_1$, when $t_1 \leq t_2$. An analysis of the argument in Lemma 2.2 in [Cr] used to prove that in the convex case BPFL is a monotone homotopy that goes through simple curves demonstrates that in our case curves γ_t will be simple, and D_t^1 will be a monotone family of discs uniquely defined for all values of t until one of the curves γ_t will "come close" to x in the following sense: The minimizing geodesic connecting two (very close) points a, b of γ_t on the next step of BPBL will intersect γ_t very closely to x, so that it will intersect an arc γ^{beg} of γ_t that leaves x and an arc γ^{end} of γ_t that comes to x (see fig. 8, 9). Also, if we choose the rate of the Birkhoff curve shortening process to be sufficiently small, then we can assume that a, b, γ^{beg} and γ^{end} are inside of a convex metric ball in M centered at x, and that this geodesic segment will intersect the curve γ_t exactly once to the either side of x, (see Fig. 9). Not only this is the only way that can lead to appearance of a self-intersection (and the loss of monotonicity of the homotopy), but a simple argument using the Jordan curve theorem shows that if this happens, a self-intersection must necessarily develop.

Let us consider the geodesic ray starting at x and bisecting the angle of D_t^1. In fact let us consider a sector of $B_x(\varepsilon)$ between γ^{beg} and γ^{end} that has a non-empty intersection with D_t^1. The geodesic ray will subdivide this sector into two convex domains Ω_1 and Ω_2, (see Fig. 10).

Note that if a, b are both either in Ω_1 or in Ω_2 then the minimal geodesic connecting them has to stay in Ω_1 or in Ω_2 respectively, because of the convexity of these domains. Therefore a, b lie in different domains. Without loss of generality, assume $a \in \Omega_1$ and $b \in \Omega_2$. Let us consider an arc γ^0 of the curve γ_t between a and b. Once again by choosing the rate of the Birkhoff curve shortening process very small, we can assume that this arc lies in $B_x(\varepsilon) \cap \bar{D}_t^1$. Thus, γ_0 must intersect the geodesic ray coming out of x and bisecting the angle of D_t^1 at some point y in $B_x(\varepsilon)$ different from x.

There are two arcs of γ_t between x and y. Denote these arcs by γ^1 and γ^2, and their lengths by $\tilde{l}_1$, $\tilde{l}_2$. Consider two closed curves formed by γ^i and the bisector xy, $i = 1, 2$. Each of these two curves that we will denote σ_1 and σ_2 will be convex to the subdomain of D_t^1 that it bounds, as the angles

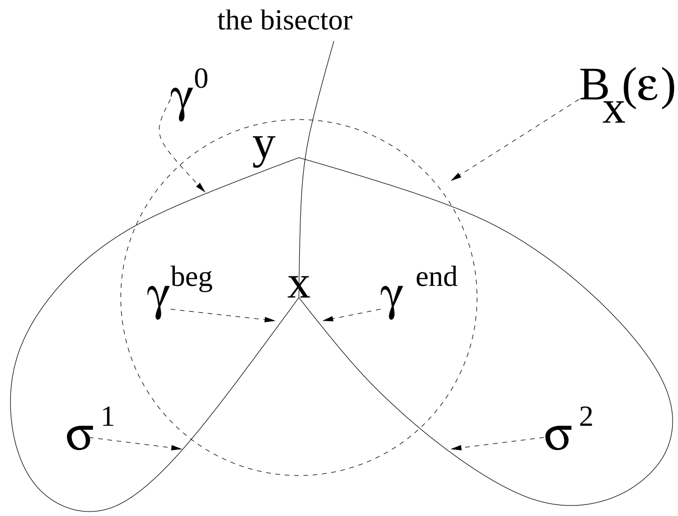

Figure 8: The case when the angle at x is greater than π

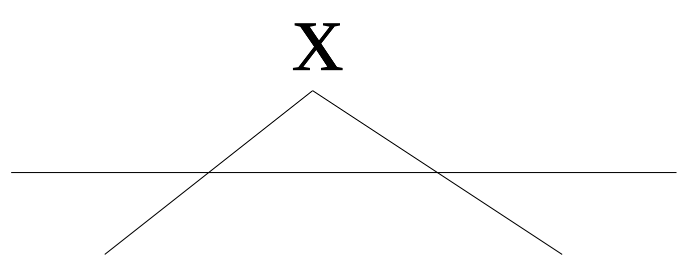

Figure 9: A possilbe self-intersection of the curve during the homotopy

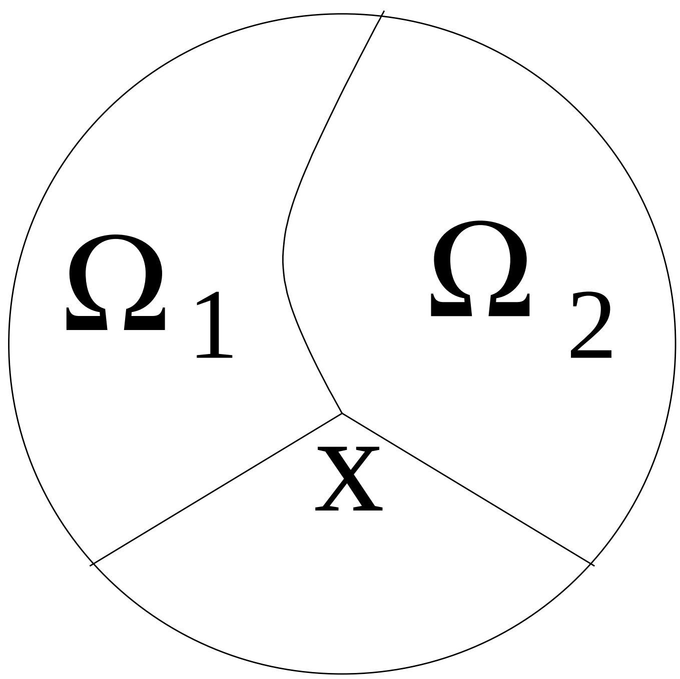

Figure 10: Domains Ω_1 and Ω_2

at x and y are less than π, and γ_t was convex to D_t^1 at all points but x.

Assume that there exist path homotopies that contract σ_i to a point via loops based at x of length $\leq 2d + 2\tilde{l}_i + o(1)$ for $i = 1, 2$.

Then these two path homotopies can be merged in an obvious way, so that we obtain a path homotopy contracting γ_t to a point via loops based at x of length $\leq \max\{2\max\{\tilde{l}_1, \tilde{l}_2\}, \max\{\tilde{l}_1, \tilde{l}_2\} + 2\min\{\tilde{l}_1, \tilde{l}_2\}\} + 2d + o(1)$. Indeed, we can first insert the segment xy travelled twice in the opposite direction, then contract the shortest of two loops σ_1, σ_2 (as a based loop), and then contract the longer one.

Now note that we can apply BPFL to σ_i. It will either converge to a non-trivial periodic geodesic of length at most $\tilde{l}_i + o(1) \leq l_1 + l_2 + o(1)$ inside $D_t^1 \subset D_1$, or to a point via curves of length at most $\tilde{l}_i + o(1)$. In the former case we are done. In the latter case, let us apply Lemma 3.4 to conclude that there is a homotopy with a fixed point x contracting σ_i to x over the loops of length at most $2\tilde{l}_i + 2d + o(1)$.

It then follows that there exist either a "short" periodic geodesic, or a fixed point homotopy connecting γ with the point x over loops of length at most $\leq \max\{2\max\{\tilde{l}_1, \tilde{l}_2\}, \max\{\tilde{l}_1, \tilde{l}_2\} + 2\min\{\tilde{l}_1, \tilde{l}_2\}\} + 2d + o(1)$. We can now apply Lemma 3.2 to obtain a path homotopy between α_1 and α_2 that passes through the curves of length $\leq \max\{2\max\{\tilde{l}_1, \tilde{l}_2\}, \max\{\tilde{l}_1, \tilde{l}_2\} + 2\min\{\tilde{l}_1, \tilde{l}_2\}\} + 2d + o(1) + \min\{\tilde{l}_1, \tilde{l}_2\} = \tilde{l}_1 + \tilde{l}_2 + \max\{\max\{\tilde{l}_1, \tilde{l}_2\}, 2\min\{\tilde{l}_1, \tilde{l}_2\}\} + 2d + o(1) \leq 2(\tilde{l}_1 + \tilde{l}_2) + 2d + o(1) \leq 2(l_1 + l_2) + 2d + o(1)$. □

Lemma 3.6 *Let T be a geodesic triangle with three vertices p, q, x such that its three sides are length-minimizing geodesics γ, connecting p and q, τ connecting q and x and σ_i that connects p and x. Then one of the following three possibilities holds true:*

1) There exists a geodesic digon connecting the points p, q, where one of the edges is a geodesic $\tilde{\gamma}$ of length at most $2d$, and the second edge is the geodesic γ of length at most d as well as a length non-increasing path homotopy between $\sigma_i * \bar{\tau}$ and $\tilde{\gamma}$;

2) There exists a path homotopy between σ_i and $\gamma * \tau$ over the curves of length at most $11d + o(1)$, or

3) There exists a closed domain C in M homeomorphic to the closed annulus, with $p, q \in C$, in which each boundary component is convex to C, and has length not exceeding $3d + o(1)$.

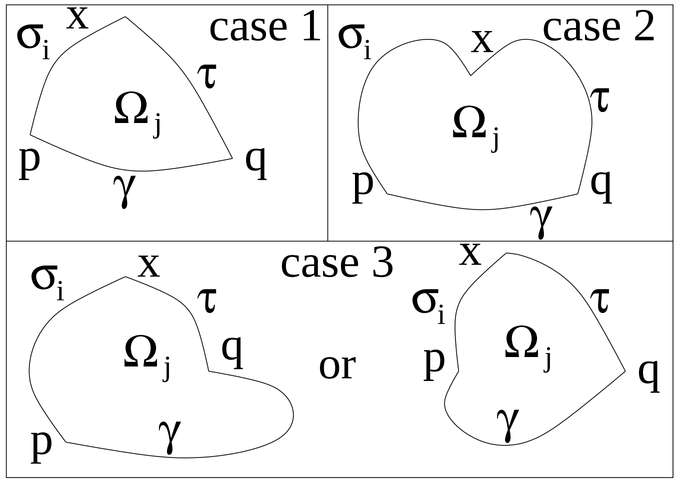

Figure 11:

Proof. Note that the geodesic triangle T is a simple closed curve, by the virtue of being made out of minimal geodesics. Thus, by the Jordan curve theorem, it subdivides M into two Riemannian disks. Let us denote them as Ω_1 and Ω_2. We will consider the following three cases, (see fig. 11):

Case 1. T is convex to one of the domains;

Case 2. There exists a domain Ω_j, ($j = 1$ or 2), such that T is convex at the points p, q with respect to Ω_j, and is concave at x with respect to Ω_j (that is, the angle in Ω_j between two sides of T adjacent to x is $> \pi$).

Case 3. There exists a domain Ω_j, such that T is convex at p, x with respect to Ω_j, but is concave at q with respect to Ω_j, or T is convex at q, x with respect to Ω_j, but is concave at p.

Case 1. This is the simplest case to consider. Let us apply the BPFL to T.

Then either it will converge to a periodic geodesic σ of length at most $3d$, (see fig. 12 (a)), or to a point, (see fig. 12 (b)). In the former case, we get C, such that $p, q \in C$ as the domain bounded by T and by this periodic geodesic. Notice that the length of both of the boundary components is bounded from above by $3d$. In the latter case, apply Lemma 3.4 to obtain a homotopy that fixes the point p over the curves of length at most $2 \times 3d + 2d = 8d$. Figure 12 (c) depicts a typical curve in this homotopy. Now apply Lemma 3.2 to a digon that consists of points p, x connected by segments σ_i and $\gamma * \tau$ to obtain a path homotopy between σ_i and $\gamma * \tau$ that passes through curves of length at most $d + 8d = 9d$, (see fig. 12 (d) that depicts a typical curve

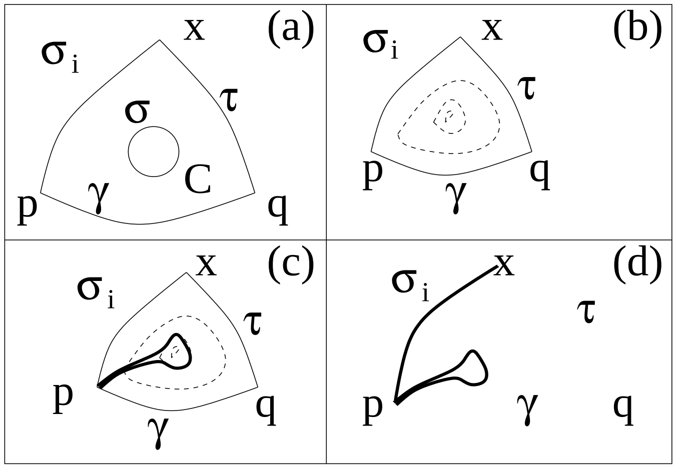

Figure 12:

in this homotopy).

Case 2. Without any loss of generality assume that Ω_2 is such a domain, (see fig. 13 (a)). In this case, let us begin by applying the BPS to $\sigma_i * \bar{\tau}$.

Possibility (i). The process converges to a geodesic $\tilde{\gamma}$ of length $\leq 2d$ that is different from γ and that does not intersect γ. In this case, we are done, (see fig. 13 (b)).

Possibility (ii). The process converges to γ, (see fig. 14 (a)). In this case, let us apply Lemma 3.1 to obtain a path homotopy between σ_i and $\gamma * \tau$ that passes through the curves of length at most $3d + d = 4d$, (see fig. 14 (b) that depicts a typical curve in this homotopy).

Possibility (iii). Intersections between the curves in the homotopy and the curve γ develop during the BPS. The only way this can possibly happen is for some arc of $(\sigma_i * \bar{\tau})_t$ is to come close to p and / or another arc of $(\sigma_i * \bar{\tau})_t$ is to come close to q (as the curves $(\sigma_i * \bar{\tau})_t$ remain convex the subdisc of Ω_1 that it bounds at all points, but p, q. A simple formal proof of this fact is analogous to the proof of Lemma 2.2. in [Cr]). Here $(\sigma_i * \bar{\tau})_t$, where $t \in [0, 1]$ denote the curves during BPS. In this case we will proceed as in the proof of Lemma 3.5. We are going to assume that the first intersection develops in the neighborhood of the point p, (see fig. 15). (The case when the first intersection develops near q can be treated in almost the same way.) That is, some arc a of $(\sigma_i * \bar{\tau})_t$ comes within distance of ε from p (where ε is much smaller than the injectivity radius of M). In this case the angle at p

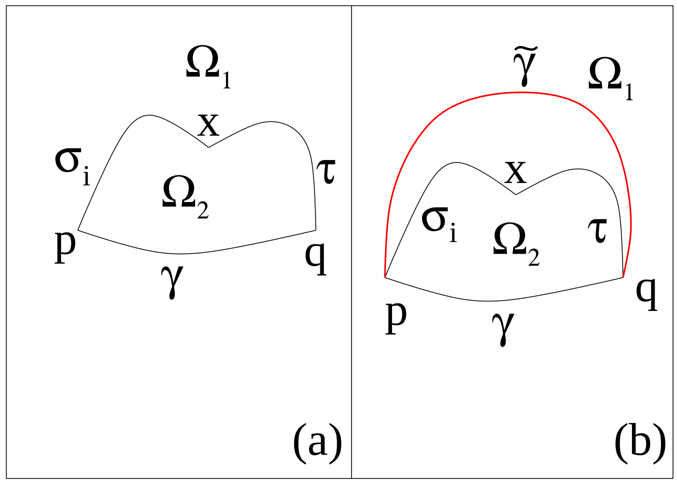

Figure 13:

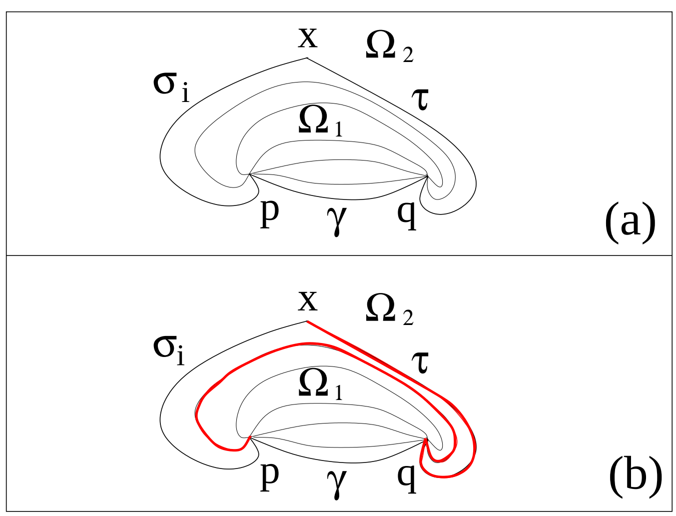

Figure 14:

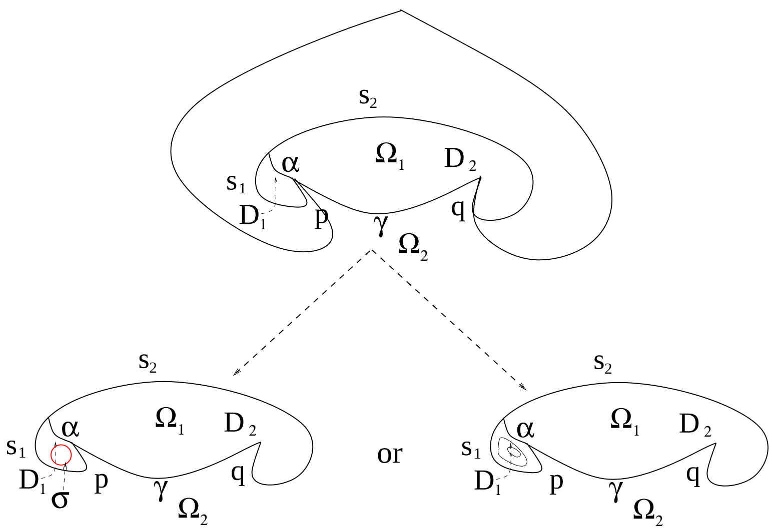

Figure 15:

between γ and $(\sigma_i * \bar{\tau})_t$ with respect to a subdomain $(\Omega_1)_t$ of Ω_1 is $> \pi$. Let us bisect the angle at π by a length minimizing geodesic α of length $< \varepsilon$ and intersecting the arc a. The point of intersection will subdivide $(\sigma_i * \bar{\tau})_t$ into two segments s_1, s_2. of length l_1, l_2 respectively. Also, the geodesic α will subdivide $(\Omega_1)_t$ into two domains D_1 and D_2. $s_1 * \bar{\alpha}$ is a loop based at p. It is convex with respect to the domain $D_1 \in \Omega_1$. The curve $\bar{\gamma} * \alpha * s_2$ is convex to D_2, except possibly at the point q.

Let us apply the BPFL to the curve $s_1 * \bar{\alpha}$. It either converges to a periodic geodesic σ of length at most $l_1 + o(1) \leq 2d + o(1)$, or to a point over the curves of length at most $l_1 + o(1)$, (see fig. 15).

In the former case we will proceed as follows:

Construction 1.

Let us apply BPBL to the geodesic triangle T regarded as a loop based at x. T is convex to Ω_2 except at x. The following are the possibilities: (a) the process converges to x, (see fig. 16); (b) the process converges to a geodesic loop β based at x, and no self-intersections develop in the process, (see fig. 17); (c) self-intersections develop during the homotopy.

First let us consider (a). The BPBL converges to x over the curves of length at most $3d$. Let us apply Lemma 3.2 to obtain a path homotopy between σ_i and $\gamma * \tau$ over the curves of length at most $4d$. In this case, we are done. Figure 16 (b) depicts a typical curve in the homotopy.

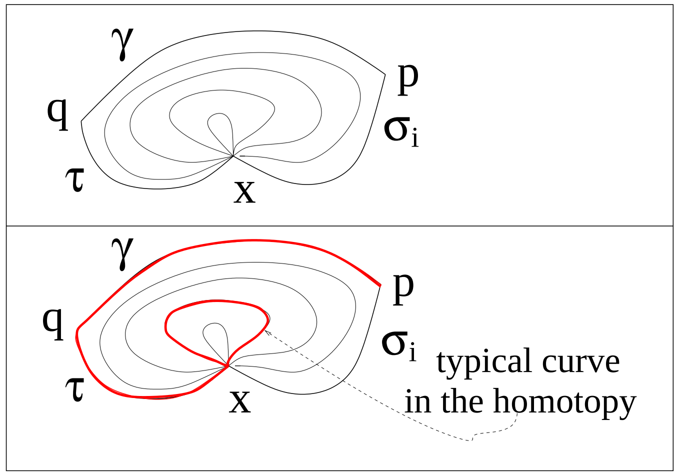

Figure 16:

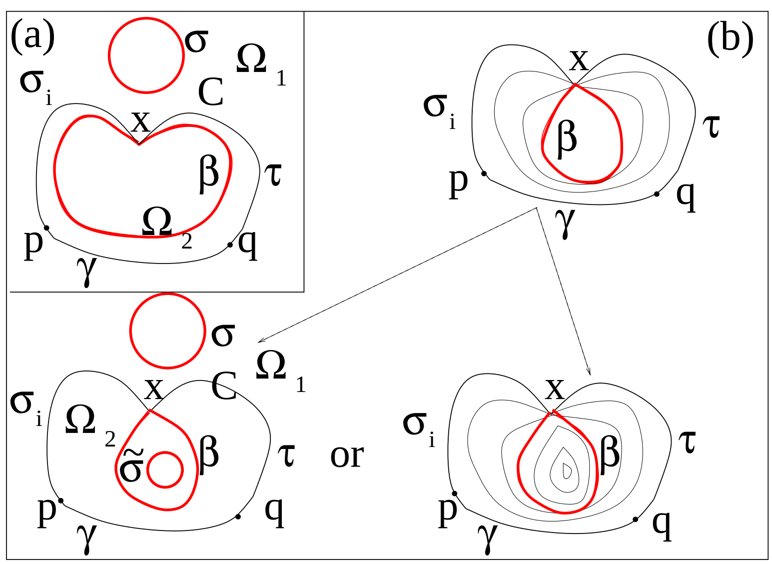

Figure 17:

Next let us consider (b). A geodesic loop β at x is either convex to some domain in Ω_2, (see fig. 17 (b)), or is concave, (see fig. 17 (a)). If it is concave, then we have obtained our C. It is bounded by β of length at most $3d$ and σ of length at most $2d + o(1)$, (see fig. 17 (a)). If it is convex, let us apply the BPFL to it. It will either converge to a periodic geodesic of length at most $3d$, or to a point via curves of length at most $3d$, (see fig 17 (b)). If it converges to a periodic geodesic $\tilde{\sigma}$, once again we obtain C. It is bounded by σ and $\tilde{\sigma}$. If it converges to a point, Lemma 3.4 implies that there is a fixed point homotopy of β to the point x over the curves of length at most $8d$. Next we apply Lemma 3.2 which tells us that there exists a path homotopy between σ_i and $\gamma * \tau$ over the curves of length at most $8d$.

Finally, we will examine case (c). In this case let us apply Lemma 3.5. We see that either there exists a periodic geodesic of length at most $3d+o(1)$ inside Ω_2, or there exists a path homotopy between σ_i and $\gamma * \tau$ over the curves of length at most $2 \times 3d+2d+o(1) = 8d+o(1)$. In the second case we are done. In the first case this periodic geodesic and σ form the boundary of a desired closed annulus C, and the assertion of the lemma is true.

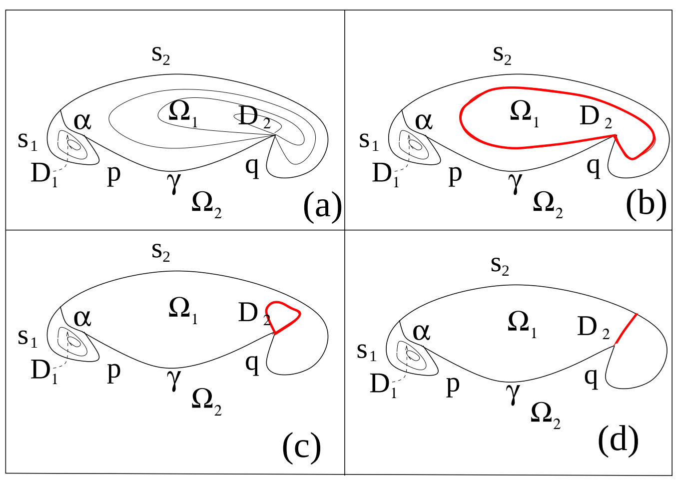

Figure 18:

Let us now go back and consider the case in which BPFL applied to curve $s_1 * \bar{\alpha}$ converges to a point. In this case there is a fixed point homotopy of this curve to the point p over the curves of length at most $2l_1 + 2d + o(1)$ by Lemma 3.4. Let us consider $\bar{\gamma} * \alpha * s_2$. This is a loop that is based at q and convex to D_2 everywhere except possibly at q. Let us apply BPBL to this curve. There are the following three possibilities. (a) BPBL converges to the point q, (see fig. 18(a)). (b) It converges to a geodesic loop in D_2, and thus in Ω_1, and no self-intersection develops in the process, (see fig. 18 (b) and (c)); (c) a self-intersection develops. Figure 18 (d) depicts a curve $(\bar{\gamma} * \alpha * s_2)_t$ coming close to itself in the neighborhood of q just before the appearance of the first self-intersection, and the point q being connected to some point in $\bar{\gamma} * \alpha * s_2$ by a minimal geodesic of length at most ε, where ε is much smaller than the convexity radius of M (compare with the proof of Lemma 3.5).

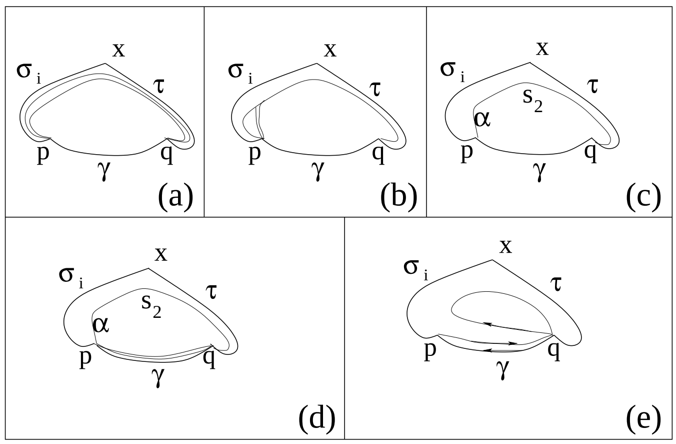

Figure 19:

Construction 2. Before we will consider the above three cases, note that one can show that if there exists a fixed point homotopy h_t of the loop $\bar{\gamma} * \alpha * s_2$ to the point q over the curves of length at most l_q for some number l_q, then there exists a path homotopy between σ_i and $\gamma * \tau$ that passes through the curves of length at most $3d + \max\{l_q,\ l_1 + 3d\} + o(1) \leq 3d + \max\{l_d,\ 6d\} + o(1)$.

To show this, we first construct the following homotopy of the loop $\sigma_i * \bar{\tau} * \bar{\gamma}$ to the point p. The based loop $\sigma_i * \bar{\tau} * \bar{\gamma}$ is path homotopic to $s_1 * s_2 * \bar{\gamma}$ over the curves of length at most $3d$, (see fig. 19 (a)) by virtue of $\sigma_i * \bar{\tau}$ being path homotopic to $s_1 * s_2$ over the curves of length at most $2d$. The loop $s_1 * s_2 * \bar{\gamma}$ is path homotopic to $s_1 * \bar{\alpha} * \alpha * s_2 * \bar{\gamma}$ over the loops of length at most $3d + o(1)$, (see fig. 19(b)). This last homotopy can be described as inserting longer and longer segments of α travelled in both directions. The latter loop is path homotopic to $\alpha * s_2 * \bar{\gamma}$ via loops based at p of length $\leq 2l_1 + l_2 + 3d + o(1) \leq 3d + o(1)$, as we are considering the case, when $s_1 * \bar{\alpha}$ is path homotopic to the trivial loop p via loops of length $\leq 2l_1 + l_2 + 2d + o(1) \leq l_1 + 5d + o(1)$ based at p (fig. 19 (c)). The loop $\alpha * s_2 * \bar{\gamma}$ is path homotopic to $\gamma * \bar{\gamma} * \alpha * s_2 * \bar{\gamma}$ over the loops of length at most $5d + o(1)$ (see fig 19 (d)). (This homotopy involves gradually inserting the segment $\gamma * \bar{\gamma}$.) Lastly, $\gamma * \bar{\gamma} * \alpha * s_2 * \bar{\gamma}$ is path homotopic to $\gamma * \bar{\gamma}$, (see fig. 19 (e)) over the loops based at p of length at most $2d + l_q + o(1)$. Obviously, $\gamma * \bar{\gamma}$ is contractible to p along itself. Thus, there exists a homotopy that fixes the point p of the loop $\sigma_i * \bar{\tau} * \bar{\gamma}$ and contracts this loop to the point p over the curves of length at most $2d + \max\{l_q, l_1 + 3d\} + o(1)$. Then by Lemma 3.2 there exists a path homotopy between σ_i and $\gamma * \tau$ that passes through the curves of length at most $3d + \max\{l_q, l_1 + 3d\} + o(1) \leq 3d + \max\{l_q, 6d\} + o(1)$.

Now let us go back to considering the different possibilities of convergence of BPBL applied to $\bar{\gamma} * \alpha * s_2$.Let us first examine case (a) In this case $l_q \leq l_2 + d \leq 3d + o(1)$, and the assertion of the lemma is true.

Next let us consider case (b). The geodesic loop that we obtain is either convex to the subdisk in D_2 that it bounds, or is concave to it. (In other words, the angle at its vertex measured in the subdisk is either $\leq \pi$, or is $> \pi$.) In the case of convexity we apply the BPFL to this geodesic loop. It will either converge to a periodic geodesic or to a point. If it converges to a point, then by Lemma 3.4 there is a fixed point homotopy of $\bar{\gamma} * \alpha * s_2$ to the point q that passes through the curves of length at most $2l_2 + 4d + o(1)$, so by Construction 2 there exists a path homotopy between σ_i and $\gamma * \tau$ over curves of length at most $7d + \max\{2l_2, 2d\} + o(1) \leq 11d + o(1)$ (as $l_q = 2l_2 + 4d + o(1)$ and $2l_2 \leq 2(l_1 + l_2) \leq 4d$). In both remaining cases, namely, of either a concave geodesic loop at q, or a periodic geodesic, let us perform Construction 1. We will get either a path homotopy between σ_i and $\gamma * \tau$ over the curves of length at most $9d + o(1)$, or a domain C homeomorphic to the annulus, such that the length of each boundary component is at most $3d + o(1)$.

Finally, let us consider case (c). Let us apply Lemma 3.5 to the curve $\bar{\gamma} * \alpha * s_2$. By this lemma either there exists a periodic geodesic of length at most $3d + o(1)$ in Ω_1, or there exists a path homotopy between α and $\gamma * \bar{s}_2$ over the curves of length at most $8d + o(1)$. Therefore, in this case there exists a path homotopy between $\alpha * \bar{s}_1$ and $\gamma * \bar{s}_2 * \bar{s}_1$over the curves of length $\leq 9d + o(1)$. Now $\gamma * \bar{s}_2 * \bar{s}_1$ is path homotopic to $\gamma * (\tau * \bar{\sigma}_i)$ via based loops of length $\leq 3d$, and, as we are considering the case, when $\alpha * \bar{s}_1$ is homotopic to the trivial loop via based loops of length $\leq 2l_1 + 2d + o(1) \leq 6d + o(1)$, we see that $\gamma * \tau * \bar{\sigma}_i$ is path homotopic to the trivial loop based at p via based loops of length $\leq 9d+o(1)$, and an application of Lemma 3.2 implies the existence of a desired path homotopy between σ_i and $\gamma * \tau$. In the former case, when there exists a non-trivial periodic geodesic of length $\leq 3d+o(1)$ in Ω_1, we find ourselves in a situation similar to the situation that was considered above, when BPFL applied to $s_1 * \bar{\alpha}$ ended at a periodic geodesic. Exactly as before we can finish the proof of the lemma applying construction 1 and the arguments described above after the text of construction 1.

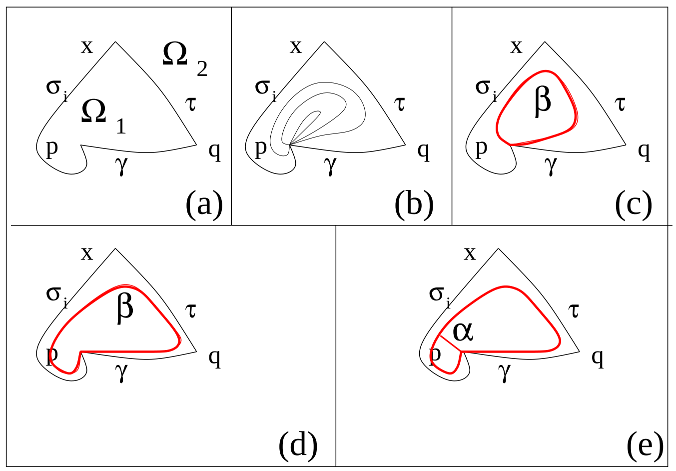

Figure 20:

Case 3. Without loss of generality, let us assume that T is convex to Ω_1 at the points x, q and concave to Ω_1 at p, (see fig. 20 (a)).

Let us first consider the loop $\sigma_i * \bar{\tau} * \bar{\gamma}$ based at p. Let us apply the BPBL to it. If the curve does not begin to self-intersect during the homotopy, it will either converge to p over the curves of length at most $3d$, (see fig. 20 (b)), or to a geodesic loop at p, (see fig. 20 (c) and (d)). Note that the curves in the homotopy will stay inside Ω_1. Figure 20 (e) depicts the curve coming close to itself in the neighborhood of p, and minimal geodesic α of length at most ε connecting p with the point on this curve that is closest to p. If the BPBL converges to the geodesic loop, this geodesic loop will be in Ω_1 and will either be convex to some domain in Ω_1 at p, (fig. 20 (c)), or not, (fig. 20 (d)). If it is convex to some domain in Ω_1, we will apply the BPFL to it. It will then either converge to a periodic geodesic, or to some point. In the last case by Lemma 3.4 there will be a fixed point homotopy of the original curve to p over the curves of length at most $8d$. So one of the following is true: either there exists a path homotopy between σ_i and $\gamma * \tau$ over the curves of length at most $9d$ by Lemma 3.2, or a geodesic loop of length at most $3d$ in Ω_1 that is concave at p with respect to some domain in Ω_1, or a periodic geodesic of length at most $3d$ in Ω_1. If a self-intersection develops, let us apply Lemma 3.5. We will conclude that either there exists a periodic geodesic in Ω_1 of length at most $3d + o(1)$, or a path homotopy between σ_i and $\gamma * \tau$ that goes through the curves of length at most $8d + o(1)$. If we obtain the required path homotopy, then we are done. In the remaining two cases, we note that it would be sufficient for us to obtain either a periodic geodesic of length $\leq 3d + o(1)$ in Ω_2, or a geodesic loop of length $\leq 3d + o(1)$ in Ω_2 concave to the subdisc of Ω_2 bounded by this loop. With this observation in mind we will proceed as follows:

First, we consider a segment $\bar{\gamma} * \sigma_i$ connecting the points q and x and apply the BPS to this segment, (see fig. 21 (a)). There are the following possible outcomes:

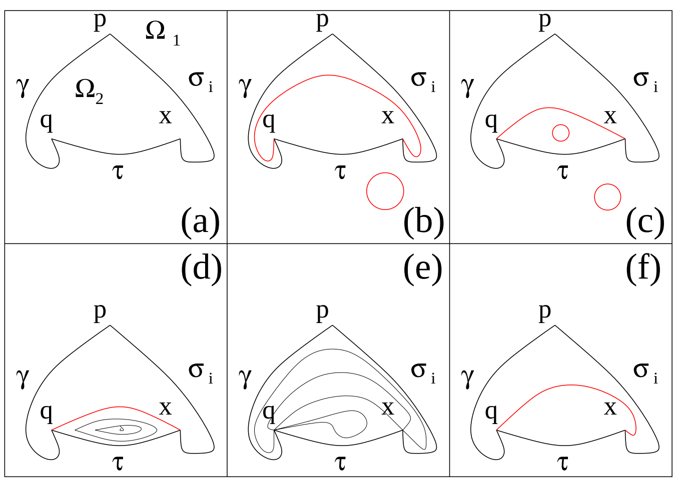

Figure 21:

Possibility (i). The process converges to a non self-intersecting geodesic segment λ of length at most $2d$ that is different from τ. Note that in this case the resulting digon will bound some domain in Ω_2. There are three cases to consider: case (a)- the angles at x and at q measured in this domain are both greater than or equal to π. In this case, we are done, since this digon, together with either a geodesic loop, or a periodic geodesic obtained in the previous paragraph, will be the boundary of the required annulus C, (see fig. 21 (b)). Case (b) is the case when both of the angles are less than *pi*, (see 21 (c)). In this case, let us apply the BPFL to the digon. It will either converge to a point, (fig. 21 (d)), or to a periodic geodesic, (fig. 21 (c)). In the case of a periodic geodesic we are done. (That is, we obtained the desired annulus C as the domain between this periodic geodesic and either a periodic geodesic or a geodesic loop obtained above.) If, however, the process converges to a point, we apply Lemma 3.4 to obtain a fixed point homotopy of the digon to q over the curves of length at most $8d$, (see fig. 21 (e)). Then by Lemma 3.2 there exists a path homotopy between $\bar{\gamma} * \sigma_i$ and τ that goes through the curves of length at most $9d$. Now by Lemma 3.1 there exists a path homotopy between σ_i and $\gamma * \tau$ that goes through the curves of length at most $10d$.

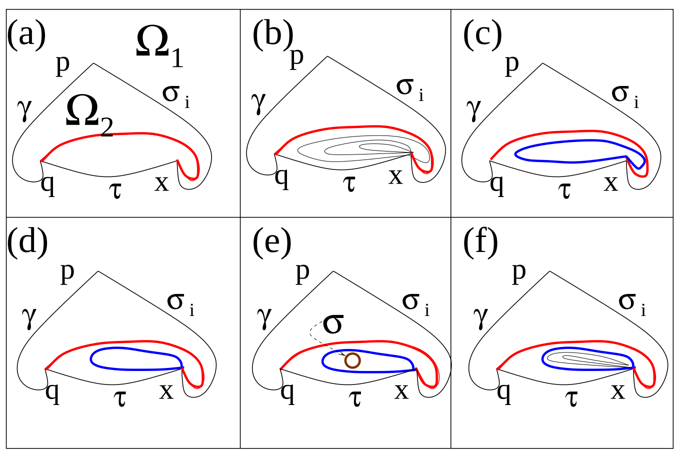

Figure 22:

Finally, we will consider case (c), when one of the angles of the digon formed by λ and τ is less than π, and another is greater than or equal to π, (see fig. 21 (f) and 22 (a)). We are going to consider only the case the curve is concave at x, and to observe that the case, when this curve is concave at q can be treated completely similarly. Let us apply the BPBL to the loop. Let us first assume that no self-intersection forms during the process. Then the process will either converge to a point, (fig. 22 (b)), to a geodesic loop in Ω_2 with an angle $\geq \pi$, (fig. 22 (c)), or to a geodesic loop with an angle $< \pi$, (fig. 22 (d)). In the latter case, apply the BPFL to this geodesic loop. The process will either converge to a periodic geodesic σ, (fig. 22 (e)), or to a point, (fig. 22 (f)). In the case of a periodic geodesic,

or the previously mentioned case of a geodesic loop with an angle $\geq \pi$, we will be done. In both the cases we obtain the desired domain C. If either the BPFL or the earlier considered BPBL leads to a point, we can easily obtain a path homotopy between $\sigma_i * \tilde{\tau}$ and γ that goes through the curves of length at most $10d$, by successful applications of Lemmae 3.4 and 3.2. Finally, we should consider what we are going to do when self-intersections develop during the considered BPBL. In this case, let us apply Lemma 3.5. We will obtain either a periodic geodesic in Ω_2 of length at most $3d + o(1)$ (and in this case we are done), or a path homotopy between σ_i and $\gamma * \tau$ that goes throught the curves of length at most $8d + o(1)$.
Possibility (ii) The process converges to γ. In this case by Lemma 3.1 there exists a path homotopy between σ_i and $\gamma * \tau$ over the curves of length at most $4d$.
Possibility (iii) Intersections between the curves in the homotopy and γ develop in the process. In this case, let us proceed as in Case 2 to obtain either a periodic geodesic in Ω_2, or a concave geodesic loop in Ω_2 of a sufficiently small length, or a path homotopy between σ_i and $\gamma * \tau$ that passes through curves of an acceptable for us length.

□

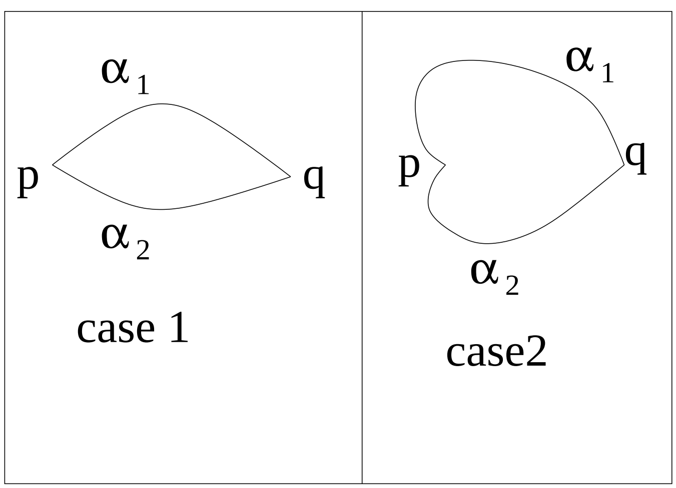

Figure 23:

Lemma 3.7 *Let α_1, α_2 be two non-intersecting geodesics connecting a pair of points $p, q \in M$. of length l_1, l_2 respectively. Then either there exists a* *path homotopy between α_1 and α_2 that passes through the curves of length at most $2(l_1 + l_2) + \max\{l_1, l_2\} + 2d$, or there exists a closed domain C homeomorphic to the closed annulus, such that $p, q \in C$, and the length of the each of the two connected components of the boundary does not exceed $l_1 + l_2 + o(1)$.*

Proof.
Let us consider the digon that is obtained from the curves α_1 and α_2. It is a simple closed curve. Therefore, it subdivides M into two connected components Ω_1 and Ω_2, each of which is homeomorphic to the 2-disc. We will consider the following two cases:
Case 1: One of the discs is convex;
Case 2: Both of the discs are not convex, (see fig. 23).

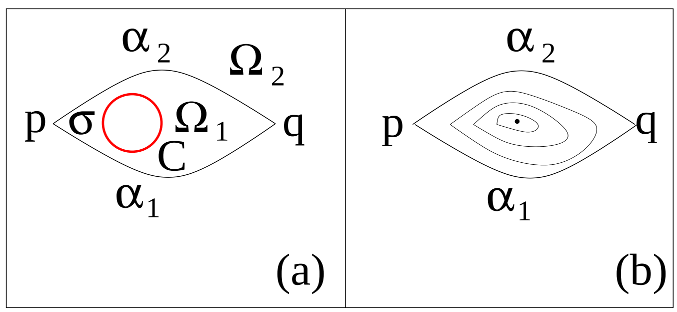

Figure 24:

Case 1. Withoout loss of generality, let us assume that Ω_1 is convex. Let us apply the Birkhoff curve shortening to the digon. It will either converge to a non-trivial periodic geodesic σ inside Ω_1, (see fig. 24 (a)), or to a point inside Ω_1, (see fig. 24 (b)). In each of the cases the length of curves in the homotopy will be at most $l_1 + l_2$. In the former case, we will obtain the desired domain C. It is the annulus bounded by this periodic geodesic and the original digon. Moreover, the length of each of the boundary components does not exceed $l_1 + l_2$. In the latter case, we will use Lemma 3.4 to conclude that there exists a fixed point homotopy starting from the curve $\alpha_1 * \bar{\alpha}_2$ and ending at the point p that passes through loops of length at most $2(l_1 + l_2) + 2d$. Furthermore, Lemma 3.2 implies that there exists a path homotopy between α_1 and α_2 that passes through curves of length at most $2l_1 + 3l_2 + 2d$.
Case 2. In this case, let us fix the point p and apply the BPBL. Without

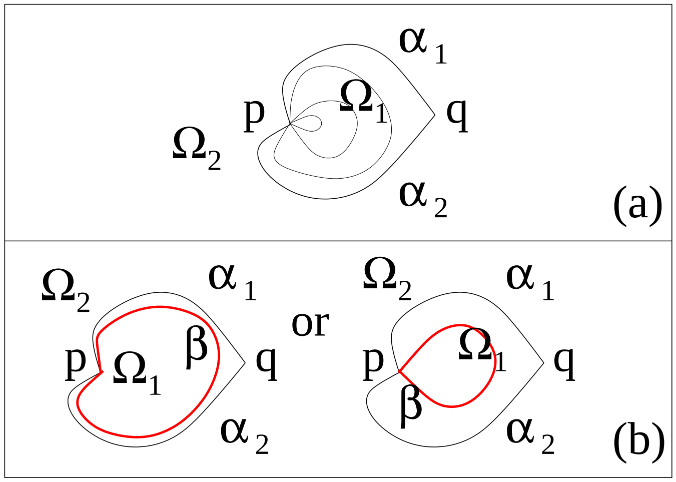

Figure 25:

loss of generality, let us assume that the angle formed by α_1 and α_2 at the point q is less than π with respect to Ω_1. Let us also assume for now that no possible self-intersections occur during the process. Then either the process will converge to p, (see fig. 25 (a)), or it will converge to a geodesic loop β_p inside Ω_1, (see fig. 25 (b)). In the former case, let us appply Lemma 3.2 to obtain a path homotopy between α_1 and α_2 that passes through the curves of lengths at most $l_1 + 2l_2$. In the latter case we need to consider the following two possibilities:

(a) the angle of β_p at p is greater than or equal to π with respect to the subdomain of Ω_1 bounded by β_p;

(b) the angle of β_p at p is less than π with respect to the subdomain of Ω_1 bounded by β_p, (fig. 25 (b)). In this case, we will proceed as in Case 1.

Let us apply the BPFL to β_p Either it converges to a point, (see fig. 26(a)) which will allow us to construct a path homotopy between α_1 and α_2 through curves of length at most $2l_1 + 3l_2 + 2d$, or it will converge to the geodesic σ_p in Ω_1, (see fig. 26 (b)).

In the case of a periodic geodesic σ_p, or if the angle of β_p is greater than π (case (a) above) we will reverse the role of p and q. That is, let us now fix the point q and apply the BPBL to the loop that is based at q defined as $\alpha_2 * \bar{\alpha}_1$. Once again, we are assuming that no intersections will occur during the homotopy. We are faced with the same possibilities. In the case of convergence to the point q, we are done, since there will be a “short”

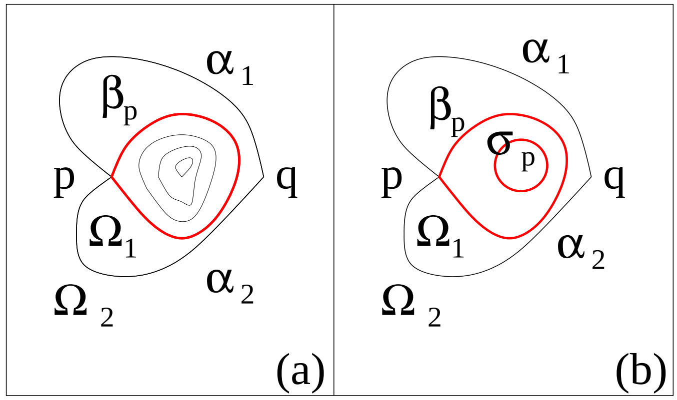

Figure 26:

path homotopy between α_1 and α_2. In the case of convergence to a geodesic loop β_q in Ω_2, we need to consider the angle of β_q at q with respect to Ω_2. If the angle is greater than or equal to π, we are done, because we obtain the required annulus that has as its boundary β_q and either β_p or σ_p. If the angle is less π than we apply the BPFL. It either converges to a point, which results in a short path homotopy between α_1 and α_2, or converges to a periodic geodesic σ_q in the domain Ω_2. Note, that in this case either σ_q together with σ_p, or β_p bounds the required annulus.

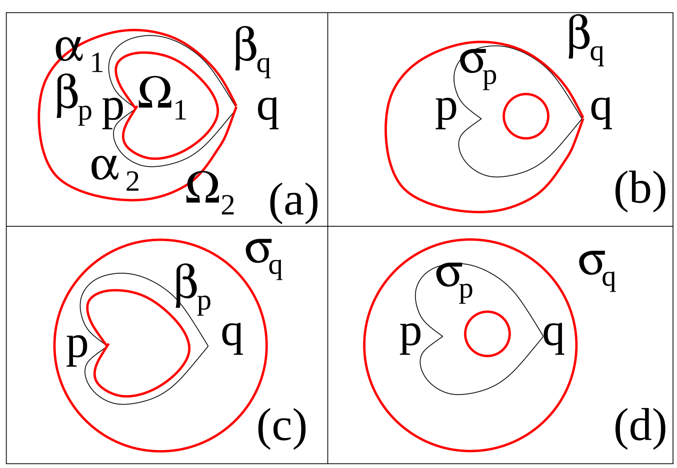

Figure 27:

Figure 27 depicts the four possible cases of the boundary of C described above. The boundary is either formed by β_p and β_q as in fig. 27 (a), or by σ_p and β_q as in fig. 27 (b), or by β_p and σ_q as in fig. 27 (c), or by σ_p and σ_q as in fig. 27 (d).

It thus, remains only to consider what happends if intersections start to form during the BPBL with fixed p and / or with fixed q. In this case, let us apply Lemma 3.5. We conclude that in this case, either there exists a periodic geodesic σ_p inside Ω_1 of length at most $l_1 + l_2 + o(1)$, or a path homotopy between α_1 and α_2 that passes through the curves of lengths at most $2(l_1 + l_2) + 2d + o(1)$. In the former case, let us reverse the role of p and q. Once again, we will either obtain a geodesic σ_q inside Ω_2, of length at most, or a geodesic loop β_q that is convex to the outside of Ω_2, or a short path homotopy between α_1 and α_2.

□

Lemma 3.8 *Let C be a closed domain in M homeomorphic to the annulus, such that both of the boundary components m and m' of C are simple curves that are convex to C. Let us assume that length$(m) \leq$ length(m'). Let $p, q \in C$, p' denotes a point of m that is the closest to p in C, and q' denotes a point of m that is the closest to q in C. For each integer i denote by m^i a path that starts at p', goes along m i times, and then continues from p' to q' along m in the direction of m. Let $\gamma_i = \tau_p * m^i * \bar{\tau}_q$, where τ_p is a shortest curve in C connecting p with p', τ_q is a shortest curve in C that connects q with q'.*

Consider a geodesic g_i obtained from γ_i by the application of the BPS. If the rate s of the process is sufficiently small, then these geodesics are distinct for different i.

Proof.

Let us consider m'. By the Jordan curve theorem it subdivides M into two connected domains. Moreover, since M is diffeomorphic to S^2, these domains are both homeomorphic to the 2-disk. Let D be the disk containing C.

Let $a \in D - C$. Choose a shortest path ω connecting q and p in C. For each pair i, j the difference of the absolute values of winding numbers of $\gamma_i * \omega$ and $\gamma_j * \omega$ with respect to a is equal to $|i - j|$. Clearly, these winding numbers do not change during any path homotopies of γ_i or γ_j that stay in C. (We keep the arc ω fixed during the considered path homotopies.)

Now note that the paths obtained during the BPS that has γ_i as a starting curve stay inside C. Indeed, if the rate of the process is sufficiently small, then new geodesic segments introduced during the process cannot intersect any connected component of the boundary of C.

This observation completes the proof of the lemma. □

4 The Proof of Theorem 0.1.

In this section we will proof Theorem 0.1. The proof will be rather short and follows easily from Lemmas 3.6, 3.7 and 3.8

Proof of Theorem 0.1. We will first consider the case when M is an analytic Riemannian manifold. Later we will explain how to reduce the general (smooth) case to the analytic case.

Assume that M is analytic Riemannian manifold. Recall that in this case for any $x \in M$ the cut locus of x is homeomorphic to a tree. Let $f : S^2 \longrightarrow M$ be a diffeomorphism between the euclidean sphere enowed with a fine triangulation and M. Also, recall that we plan to obtain k short geodesics between $p, q \in M$ as obstructions to an extension of the map $f : S^2 \longrightarrow M$ to the disk D^3 triangluated as a cone over S^2. Recall also, that while the extension to the 0-and 1-skeleta of D^3 is trivial, (we map the center $\tilde{p}$ of the disk to the point p and the edges to some minimal geodesics connecting the point p with the corresponding vertices), the real core of the matter is in how we extend to the 2-skeleton of the disk.

Without any loss of generality, let us assume that the 1-skeleton of the induced triangulation of M intersects the cut locus of p only at the edges in a finitely many number of points. Note that there are exactly two minimizing geodesics that connect p with each point in the interior of one of the edges of the cut locus of p. We will then extend to the 2-skeleton of D^3 in the following manner. Let $[\tilde{p}, \tilde{v}_i, \tilde{v}_j]$ be a typical 2-simplex of D^3. Let us try to continuously connect p with $[v_i, v_j] = f([\tilde{v}_i, \tilde{v}_j])$ by minimal geodesics. In general, this is impossible to do. The discontinuity will occur precisely at the points of intersection of the 1-skeleton of M with the cut locus of p. Let x be such an intersection point. The two minimal geodesics σ_1 and σ_2 connecting pd , and x will form a digon that we would like to fill continuously by short segments connecting p and x. Either we will always succeed at doing this, and will obtain a desired extension to the considered 2-simplex, or we will obtain a domain C, homeomorphic to the annulus with short convex simple boundary components.

The filling will be constructed as follows. Let us connect p with q by a length minimizing geodesic γ and q to x by length minimizing geodesic τ. We will try to construct two path homotopies between σ_i and $\gamma * \tau$, $i = 1, 2$. Combining these two homotopies together will give us the required filling.

Now let us apply Lemma 3.6 to both of the geodesic triangles formed by σ_i, γ and τ. Either there exists a path homotopy between σ_i and $\gamma * \tau$ passing through curves of length at most $11d + o(1)$ or an annulus C satisfying the necessary hypothesis, such that the maximal length of the boundary component is at most $3d + o(1)$, or a geodesic digon between with the vertices p and q, such that one of the geodesics has length at most $2d$, while the second is of length at most d. In the first two cases, we are done. In the third case, let us apply Lemma 3.7 to this geodesic digon. This lemma implies that either there exists a path homotopy between α_1 and α_2 that passes through the curves of length at most $2(3d) + 2d + 2d + o(1) = 10d + o(1)$, or there exists a closed convex domain C that is homeomorphic to an annulus, such that $p, q \in C$, and the length of the maximal boundary componenet is at most $3d + o(1)$. In the latter case, we are done. In the former case we can conclude that there exists a path homotopy between $\sigma_i * \bar{\tau}$ and γ that passes through the curves of length at most $10d + o(1)$. Let us now apply Lemma 3.1, which implies that there exists a path homotopy between σ_i and $\gamma * \tau$ that passes through curves of length at most $11d + o(1)$. Thus, either there exists an annulus C satisfying the hypothesis of Lemma 3.8, and we are done (that is, we have the desired geodesics between p and q by virtue of Lemma 3.8), or we can extend the map f to each simplex of the form $[\tilde{p}, \tilde{v}_i, \tilde{v}_j]$ by mapping this simplex to the disk $[p, v_i, v_j]$ that is generated by continuous family of curves that connect the point p with the edge $[v_i, v_j]$, the length of which does not exceed $11d + o(1)$. (Here is a brief recap of the construction of this continuous family of curves connecting p with all points on $[v_i, v_j]$: We connect p with the points on $[v_i, v_j]$ by families of minimal geodesics of length at most d, which vary continuously unless the point $x \in [v_i, v_j]$ belongs to the interior of an edge of the cut locus of p. In this case, we obtain digons formed by two minimal geodesics σ_1 and σ_2, (see fig. 28). We fill in these digons by constructing path homotopies between σ_1 and $\gamma * \tau$ and σ_2 and $\gamma * \tau$. The length of curves in these homotopies does not exceed $11d + o(1)$.)

Now note that we cannot extend to the 3-skeleton of D^3, since f is a homeomorphism. Thus there exists a 3-simplex $s_{ijk} = [\tilde{p}, \tilde{v}_i, \tilde{v}_j, \tilde{v}_k]$ in the triangulation of D^3, such that $f|_{\partial s_{ijk}}$ has a non-zero degree. The image of each face of this simplex, except for $[\tilde{v}_i, \tilde{v}_j, \tilde{v}_k]$ is meridianally swept out by short curves by the previous step. This can easily be extended as it was described in the introduction to a meridianal sweep out of the boundary of the whole simplex. Note that the curves in the constructed meridianal sweep-out connect the point p, with some other point y , (not necessarily q).

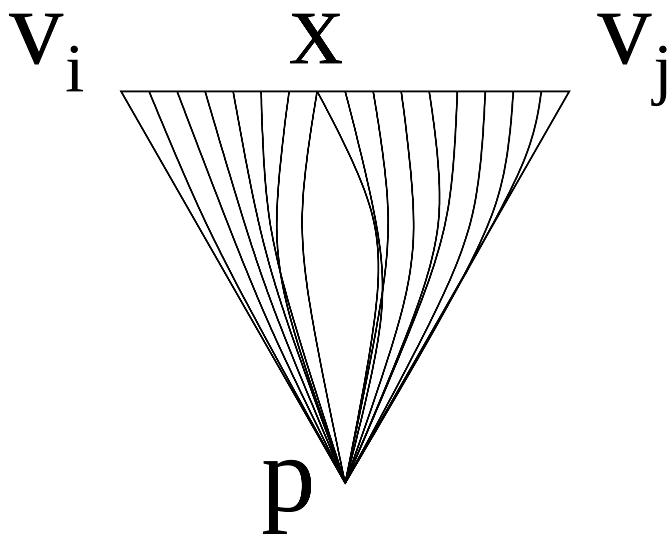

Figure 28: Contracting loops can be reduced to contracting geodesic digons

Finally, let us explain how we pass to the case of a manifold with an arbitrary smooth metric. Let us approximate this Riemannian metric by a sequence of analytic Riemannian metrics (in C^2-topology). For each of these Riemannian manifolds M_n we either can found a meridianal sweep-out of M by short curves, or a closed annulus with boundary formed by two closed curves of length $\leq 3d + o(1)$ such that each of these two curves is a geodesic loop, or a geodesic digon, or a geodesic triangle convex to the annulus. If there is an infinite subsequence of the sequencei $\{M_n\}$, where one has a desired meridianal sweep-out, then we can take a sufficiently close Riemannian metric M_n with the sweep-out, to discretize it and to "transfer" the sweep-out to the limit Riemannian manifold. One chooses the parameter of the discretization equal to the Gromov-Hausdorff distance between M_n and M that is assumed to be less than one tenth of the injectivity radius of M. As the result, one will obtain a meridional sweep-out of M, where the lengths of the meridians can increase in comparison with the upper bound for the length of meridians in the sweep-out of M_n only by a small summand. (This summand is bounded by a multiple of the Gromov-Hausdorff distance between M_n and M, and, therefore, tends to zero, as $n \longrightarrow \infty$.)

Now assume that starting from some n all manifolds M_n contain closed annuli with short connected components of the boundary that are convex to the annuli. We would like to obtain a similar closed annulus in M by passing to the limit in an appropriate subsequence. The angles at vertices of geodesic loops or digons that were $\leq \pi$ will remain $\leq \pi$ in the limit. As our convergence of Riemannian metrics is in C^2 -topology, sequences of geodesics converge to geodesics. Further, Cheeger's inequality implies a uniform positive bound for the injectivity radii of M_n, so sequences of non-trivial geodesic loops on M_n have lengths uniformly bounded from below, and cannot converge to trivial geodesic loops. If the boundary of C consists of a geodesic triangle (convex to C) and a periodic geodesic obtained from the geodesic triangle as the result of an application of BPFL (as at the beginning of the proof of Lemma 3.6) , and this situation occurs for an infinite sequence of values of n, then, when we pass to the limit of an appropriate subsequence, the periodic geodesics cannot collapse to a point. Therefore, the geodesic triangles cannot collapse to a point either. (The situation, when the geodesic triangles converge to the limit periodic geodesic, and C degenerates to a non-trivial periodic geodesic (in the limit) is acceptable for us, as in this case, we obtain infinitely many "short" geodesics between p and q that follow this periodic geodesic.) However, geodesic digons, in principle, can converge to degenerate digons formed by a geodesic segment travelled twice in opposite directions. There are two different ways to exclude this undesirable for us possibility. First, note that analyzing our proof we see that at most one connected component of a closed annulus C can be a geodesic digon. So, the convex discs bounded by the geodesic digons in M_n contain geodesic loops in their interiors. These geodesic loops cannot degenerate in the limit, and, therefore, prevent the geodesic digons from collapsing to a geodesic segment. Second, we observe that the geodesic digons forming a part of the boundary of C appeared as obstructions to contractibility of some paths by path homotopies that are length non-increasing. If the angles between two sides of a geodesic digon are sufficiently small, then the longer side can be path homotoped to the shorter side without length increase. (This follows from the already mentioned fact that the injectivity radii of M_n have a uniform positive lower bound.) We can incorporate this length non-increasing path homotopy into our proof, making sure that the geodesic digon will not appear as a part of boundary of a convex annulus

C. Thus, we can assume that at least one angle in each geodesic digon that appears as a connected component of the boundary of a closed disc C in M_n is greater than some positive number that does not depend on n. Therefore the geodesic digons cannot collapse to a geodesic segment travelled twice in opposite directions. Thus, the desired properties of the annuli persist, when we pass to the limit. This completes the proof of Theorem 0.1 in the smooth case. $\square$

Acknowledgements: Both authors acknowledge a partial support of this work by their corresponding NSERC Discovery Grants. This paper was partially written during the authors' visit of the Max-Planck Institute for Mathematics in Bonn in July-August, 2012. The authors would like to thank the Max-Planck Institute for its kind hospitality.

References

[BCK] F. Balacheff, C. Croke, M. Katz, A Zoll counterexample to a geodesic length conjecture, Geom. Funct. Analysis (GAFA) 19(2009), 1-10, MR2507217, Zbl 1201.53050.

[Cr] C. Croke, Area and the length of the shortest periodic geodesic, J. Differential Geometry, 27(1088), 1-21, MR0918453, Zbl 0642.53045.

[FHT] Y. Felix, S. Halperin, J.-C. Thomas, Rational Homotopy Theory, Springer, 2001.

[FK] S. Frankel, M. Katz, The Morse landscape of a Riemannian disk, Annales de l'institut Fourier, 43 no. 2 (1993), 503-507.

[Gr] M. Gromov, Metric structures for Riemannian and non-Riemannian spaces, Birkhauser, 1991, MR2307192, Zbl 1113.53001.

[L] Y. Liokumovich, Spheres of small diameter with long sweep-outs, to appear in Proceeding of the AMS.

[M] M. Maeda, The length of a closed geodesic on a compact surface, Kyushu J. Math. 48 (1994), no. 1, 9-18, MR1269063, Zbl 0818.53064.

[NR 1] A. Nabutovsky, R. Rotman, Lengths of geodesics on a two-dimensional sphere, Amer. J. Math., 131(2009), 545-569, MR2503992, Zbl 1170.53023.

[NR 2] A. Nabutovsky, R. Rotman, Length of geodesics and quantitative Morse theory on loop spaces, preprint, available at www.math.toronto.edu/alex/morse09.pdf

[NR 3] A. Nabutovsky, R. Rotman, Linear bounds for lengths of geodesic loops on Riemannian 2-spheres, Journal of Diff. Geom., 89 (2011), 217-232.

[R] R. Rotman, The length of a shortest geodesic loop at a point, J. Differential Geom., 78:3(2008), 497-519, MR2396252, Zbl 1143.53038.

[Se] J.-P. Serre, Homologie singulière des espaces fibriés. Applications. Ann. of Math. 54(1951), 425-505, MR0045386, Zbl 0045.26003.

[Sch] A. S. Schwartz, Geodesic arcs on Riemannian manifolds", Uspekhi Math. Nauk (translated in Russian Math. Surveys) 13(6) (1958), 181-184, MR 0102076.

Alexander Nabutovsky
Department of Mathematics
University of Toronto
Toronto, Ontario M5S 2E4
Canada
e-mail: alex@math.toronto.edu

Regina Rotman
Department of Mathematics
University of Toronto
Toronto, Ontario M5S 2E4
Canada
rina@math.toronto.edu