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

Alexander Nabutovsky and Regina Rotman

March 8, 2011

Abstract

Let M be a closed surface diffeomorphic to S^2 endowed with a Riemannian metric. Denote the diameter of M by d. We prove that for every $x \in M$ and every positive integer k there exist k distinct geodesic loops based at x of length $\leq 20kd$.

Introduction.

A well-known result of J. P. Serre asserts that for every closed Riemannian manifold M^n and an arbitrary pair of points $x, y \in M^n$ there exist an infinite set of distinct geodesics connecting x and y, (see [Se]). Later A. Schwartz proved 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 x and y of length $\leq c(M^n)k$, (see [Sch]). In particular, if $x = y$ those geodesics are geodesic loops based at x. Note that one can write $c(M^n)$ as $c_0(M^n)d$, where d is the diameter of M^n and $c_0(M^n)$ is a *scale-invariant* constant.

But what Riemannian invariants are required in order to majorize the lengths of k distinct geodesics connecting a fixed pair of points? In [NR 2] the authors proved that one can get such an estimate using only the diameter and the dimension of M^n. Namely, we proved that the lengths of k distinct geodesics connecting x and y do not exceed $4k^2nd$. (We do not know if there exists such an upper bound that does not involve the dimension, i. e. an estimate of the form $f(k)d$.) A comparison of our result with the result of Shwartz leads to the following natural question: Is it possible to replace the upper bound that is quadratic in k by a linear bound $c(n)kd$? Whenever we do not believe that this is possible for Riemannian manifolds of all dimensions, we prove that this is, indeed, so in the two-dimensional case:

Theorem 0.1 *Let M be a closed Riemannian manifold diffeomorphic to S^2 of diameter d. Let x be a point in M, and k be a positive integer number. Then there exist at least k distinct non-trivial geodesic loops based at x of length not exceeding $20kd$.*

Remark 0.2. If $k = 1$, one can get a better upper bound $4d$ as a corollary of the main result in [R]. Also, a better bound for small values of k follows from the main result of our paper [NR 1], where we proved that there exist $(4k^2 + 2k)d$ distinct non-trivial geodesic loops. Moreover, it was noted in [NR 2] that the last estimate can be immediately improved to $(k^2+3k+2)d$, if one notices that cycles of non-zero dimensions in the space of loops on M based at x used to prove the existence of the geodesic loops never "hang" at critical points of index zero. On the other hand, F. Balacheff, C. Croke and M. Katz demonstrated that it is *not* true that the length of a shortest non-trivial geodesic loop is always $\leq 2d$, (see [BCK]).

Remark 0.3. The linear upper bound of Theorem 0.1 will hold for all closed 2-dimensional Riemannian manifolds and not only those diffeomorphic to S^2. To prove this assertion first assume that M has an infinite fundamental group. Then there exists a non-contractible loop γ based at x of length $\leq 2d$. The existence of such a loop is easy and well-known, (cf. [Gr]). Then one can consider the iterates γ^i of this loop. They all will be non-contractible and non-pairwise homotopic. Now we can obtain the desired geodesic loops by applying a curve-shortening process with the fixed base point x to γ^i. The case, when M is diffeomorphic to $\mathbf{R}P^2$, can be reduced to the spherical case by passing to the double covering of M with the induced Riemannian metric. As the diameter of the double covering of M does not exceed $2d$, we will obtain k geodesic loops of length $\leq 40kd$, that can be then projected to M. Finally, note that our proof of Theorem 0.1 does not seem to apply to the situation when there are k geodesics connecting distinct points $x, y \in M$. (The argument in Section 1.3 below does not seem to work, when one tries to modify it in the case of $x \neq y$.) Nevertheless, we plan to establish a similar linear bound for lengths of geodesics connecting distinct points $x, y \in M$ using a different more complicated argument (that also yields a worse constant) in a sequel to this paper. Also, our proof can be adapted (with modifications) to the case, when x and y are the most distant points in M, (see Theorem 2.1 in Section 2).

1 Proof of the main result.

Proof of Theorem 0.1.

1.1. The length of a meridianal sweep-out and lengths of geodesics.

Definition 1.1 *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 does not exceed L. We will refer to the images of meridians of S^2 under f as meridians of M.*

The proof of the existence of infinitely many geodesic loops based at a prescribed point on $x \in M$ given by A. Schwartz in [Sch] easily implies that if both poles of the sphere are mapped into x, then the lengths of the first k of these loops (including the trivial loop) do not exceed $2kL$ (and do not exceed $(2k+2)L$ if the poles are mapped into arbitrary points of M. See, for example, [NR 1] for a detailed explanation of Schwartz's proof in the case of a 2-sphere.)

Here is a brief explanation of this result. Homology groups $H_i(\Omega S^2, \mathbf{R})$ of the space of based loops on S^2 are all isomorphic to $\mathbf{R}$, and are generated by Pontryagin powers of the generator of $H_1(\Omega S^2, \mathbf{R})$. (Recall, that Pontryagin product in homology groups of loops spaces is induced by the operation of taking the join of loops.) The generator of $H_1(\Omega S^2, R)$ can be represented by a map of a circle into $H_1(\Omega S^2, \mathbf{R})$, where each point of a circle is mapped into the image under f of the meridian of S^2 with the corresponding longitude. (Recall that we assumed that both poles are mapped to x, so every meridian is mapped into a loop in M based at x.) Its Pontryagin powers can be represented by tori in ΩS^2, where each point of a k-torus is mapped into a loop that is obtained by going along the images of k corresponding meridians one after the other. Thus, given a meridianal sweep-out of M by loops of length $\leq L$ based at x, we obtain a geometric realization of generators of $H_k(\Omega M, \mathbf{R})$ by maps of k-tori, where each point is mapped into the loop on M made of k meridianal loops of M of total length $\leq kL$.

If the length functional is a Morse function, then those classes correspond to distinct critical points (of different indices), and we are done. In the degenerate situation Schwartz have proceeded as follows: For an *even-dimensional* homology class c he considered a dual cohomology class u of the same dimension and observed that cup powers of u are dual up to some constant factors to Pontryagin powers of u. Then he used a theorem proven by Lyusternik and Shnirelman asserting that if cohomology classes u and $u \cup v$, $v \neq 0$, correspond to the same critical value, then there exist a whole critical level (made of uncountably many geodesic loops) corresponding the the same critical value.

As one is using only the even-dimensional classes, one gets $2kL$ as the upper bound for the values of the length functional at first k non-trivial critical points.

1.2. Meridianal sweep-out via filling. Therefore, our first intention would be to construct a meridianal sweep-out of M by curves of length $10d$ that maps both poles of S^2 into x. In the case of success we would obtain infinitely many geodesic loops such the the length of the kth of them does not exceed $20kd$. To be more precise we will either obtain such a meridianal sweep-out by curves of length $\leq 10d + \varepsilon$, where $\varepsilon > 0$ is a parameter that we can make arbitrarily small (as this upper bound still yields the desired upper bound for the lengths of k geodesic loops based at x), or we discover that there exists a geodesic loop α and a non-trivial periodic geodesic β, both of length $\leq 2d + \varepsilon$ such that β is contained in a domain D bounded by α such that the angle of D at x is $\leq \pi$. It turns out that in this last case one obtains infinitely many distinct geodesic loops based at x as follows: First, one connects x inside D with a point $y \in \beta$ by a path τ of length $\leq 2d + \varepsilon$, then considers loops $\tau * \beta^i * \tau^{-1}$ formed by travelling along τ, then along β i times and then returning along τ^{-1}, (see Fig. 1). (Here i can be any positive integer number.)

Finally, one applies a Birkhoff curve-shortening process with fixed end-points to these curves. It turns out that the process terminates at different geodesic loops based at x for different values of i, (see Section 1.3 for the details).

To construct a meridianal sweep-out of S^2 we start from a diffeomorphism $F : S^2 \longrightarrow M$. We represent S^2 as the boundary of a 3-ball D^3. We consider a very fine triangulation of S^2 such that the images of the simplices under F are contained within the ball of a very small radius ε not exceeding $inj(M)/100$, where $inj(M)$ denotes the injectivity radius of M. Moreover, later we will pass to the limit as $\varepsilon \longrightarrow 0$. We triangulate D^3 as the cone over the chosen triangulation of S^2.

Now we are going to try to extend F from $S^2 = \partial D^3$ to the whole D^3. (Of course, there is no such extension.) For this purpose we map the center c of D^3 to x, then map all 1-dimensional simplices of D^3 that connect c with the vertices v_i of the chosen triangulation of S^2 into (some) minimal

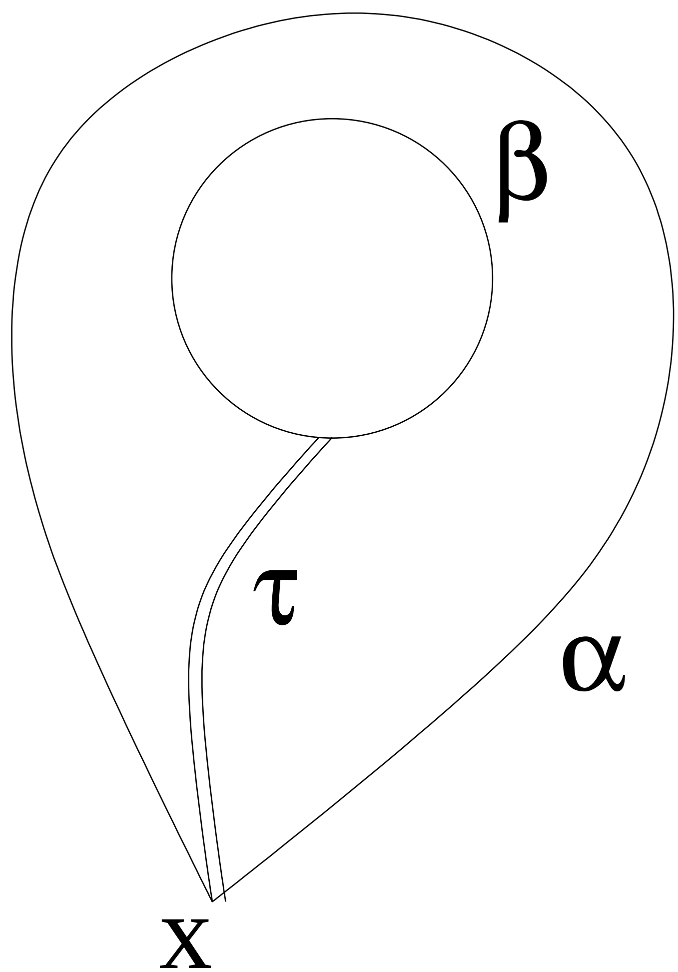

Figure 1: Periodic geodesic β is inside the domain bounded by a geodesic loop α. The angle at x is less than π

geodesics connecting x with $F(v_i)$.

Our next step will be to construct the extension to the 2-skeleton of the chosen triangulation of D^3. Once this step will be accomplished, we are going to have a collection of maps π_{ijk} of 2-spheres in M corresponding to the boundaries S^2_{ijk} of 3-simplices $[cv_iv_jv_k]$, where $[v_iv_jv_k]$ runs over all 2-simplices of the chosen triangulation of S^2. As we are not able to extend F to D^3, at least one of these 2-spheres is mapped into M by a map of a non-zero degree.

Let z_0 be a point in the center of the triangle $F([v_iv_jv_k])$, z_1, z_2, z_3 denote midpoints of geodesic segments $F(v_i)F(v_j)$, $F(v_j)F(v_k)$ and $F(v_i)F(v_k)$. Extend the geodesic segment $xF(v_i)$ by adding the minimal geodesic segment $F(v_i)z$, $xF(v_j)$ by adding the minimal geodesic segment $F(v_j)z$ and $xF(v_k)$ by adding the minimal geodesic segment $F(v_k)z$. Consider three loops obtained from three pairs of these three broken geodesics connecting x with z through $F(v_i), F(v_j)$ or $F(v_k)$. We are going to attempt to contract them to x as loops based at x through based loops of length $\leq 6d + O(\varepsilon)$. The desired 2-sphere will be obtained by gluing three maps of D^2 generated by these contractions.

We are going to describe how we will be contracting loop $l_1 = xF(v_i)zF(v_j)x$, (see Fig. 2).

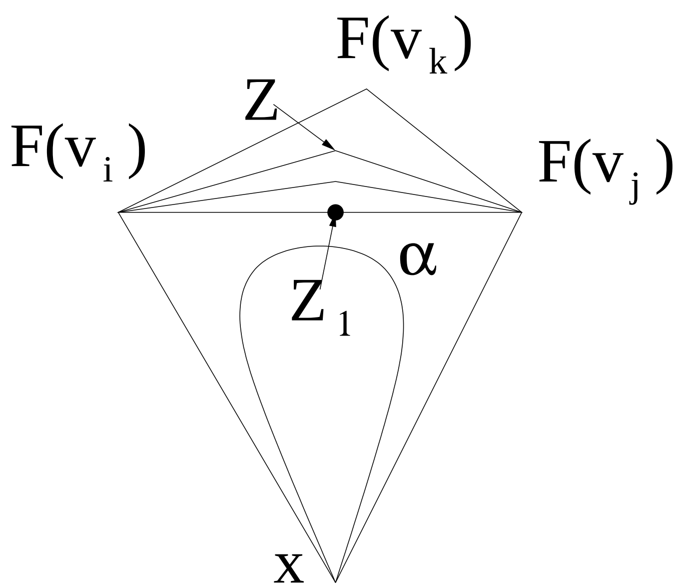

Figure 2: Contracting loop l_1

The contraction of two other loops will be performed in exactly the same way.

On the first stage we will contract this loop to $xF(v_i)z_1F(v_j)x$ through the small triangle $F(v_i)F(v_j)F(v_k)$. Now we need to contract the loop $l_{ij} = xF(v_i)F(v_j)x = xF(v_i)z_1f(v_j)x$ to a point. To achieve this goal we first apply the Birkhoff curve-shortening process with the fixed base point to this loop, (cf. [Cr] for a good description of the Birkhoff curve-shortening process and its properties for free loops. In the nutshell the Birkhoff curve-shortening process works as follows: Parametrize a given curve γ_0 by its arclength. Divide this curve into N intervals of equal length for some very large N. Connect the endpoints of these intervals by (unique) minimal geodesics. The resulting broken geodesic will be a curve $\gamma_{1/2}$. Now connect the midpoints of adjacent geodesic segments of $\gamma_{1/2}$ (including the first and the last segments) by minimal geodesics. These minimal geodesic will form a new curve $\gamma_1 = \beta(\gamma_0)$. Then we will be repeating this process, and inductively define γ_{i+1} as $\beta(\gamma_i)$. If N is sufficiently large, it is not difficult to 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]). In the version of this process for based loops we *do not* connect the midpoints of the first and the last geodesic segments, so that all curves will be loops based at x. 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. We always choose N sufficiently large to ensure that the rate does not exceed the injectivity radius of M, $inj(M)$, but sometimes below we will need to choose the rate to be very small.) This process ends at a point or at a non-trivial geodesic loop α based at x. Note that so far we obtained a one-parametric family of loops based at x of length $\leq 2d+O(\varepsilon)$. Without any loss of generality we can assume that α is not a periodic geodesic as in this case we are able to obtain the desired geodesic loops as iterates of α.

First, let us assume that α is a simple (i.e. nonself-intersecting) loop. We will amend our proof to encompass the case when the loop l_{ij} develops self-intersections during the Birkhoff curve-shortening process in Section 1.4.

Now let us apply the Birkhoff curve-shortening process for free loops to α. That is, we do not keep the base point x fixed anymore. This process ends either at a point y inside a domain bounded by α or at a non-trivial periodic geodesic β. In the second case note that β is contained inside a domain D on M bounded by α, and that the angle of D (i.e. of α) at x is less than π. (Cf. [Cr] for almost obvious details of the proof of this assertion.)

If we obtain a non-trivial periodic geodesic β for at least one of the loops that we are going to contract, we immediately stop our attempts to extend F to the 2-skeleton of the chosen triangulation of D^3. In this case we will obtain the desired geodesic loops using an entirely different idea described in Section 1.3 below.

Therefore we can assume that α contracts to a point y as a free loop. At this stage we already obtain the extension of F to the 2-skeleton of the chosen triangulation of D^3. However, our goal will be to find a meridional sweep-out of each sphere $xF(v_i)F(v_j)F(v_k)$ into loops (meridians) of length $\leq 8d + O(\varepsilon)$. As one of those spheres is mapped into M by a map of a non-zero degree, this completes the proof of the theorem.

For this purpose we are going to convert the free loop part of the process into a contraction of α to a point through based loops of length $\leq 6d+O(\varepsilon)$.

It is not difficult to see that the contraction α_t of $\alpha = \alpha_0$ to $y = \alpha_1$ is monotone, that is the domain bounded by α_s containing y contains all domains bounded by α_t and containing y for all $t > s$, (cf. [Cr]). It seems rather obvious that we can perform an arbitrarily small perturbation of this homotopy to make curves α_t disjoint. (We leave α intact but might need to perturb y.) Yet we are not going to prove or to use this assertion. Let us assume first that closed curve α_t already have this property that we are going to call *strict monotonicity*.

Let D denote the domain bounded by α in M that contains y. Note that the $dist_D(x, y) \leq 2d + O(\varepsilon)$ as we can connect y with the closest point $w \in \partial D$ by a geodesic of length $\leq d$ and then connect w and x along α by an arc of length $\leq d + O(\varepsilon)$. Let τ denotes a geodesic in D that connects x and y and is parametrized by $[0, 1]$, (see Fig. 3).

The strict monotonicity of the contraction implies that: 1) For every t there exists a unique $\lambda(t)$ such that $\tau(t) \in \alpha_{\lambda(t)}$; 2) $\lambda(t)$ continuously depends on t. For every t denote the arc obtained from τ by restricting to $[0, t]$ interval by τ_t, and the loop that first goes along τ_t, then along $\alpha_{\lambda(t)}$ and then returns to x along τ_t in the opposite direction by β_t. We can assume that β_t is parametrised by $[0, 1]$ proportionally to the arc length. Note that $\beta_0 = \alpha$ and β_1 is made of two copies of τ traversed in the opposite direction. We can extend the homotopy β_t by contracting these two copies of τ to x along $\tau_t * \tau_t^{-1}$. The result will be the desired path homotopy contracting α to a point via loops of length $\leq 6d + O(\varepsilon)$ based at x.

To deal with the general case, when the homotopy is only monotone, but not necessarily strictly monotone, observe that the convexity of all curves α_s for every t implies that the set $A(t)$ of all s such that $\tau(t) \in \alpha_s$ is

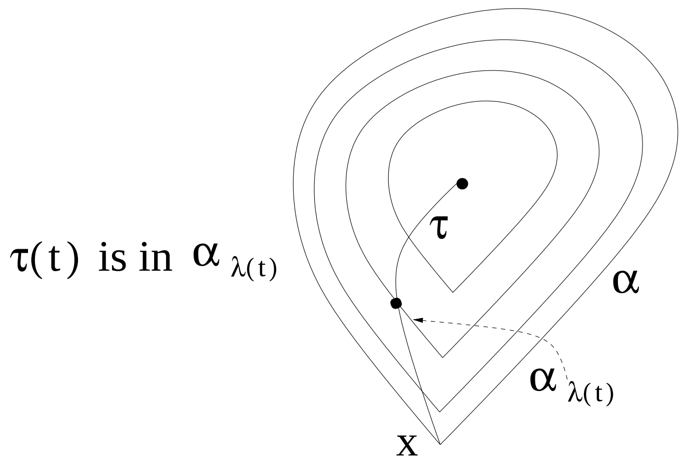

Figure 3: Transforming a free loop homotopy into a based loop homotopy

either a point $\lambda(t)$, or an interval $[\lambda_1(t), \lambda_2(t)]$. Moreover, the sets $A(t)$ are disjoint for different values of $t < 1$. (This assertion follows from the proof of Lemmae 6 and 7 in [M]). Therefore, we can modify the homotopy described in the strictly monotone case by using all curves obtained by travelling along τ_t, then along curves $\alpha_s, s \in [\lambda_1(t), \lambda_2(t)]$, and then along τ_t in the opposite direction in the case, when $A(t)$ contains more than one point. (If $A(1)$ has more than one point, then, when $t = 1$, we might need to start not with $s = \lambda_1(1)$ but with $s = \lim_{t \longrightarrow 1^-} \lambda(t)$, where the limit is taken over the set of t such that $A(t)$ contains only one point. In principle, this limit can be greater than $\lambda_1(1)$.)

Thus, we can assume that we contracted each of three loops l_1, l_2 and l_3 through loops based at x of length $\leq 6d + O(\varepsilon)$. It remains to merge these three 1-parametric families of loops into one continuous family of loops based at x of length $\leq 8d + O(\varepsilon)$ parametrized by S^1 to complete the proof of the theorem, as we can regard these loops as images of meridians of S^2 under a meridianal sweep-out of M.

We proceed as follows. We start at the constant loop x. We perform the path homotopy contracting l_1 in reverse to obtain l_1. Note that the lengths of all three loops l_i do not exceed $2d + O(\varepsilon)$. Assume that l_2 was contracted via based loops γ_t to a constant loop γ_1. Now we consider loops $l_1 * \gamma_{1-t}^{-1}$, $t \in [0, 1]$. This part of the homotopy starts at $l_1 = l_1 * \gamma_1^{-1}$ and ends at $l_1 * l_2^{-1} = l_1 * \gamma_0^{-1}$. Note that l_1 and l_2 have a common arc $xF(v_j)z$. After contracting this arc traversed in opposite directions over itself we obtain

l_3, which then can be contracted to the constant loop using the already constructed path homotopy. This completes the proof of Theorem 0.1 in the case when α is nonself-intersecting and can be contracted to a point by a Birkhoff curve-shortening process for free loops.

1.3. The case of a non-trivial periodic geodesic. Here we are going to prove Theorem 0.1 in the case when there exists a nonself-intersecting geodesic loop α of length $\leq 2d + O(\varepsilon)$ based at x and a non-trivial periodic geodesic β of a smaller length contained in a domain D bounded by α, and the angle of D at x is $< \pi$, (see Fig. 1).

Consider the closed domain $T \subset D$ contained between α and β. Denote a shortest curve in T connecting x and a point of β by ϱ. The length of ϱ does not exceed $2d + O(\varepsilon)$. Indeed, denote the other endpoint of ϱ by q. Let q_0 denote the last point of intersection of a minimal geodesic g connecting x and q in M with α. The length of ϱ does not exceed the length of a curve obtained by going from x to q_0 along the shortest arc of α, and then to q along g. For every positive integer m consider loops γ_m based at x and obtained by going from x to q along ϱ, then going along β m times, and finally returning to x along ϱ.

Recall that Birkhoff curve-shortening process involves subdividing a curve into segments of a small length $\leq s < inj(M)$, replacing these segments by minimal geodesics, then connecting the midpoints of these segments by another set of the minimal geodesics, etc. Once the parameter s (that we call the rate of the process) is chosen, we never need to connect points that are situated at a distance greater than s from each other by a minimal geodesic. As the result, we obtain a sequence of very close closed curves which then can be connected by obvious homotopies filling added (or subtracted) "triangles". Theorem 0.1 immediately follows from the following lemma:

Lemma 1.2 *Consider geodesic loops g_m obtained from γ_m by the application of a Birkhoff curve-shortening process with fixed basepoint x. If the rate s of the process is sufficiently small, then these geodesic loops are distinct for different m.*

Proof. Let a denote a point inside the subdomain of D bounded by β. The absolute value of the winding number of γ_m around a is equal to m. Clearly, this number does not change during any homotopy of γ_m in T.

Thus, lemma follows from the following key fact: The loops obtained during the Birkhoff curve-shortening process applied to g_m stay inside the closed domain T.

To prove this observation notice that there exists a positive number $s(\alpha)$ such that if $r_1, r_2 \in T$ are two points such that $dist_M(r_1, r_2) < \min\{inj(M),\ s(\alpha)\}$, then the minimal geodesic connecting r_1, r_2 is contained in T. This geodesic cannot transversely intersect β or α "far" from x as in this case two points of intersection would be connected by two minimal geodesics, namely, an arc of β (or α) and a segment of $r_1 r_2$, which yields a contradiction.

It remains to note the following well-known fact: There exists a positive $s(\alpha)$ such that for every two points x_1, x_2 on α $s(\alpha)$-close to x the minimal geodesic that connects them lies in the closure of D. Indeed, it cannot intersect α at other points due to the uniqueness of minimal geodesics connecting sufficiently close points of M. And it cannot be contained in the closure of $M \setminus D$ as a consequence of the fact that the angle of D at x is less than π, and the angle of its complement at x is greater than π.

Therefore every step of the Birkhoff curve shortening process yields closed curves that are contained in T, which completes the proof of the lemma and Theorem 0.1. □

1.4. The general case. In Sections 1.2 and 1.3 we dealt with the case when a loop l_{ij} formed by two minimizing geodesics from x to two very close points $F(v_i)$, $F(v_j)$ and the minimal geodesic segment $\nu = F(v_i)F(v_j)$ can be contracted to a *nonself-intersecting* geodesic loop based at x through loops based at x of length not exceeding the length of l_{ij}.

As we mentioned in Section 1.2, we will now consider the general case, and intend to achieve either one of the following two goals:

1.4.1. To contract l_{ij} to a point through loops based at x of length $\leq 8d + o(1)$ (which then can be merged into a meridianal sweep-out of M by curves of length $\leq 10d + o(1)$ exactly as it was described in the last paragraph of Section 1.2);

1.4.2. To find a simple geodesic loop α based at x and a non-trivial simple periodic geodesic β contained in a domain D bounded by α such that (a) The lengths of α and β do not exceed $2d + o(1)$; (b) The angle of D at x is less than π (or equivalently, α is convex to D in terminology of [Cr]. A closed curve α that bounds D is convex to D if the minimal geodesic in M that connects each pair of sufficiently close points in α is contained in D. We use the notation $o(1)$ for terms that can be made arbitrarily small by choosing arbitrarily fine triangulation of M and a very small rate of a Birkhoff curve-shortening process used in the course of our construction.)

Again we are going to attempt to contract all loops l_{ij}. Our first step will be to reduce contracting the triangle l_{ij} to contracting several of geodesic digons. For this purpose consider minimal geodesics connecting x with points of the interval $F(v_i)F(v_j)$, (see Fig. 4). This interval can be

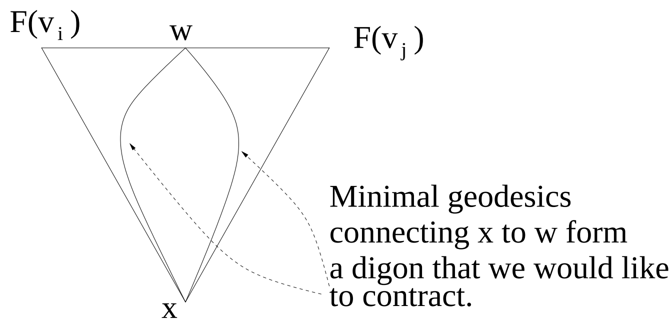

Figure 4: Contracting loops can be reduced to contracting geodesic digons

subdivided into open intervals I, where the minimal geodesic from x to a point $w \in I$ is unique, and continuously varies with w. Moreover, the continuous family of minimizing geodesics connecting x with points of I can be extended to endpoints of I. However, the minimal geodesics from two continuous families corresponding to two adjacent open intervals meeting at a common endpoint O can be distinct and form a minimal geodesic digon xO. Our goal is to construct a continuous one parametric family of curves connecting x with all points of $F(v_i)F(v_j)$ of length $\leq 7d + o(1)$. Once this goal is achieved, we immediately get the desired contraction of l_{ij} via loops of length $\leq 8d + o(1)$. But it is sufficient to learn to contract the digons xO via loops of length $\leq 6d + o(1)$ based at x in order to achieve this goal.

Consider one of these digons xO formed by two minimal geodesics λ_1 and λ_2 connecting x and O. As it had been already noticed, we can assume that this digon is not a periodic geodesic. It divides M into two domains. Denote the domain with the angle at O less than π by D_1. If D_1 has the angle at x less than π as well, then the Birkhoff curve-shortening process will contract xO as a based at x loop inside D_1 and via convex curves. The homotopy will end either at a point or at a nonself-intersecting geodesic loop α (because of the convexity of the boundary D_1 - cf. [Cr] for the details). Then we can continue as in Sections 1.2, 1.3.

However, if D_1 has the angle at x that is greater than π, then we are not guaranteed that the curve will not develop a self-intersection during the Birkhoff curve-shortening process. To be more precise note that the curve will be contracted inside D_1 via curves that are convex to D_1 at all points other than x until a self-intersection develops - if it develops at all. The only possibility for development of a self intersection is the case, when an arc of the curve will come close to x during the homotopy, and two points on this arc will be connected by a geodesic segment cutting through both the initial and final segments of the curve very closely to x, (see Fig. 5).

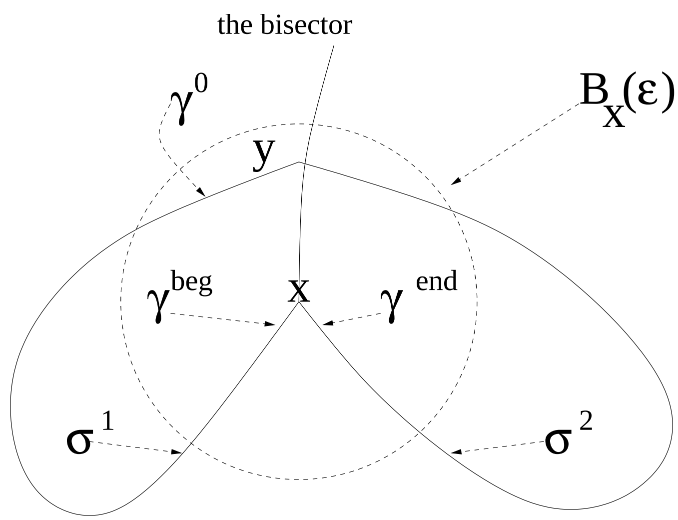

Figure 5: The case when the angle at x is greater than π

Denote the closed curve that appears during the Birkhoff curve-shortening process right before the intersection is formed on the next stage of the process by γ_t.

By choosing the rate of the Birkhoff curve shortening process very small we can ensure that the metric ball of radius $\varepsilon < inj(M)$ centered at x will intersect γ_t via two arcs γ^{beg} and γ^{end} that meet at x as well as another arc γ^0 that does not intersect them. Here we can make ε as small as we wish by choosing the rate of the Birkhoff curve-shortening process to be sufficiently small. A crucial observation now is that the geodesic ray starting at x and bisecting the angle of D_1 at x must intersect γ_0. Otherwise, γ_0 will be contained inside a convex curve formed by the boundary of the metric ball and either the bisector and one of the arcs γ^{beg}, γ^{end}, or by both arcs γ^{beg}, γ^{end}. In this case the minimal geodesic connecting any pair of points of

γ^0 will be in this convex domain and cannot intersect γ^{beg} and γ^{end}, so, contrary to our assumptions, the self-intersection will not be formed on the next stage of the Birkhoff process.

Let y denotes the point of intersection of the bisector of the angle at x with γ_0. There are two arcs of γ_t between x and y. Denote these arcs by γ^1 and γ^2, and their lengths by l_1, 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_1 that it bounds, as the angles at x and y are less than π, and γ_t was convex to D_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 $\leq 4d + l_i + o(1)$ for $i = 1, 2$. (As usual, $o(1)$ term contains quantities that we can make arbitrarily small.) 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 $\leq l_1 + l_2 + 4d + o(1) \leq length(\gamma_t) + 4d + o(1) \leq 6d + o(1)$. Indeed, we can first insert the segment xy travelled twice in the opposite direction, then contract σ_1 (as based loop) and the contract σ_2.

Now note that we can homotope σ_i first to a nonself-intersecting geodesic loop based at x by a path homotopy, and then either to a point or a non-trivial simple periodic geodesic by the application of a curve-shortening process. If we obtain a non-trivial periodic geodesic, then we find ourselves in the situation of 1.4.2. Otherwise, we can proceed exactly as in Section 1.2 to obtain a path homotopy that contracts σ_i via loops based at x of length $\leq l_i + 4d + o(1)$ as desired. □

2 Geodesics between distinct points.

The purpose of this section is to prove the following theorem:

Theorem 2.1 *Let M be a Riemannian manifold diffeomorphic to S^2. Let d denote the diameter of M, and let $x, y \in M$ be two points of M such that $dist(x, y) = d$. Then for every $k > 3$ there exist k distinct geodesics of length $\leq (12k - 37)d$ connecting x and y for every positive integer k.*

Remark 2.2. Classical Berger's lemma implies that x and y are connected by at least two distinct minimal geodesc segments. Moreover, if they are connected by exactly two minimal geodesic segments, then these two segments form a periodic geodesic γ of length $2d$. In this last case one can construct infinitely many distinct geodesics between x and y by going along γ a variable number of times, and then going to y along γ. In particular, the length of the third geodesic connecting x and y does not exceed $3d$.

Proof. C. Croke observed that if x, y, M are as in the text of the theorem, then there exists a finite number N of minimal geodesic segments l_i connecting x and y such that they divide M into digonal domains D_i with angles at x and $y \leq \pi$ (see [Cr]), ($\partial D_i = l_i \bigcup l_{i+1}$, $l_{N+1} = l_1$). Therefore the Birkhoff curve-shortening process contracts the boundary of D_i either to a simple periodic geodesic inside D_i or to a point inside D_i.

In the case, when one of those geodesic digons α_i is contracted to a non-trivial periodic geodesic β we can proceed as follows: Connect x with a closest point of β by a geodesic (in M) that we will denote τ_1. Note that τ_1 must be in D_i as the boundary of D_i is formed by two minima geodesics from x to y which because of their minimality cannot intersect τ_1. Therefore the length of τ does not exceed d. Similarly connect y with a closest point of β by a geodesic τ_2. Now consider paths $\gamma_k = \tau_1 * \beta^k * \tau_2^{-1}$ for all $k = 1, 2, \ldots$. Procceding as in the proof in Section 1.3 we can prove that the Birkhoff curve-shortening process *with fixed end points* produces distinct geodesics connecting x and y as it happens inside D_i, and, therefore, does not change the winding number with respect to a point inside the subdomain of D_i bounded by β. (In order to define the winding number we can transform these paths into closed curves by attaching a minimal geodesic from y to x forming a part of the boundary of D_i.)

Assume now that for every i the Birkhoff curve-shortening process contracts the boundary d_i of D_i to a point $y_i \in D_i$ through a monotonous (in the sense of section 2.2) family of curves d_{it}. Connect x and y_i by a minimal geodesic τ. This geodesic cannot intersect the boundary of D_i as it is formed by two minimal geodesic from x to y. Therefore, τ is contained in D_i. Now we transform the family of curves d_{it} into a strictly monotonous family of curves that we will denote d_t. Consider loops ϱ_t formed by following τ for time t, then travelling around a curve from the family d_t, and then returning back to x along τ. The lengths of these loops do not exceed $4d$. The last of them is formed by tracing τ from x to y, and then returning back to x along τ. It can be contracted to x by gradually cancelling its longer and longer segments. So, we obtain a homotopy contracting d_i via loops of length $\leq 4d$ based at x. Denote loops in this homotopy by b_i^t, $t \in [0, 1]$. Without any loss of generality we can assume that the boundary of D_i is oriented as follows: One first goes from x to y along l_{i+1} and then returns along l_i to x. Now we can construct a meridional sweep-out of M such that one of the poles is mapped to x, another to y, and the lengths of the meridians are bounded by $5d$ as follows: Start from l_1. For every t consider $b_1^{1-t} * l_i$. Of course, $b_1^1 * l_1 = l_1$ and $b_1^0 * l_1 = l_2 * l_1^{-1} * l_1$, where l_1^{-1} denotes l_1 travelled in the opposite direction, that is from y to x. Now cancel $l_1^{-1} * l_1$ alomg itself. We end up with l_2. So, we constructed a path homotopy between l_1 and l_2. Now we can construct a path homotopy between l_2 and l_3 using b_2^{1-t} in exactly the same way. Then we proceed by induction constructing apth homotopies between l_i and l_{i+1} using b_i^{1-t} for $i = 3, \ldots, N$ and end up with $l_{N+1} = l_1$. Combining those path homotopies we obtain a meridional sweep-out f of M by curves of length $\leq 5d$, where one of the poles is mapped into x and the other into y.

Now one can homotope this sweep-out into another sweep-out that will map both poles of S^2 into x and where every meridian will be mapped into the join of its image under f and l_1. The lengths of curves in this sweep-out will not exceed $6d$. Now we use the even-dimensional homology classes of $\Omega_x M$ of the space of loops based at x and the well-known fact that attaching any path (e.g. l_1) from x to y to all loops based at x yields a homotopy equivalence from $\Omega_x M$ to the space $\Omega_{x,y} M$ of paths from x to y.

In order to get k distinct geodesics between x and y we can use $N \geq 3$ minimal geodesics between x and y. (If $N = 2$, then there is a periodic geodesic γ through x and y, and the geodesic segments $\gamma^i * l_1$ will satisfy the conditions of the theorem.) Then we will need $k - N \leq k - 3$ geodesics of non-zero indices corresponding to the homology classes $H_{2m}(\Omega_{x,y}, R)$ of non-zero even degrees. Note that the lengths of paths in the constructed explicit realizations of these cycles do not exceed $(2m)(6d)+d$, but each path ends by two copies of l_1 travelled in different directions. We can cancel these copies and will obtain families of paths from x to y of length $\leq (12m - 1)d$. Thus, the lengths of these $k-3$ geodesics connecting x and y do not exceed $(12(k - 3) - 1)d = (12k - 37)d$ for $k > 3$. □

Remark 2.3. It was essential in this proof that in the case, when the Birkhoff curve-shortening process ends at a periodic geodesic x and y are on the same side of this geodesic. This prevents us from directly generalizing our proof for the case of arbitrary x and y.

Acknowledgements: Both authors acknowledge a partial support of this work by their corresponding NSERC Discovery Grants.

References

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

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

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

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

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

[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

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

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

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

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