Length of Geodesics and Quantitative Morse Theory on Loop Spaces.

Alexander Nabutovsky, Regina Rotman

October 30, 2012

Abstract

Let M^n be a closed Riemannian manifold of diameter d. Our first main result is that for every two (not necessarily distinct) points $p, q \in M^n$ and every positive integer k there are at least k distinct geodesics connecting p and q of length $\leq 4nk^2 d$.

We demonstrate that all homotopy classes of M^n can be represented by spheres swept-out by "short" loops unless the length functional has "many" "deep" local minima of a "small" length on the space $\Omega_{pq}M^n$ of paths connecting p and q. For example, one of our results implies that for every positive integer k there are two possibilities: Either the length functional on $\Omega_{pq}M^n$ has k distinct non-trivial local minima with length $\leq 2kd$ and "depth" $\geq 2d$; or for every m every map of S^m into $\Omega_{pq}M^n$ is homotopic to a map of S^m into the subspace $\Omega_{pq}^{4(k+2)(m+1)d}M^n$ of $\Omega_{pq}M^n$ that consists of all paths of length $\leq 4(k+2)(m+1)d$.

1 Main results.

One of the goals of this paper is to prove an effective version of a famous theorem published by J.P. Serre in 1951 ([Se]) that asserts that for every pair of points on a closed Riemannian manifold there exist infinitely many distinct geodesics connecting these points. Here and below two geodesics or geodesic loops are regarded as distinct if they do not differ by a reparametrization.

In our paper [NR0] we have conjectured that there exists a function $f(k, n)$ such that for every positive integer k and every pair of points p, q on a closed n-dimensional Riemannian manifold of diameter d there exist at least k distinct geodesics connecting p and q of length $\leq f(k, n)d$.

In the present paper we prove this conjecture for $f(k,n) = 4k^2n$. We first prove it in the case of simply connected manifolds. The general case will then easily follow.

The starting point will be a proof of Serre's theorem by Albert Schwarz ([Sc]). In this paper Schwarz also demonstrates that the length of the kth geodesic can be bounded above by $C(M^n)k$, where $C(M^n)$ does not depend on k but only on the Riemannian manifold M^n. (This estimate was later improved by M. Gromov in section 1.4 of [Gr0] in the situation, when p and q are not conjugate allong any geodesic. Gromov proved that in this case the number of geodesics of length $\leq x$ connecting p and q is at least the sum of Betti numbers $b_i(\Omega_p M^n)$ over i ranging from 1 to $[c(M^n)x]$ for an appropriate constant $c(M^n)$. Although for some manifolds (e.g. S^n) this still provides only a linear upper bound in k for the length of the kth shortest geodesic between p and q, for "many" manifolds the sum of the Betti numbers of the loop space grows exponentially in x, and one obtains a logarithmic upper bound in k for the length of the kth shortest geodesic.)

Here is a somewhat modernized sketch of the proof of Serre's theorem given by A.Schwarz, where we make references to the rational homotopy theory (that did not exist when this proof was invented). Let us consider the space $\Omega_p M^n$ of loops based at p on a closed simply-connected Riemannian manifold M^n. One would like to show that the sum of its Betti numbers is infinite. Then the existence of infinitely many geodesic loops based at p would follow from a standard Morse-theoretic argument.

First, note that there exists a non-trivial even-dimensional real cohomology class of the loop space $\Omega_p M^n$. To prove its existence observe that there exist a non-zero rational homotopy class of M^n of an *odd* dimension N. Otherwise, the rational homotopy theory would immediately imply that the sum of Betti numbers of M^n in infinite, which is impossible. Therefore, there exists a non-zero rational homotopy class of $\Omega_p M^n$ of even dimension $N-1$. Now, the Cartan-Serre theorem (cf. [FHT], Theorem 16.10) implies the existence of a non-trivial real cohomology class of $\Omega_p M^n$ of dimension $N-1$.

Denote a non-trivial real cohomology class of $\Omega_p M^n$ of the smallest positive dimension by u. The Cartan-Serre theorem implies that the cup powers u^i are non-trivial, and therefore the sum of Betti numbers of $\Omega_p M^n$ is infinite. Applying Morse theory one obtains a critical point of the length functional corresponding to each power of u. If the critical points are not distinct, i.e. there is a critical point corresponding to u^i and u^j for $i \neq j$, the standard Lyusternik-Schnirelman argument, (see [Kl]), implies that the critical level that corresponds to u^i contains a set of critical points of dimension $\geq dim\ u > 0$, implying the existence of infinitely many geodesic loops based at p. (Schwarz also noticed that such a degenerate situation cannot occur at all, if $dim\ u \geq n$, as the dimension of the set of all geodesics between p and q cannot exceed $n-1$.)

Thus, it is enough to consider the situation when the critical points are distinct. Note also that an easy argument involving the basics of rational homotopy theory implies that the dimension of u is not greater than $\leq 2n-2$.

Now recall that the Pontryagin product in the rational homology group of the loop space is the product induced by the geometric product $\Omega_p M^n \times \Omega_p M^n \longrightarrow \Omega_p M^n$. (By the geometric product of two loops α and β we just mean their join $\alpha * \beta$.) To estimate the length of the geodesics corresponding to u^i Schwarz defines a "dual", (meaning $\langle u, c\rangle = 1$), homology class c of u of the same dimension. Then he proves that for every positive i the ith Pontryagin power of c and a rational multiple of u^i are dual. So, the critical point corresponding to u^i also corresponds to c^i.

One can now see that in order to estimate lengths of geodesic loops based at p it is enough to find a representative of c that is is contained in the set of loops based at p of length $\leq L$ for some L. Then c^i can be represented by a chain contained in the set of loops of length $\leq iL$.

To obtain an upper bound for geodesics connecting distinct points $p, q \in M^n$, one considers an explicit homotopy equivalence $h : \Omega_p M^n \longrightarrow \Omega_{p,q} M^n$ that is constructed by fixing a minimizing geodesic between p and q and attaching it at the end of each loop based at p. Then $h_*(u^i)$ can be represented by a chain contained in the set of paths of length $\leq iL + d$ between p and q, whence the length of the ith shortest geodesic between p and q does not exceed $iL + d$.

It is natural to make a conjecture that the length of a "kth-shortest" geodesic between two arbitrary points p, q on an arbitrary closed Riemannian manifold M^n should not exceed kd, where d is the diameter of M^n. Indeed, this conjecture is obviously true for round spheres. On the other end of the spectrum the conjecture is true for closed Riemannian manifolds with torsion-free fundamental groups (Proposition 2 in [NR0]). Yet this conjecture was disproved by a recent example of F. Balacheff, C. Croke, M. Katz ([BCK]). They have proved the existence of Zoll metrics on the 2-sphere that are arbitrarily close to the round metric and for which the length of a shortest periodic geodesic, (and thus, trivially, a shortest non-trivial geodesic loop based at any point) is greater than twice the diameter of the Zoll sphere. As a shortest non-trivial geodesic loop is a second shortest geodesic from its base point to itself, this example shows that the conjecture is false even if $n = k = 2$, the Riemannian manifold is convex and arbitrarily close to a round 2-sphere, and $p = q$ is an arbitrary point of the manifold.

Our proof of the upper bound that is quadratic in k works as follows. We demonstrate that for every l there are two classes of Riemannian metrics on each closed manifolds: "nice" metrics, where for every m every m-dimensional homotopy class of the manifold can be "swept-out" by "short" loops (of length not exceeding $\sim lmd$), and "bumpy" metrics, where the length functional on every space of all paths connecting a pair of points has l ("deep") local minima of a controlled length. If a Riemannian metric is very "nice", then one immediately obtains an upper bound for the lengths of N distinct geodesic loops linear in $lmdN$ from the proof of Serre's theorem by Schwarz. If the metric is very "bumpy", then one immediately obtains many short geodesic loops from the definition of "bumpiness".

The case when our estimate becomes quadratic in k, is the case of Riemannian metrics that are neither "bumpy" enough, nor "nice" enough, so that there are approximately $l = \frac{k}{2}$ "deep" local minima of the length functional on $\Omega_p M^n$ with lengths $\leq 2ld$. These $\frac{k}{2}$ local minima could prevent us from sweeping-out the cycle of interest by loops of length smaller than $c(n)kd$ (for an appropriate $c(n)$). As the result the bound for the length of the longest of remaining $\frac{k}{2}$ geodesic loops that follows from the proof of J.-P. Serre's becomes quadratic in k.

Although we do not have any actual examples of families of Riemannian metrics demonstrating that the quadratic dependence of our estimate on k is optimal, we believe that they exist - at least in dimensions > 3. So, we think that, in general, there is no upper bound for the length of the k shortest geodesic loop based at a prescribed point of the form $f(k,n)d$, where f grows slower than a quadratic function of k. However, in [NR4] we proved that, if $n = 2$, then there exists a linear in k upper bound for the length of the kth shortest geodesic loops based at an arbitrary point. (If M^n is diffeomorphic to S^2, then our upper bound for the length of the kth shortest non-trivial geodesic loop is $20kd$. Our proof is heavily based on the two-dimensionality of the manifold.)

Note also that even in the case of a 2-sphere one cannot hope to find a sweep-out of the cycle c from Schwarz's proof of Serre's theorem by "short" loops due to a counterexample of S. Frankel and M. Katz ([FK]), who found a family of Riemannian metrics on the 2-disc with uniformly bounded diameter and the length of the boundary but such that for every fixed value of τ it is impossible to contract boundaries of each of these 2-discs via closed curves of length $\leq \tau$. Taking the doubles of these 2-discs one obtains a family of Riemannian metrics on S^2 with uniformly bounded diameter that do not admit sweep-outs into loops with uniformly bounded lengths. This fact has recently been proven by Y. Liokumovich ([Li]).

We will, however, demonstrate that sweep-outs by short loops can only be obstructed by the existence of many short geodesic loops at each point of a manifold.

To state the first of our main results denote the space of loops of length $\leq L$ based at p on M^n by $\Omega_p^L M^n$.

Theorem 1.1 *Let M^n be a closed Riemannian manifold of dimension n and diameter d, p a point of M^n, k a positive integer number, and ν an aritrarily small positive real numer. Then either:*
1) There exist non-trivial geodesic loops based at p with lengths in every interval $(2(i-1)d, 2id]$ for $i \in \{1, 2, ..., k\}$. Moreover these geodesic loops are local minima of the length functional on $\Omega_p M^n$;
or
2) For every positive integer m every map $f : S^m \longrightarrow \Omega_p M^n$ is homotopic to a map $\tilde{f} : S^m \longrightarrow \Omega_p^{((4k+2)m+(2k-3))d+\nu} M^n$. Furthermore, every map $f : (D^m, \partial D^m) \longrightarrow (\Omega_p M^n, \Omega_p^{((4k+2)m+(2k-3))d} M^n)$ is homotopic to a map $\tilde{f} : (D^m, \partial D^m) \longrightarrow \Omega_p^{((4k+2)m+(2k-3))d+\nu} M^n$ relative to ∂D^m. In addition, if for some L the image of f is contained in $\Omega_p^L M^n$, then the homotopy between f and $\tilde{f}$ can be chosen so that its image is contained in $\Omega_p^{L+(4k+2)md+\nu} M^n$. Also, in this case for every L every map f from S^0 to $\Omega_p^L M^n$ is homotopic to a map $\tilde{f}$ from S^0 to $\Omega_p^{(2k-1)d} M^n$ by a homotopy such that its image is contained in $\Omega_p^{L+2d} M^n$.

If the Riemannian manifold M^n is analytic, then for each S there exists $\nu(S) > 0$ such that $\Omega_p^S M^n$ is the deformation retract of $\Omega_p^{S+\nu} M^n$. (Indeed, a standard argument, where one first replaces the space of all loops by a finite-dimensional space of piecewise geogesic loops and then uses the sub-analiticity of the distance function implies that there is no infinite sequence of geodesic loops based at p with lengths that strictly decrease and tend to S.) Therefore, in the analytic case one can drop ν in the text of this theorem. Moreover, if there is no positive ν such that $\Omega_p^S M^n$ is the deformation retract of $\Omega_p^{S+\nu} M^n$, then there is an infinite sequence of distinct geodesic loops based at p such that the sequence of their lengths converges to S.

From now on we would like to adopt the following convention. Instead of saying that for some S and every positive ν there exists a map into $\Omega_p^{S+\nu}M^n$ with some desirable for us properties, we will say that there exists a map into $\Omega_p^{S+o(1)}M^n$ with the desirable properties.

The previous theorem almost immediately leads to a quadratic bound for the lengths of geodesic loops based at p. Indeed, suppose that for some $s<k$ there are $s-1$ non-trivial geodesic loops based at p with lengths in the intervals $(0,2d],(2d,4d],...,(2(s-2)d,2(s-1)d]$, but no geodesic loops based at p of length in the interval $(2(s-1)d,2sd]$. Then there either exists a representation of an even-dimensional cycle c in the loop space that appears in the proof of Serre's theorem given by A. Schwarz by a spherical cycle that can be formed only by loops of length at most $((8n-6)s+(4n-7))d$ based at p, or for each $\nu>0$ there exists infinitely many geodesic loops based at p of length $\leq((8n-6)s+(4n-7))d+\nu$. In the first case we obtain at least s geodesic loops based at p of length $\leq 2(s-1)d$ (including the trivial loop), and either $s+1$ loops of length $\leq((8n-6)s+(4n-7))d$, $s+2$ loops of length $\leq 2((8n-6)s+(4n-7))d$,..., $s+i$ loops of length $\leq i((8n-6)s+(4n-7))d$,..., k loops of length $\leq(k-s)((8n-6)s+(4n-7))d$. This expression attains its maximum at $s=\lfloor\frac{k}{2}\rfloor$. The maximal value is $((2n-\frac{3}{2})k^2+(2n-\frac{7}{2})k-(1-(-1)^k))d$. Denote this value by $L(n,k,d)$. Note that none of the cycles c^i from the proof of Serre's theorem by A. Schwarz can "hang" at a local minimum of the length functional on Ω_pM^n or at the same critical point as c^j for some $j\neq i$, unless there is a critical level of a dimension $\geq dim\ c$ but $\leq n-1$ at this critical level. (In this last case one of the critical points will be "lost" due to the coincidence, but we will immediately get infinitely many distict geodesics of the same length, which results in a much better upper bound for the length.) Therefore wilthout any loss of generality we can assume that these geodesic loops are distinct. Thus, we are guaranteed to have at least k distinct geodesic loops based at a point p of length $L(n,k,d)$. (including the trivial loop). In the second case we will obtain even a better estimate for the lengths of the first k geodesic loops based at p providing that $k\geq 3$. Yet, if $k=2$, then it is known that there exists two distinct geodesic loops based at p of length $\leq 2nd$ ([R]). Thus, one obtains the following theorem in the case when M^n is simply-connected, and $p=q$:

Theorem 1.2 *Let M^n be a closed Riemannian n-dimensional manifold with diameter d. Then for every point $p\in M^n$ there exist at least k distinct geodesic loops of length at most $L(n,k,d)=((2n-1.5)k^2+(2n-3.5)k-$*

$(1-(-1)^k))d < 2n(k^2+k)d$. More generally, for each pair of points $p, q \in M^n$ there exist at least k geodesics starting at p and ending at q of length $L(n,k,d)+(2n-1.5)kd(p,q)$, if k is even, and $L(n,k,d)+(2n-1.5)(k+1)d(p,q)$, if k is odd. (Here $d(p,q)$ denotes the distance between p and q in M^n.) In both cases this upper bound does not exceed $((2n-1.5)k^2+(4n-5)k+(2n-3.5))d < 2n(k+1)^2d$.

Remark. Denote the smallest *odd* number l such that there exists a non-trivial *rational* homotopy class of M^n by l. An elementary rational homotopy theory argument (cf. [FHT]) implies that $l \leq 2n-1$. Our proof of Theorem 1.2 implies upper bounds $L(\frac{l+1}{2},k,d) = ((l-0.5)k^2+(l-2.5)k-(1-(-1)^k))d$ for the lengths of k distinct geodesic loops based at an arbitrary point p of M^n. Similarly for arbitrary $p, q \in M^n$ and arbitrary k there exist at least k distinct geodesics of length not exceeding $L(\frac{l+1}{2},k,d)+(l-0.5)kd(p,q)$, if k is even, and $L(\frac{l+1}{2},k,d)+(l-0.5)(k+1)d(p,q)$, if k is odd. These estimates do not exceed $((l-0.5)k^2+(2l-3)k+(l-2.5))d < l(k+1)^2d$. Note that in [NR3] we proved a version of Theorem 1.2 in the case $l=3$, but with a worse upper bound that depended factorially on k.

Note also that $2n(k+1)^2d < 4nk^2d$ for all $k \geq 3$, and that we have a better bound $2nd (< 4nk^2d)$, when $k=2$, proven in [NR1], Therefore, if desired, one can replace the upper bounds for the lengths of k shortest geodesics between p and q in M^n provided by Theorem 1.2 by a simpler looking expression $4nk^2d$.

To prove Theorem 1.2 in the case when M^n is simply-connected, but $p \neq q$, we prove a generalization of Theorem 1.1, where Ω_pM^n is replaced by the space $\Omega_{p,q}M^n$ (Theorem 5.3). It immediately yields Theorem 1.2 in the case when $p \neq q$, exactly as Theorem 1.1 implied the case $p=q$.

To obtain Theorem 1.2 in the nonsimply-connected case we will consider the universal covering of M^n constructed from the space of all paths starting at p via the standard identification and endowed with the pull back Riemannian metric. According to the standard argument that can be found in many textbooks on Riemannian geometry one can choose the fundamental domains in the universal covering so that their interiors are all isometric to the complement of the cut locus of the base point p, and, therefore, their diameter does not exceed $2d$. If the cardinality of $\pi_1(M^n)$ is infinite or finite but $\geq k$, we will connect the base point $\tilde{p}$ in the universal covering $\tilde{M}^n$ of M^n with k closest liftings of q by shortest geodesics. The projections of these geodesics to M^n will have lengths $\leq (2k-1)d$, and the theorem follows. If the cardinality of $\pi_1(M^n)$ is less than k, then we observe that $\tilde{M}^n$ is a simply-connected manifold of diameter $\tilde{d} \leq 2|\pi_1(M^n)|d$ (as the diameter of each fundamental domain does not exceed $2d$). Let k_s denote the smallest integer number which is not less than $\frac{k}{|\pi_1(M^n)|}$. We are going to connect $\tilde{p}$ with each lifting of q by k_s or $k_s - 1$ distinct geodesics, so as to obtain the required number k of distinct geodesics between p and q after projecting down to M^n. (Obviously, if we need to connect $\tilde{p}$ with some points in the lifting of q to $\tilde{M}^n$ with k_s geodesics, and with some other points in the lifting of q with $k_s - 1$ geodesics, we choose points that we connect with $\tilde{p}$ by k_s geodesics to be the points that are the closest to $\tilde{p}$.) If one knows how to prove Theorem 1.2 in the simply-connected case, then one can get a slightly worse upper bound (but still with the leading term $\frac{2}{|\pi_1(M^n)|}(2n-1.5)k^2d \leq (2n-1.5)k^2d)$ in the nonsimply-connected case. (Indeed, asymptotically k^2 will be divided by $|\pi_1(M^n)|^2$ and multiplied by $2|\pi_1(M^n)|$.) To prove a better upper bound we will need the following:

Theorem 1.3 *Let M be a closed Riemannian manifold of diameter d with a finite fundamental group of cardinality C. The the diameter of the universal covering space $\tilde{M}$ of M endowed with the pull back Riemannian metric does not exceed Cd.*

It is hard to believe that Theorem 1.3 is not known, yet we were not able to find any mention of it in the literature. Therefore we will prove it in Section 6 of this paper. (Remark: After we have completed the first draft of this paper we mentioned Theorem 1.3 to Anton Petrunin. Anton observed that our proof is valid not only for universal coverings but for all C-fold regular coverings. He asked if this theorem can be generalized for (not necessarily regular) finite coverings and posted this question at Mathoverflow. Soon afterwards Sergei Ivanov proved that this is, indeed, the case. His proof is available at /mathoverflow.net/questions/7732/diameter-of-m-fold-cover . Anton also conjectured that the factor C in the upper bound Cd provided by Theorem 1.3 for *universal* coverings is, possibly, quite far from the optimal. A discussion of this conjecture can be found at /mathoverflow.net/questions/8534/diameter-of-universal-cover .) In section 6 we will also present a proof the following generalization of Theorem 1.3:

Theorem 1.4 *If the fundamental group of a closed Riemannian manifold M of diameter d is either infinite or finite of order $\geq k$, then for every pair of points $p, q \in M$ and every k there exist at least k geodesics connecting p and q of length $\leq kd$ that represent different path homotopy classes.*

We can combine Theorem 1.3 with the described simple procedure that allows one to reduce Theorem 1.2 for a nonsimply-connected M^n to Theorem 1.2 for its universal covering $\tilde{M}^n$. As the result, we obtain upper bounds for the nonsimply-connected case that are not worse than the estimates in the simply-connected case. As it was already mentioned, the verification mostly involves checking of what happens for small values of k. We are not going to present the details of the straightforward and elementary but tedious calculations here.

The proof of Theorem 1.1 is based on a new curve shortening process. This process will be introduced in the proof of the following theorem at the beginning of section 3. Before stating this theorem recall that a path homotopy between two curves β and γ is a homotopy that preserves the end points. In other words, it is a family of curves $\alpha_\tau(t)$ that continuously depends on $\tau \in [0,1]$ such that $\alpha_0 = \beta,\ \alpha_1 = \gamma$, and such that for every $\tau \in [0,1]$ $\alpha_\tau(0) = \alpha_0(0)$ and $\alpha_\tau(1) = \alpha_0(1)$.

Theorem 1.5 *Let M^n be a closed Riemannian manifold of diameter d, and p, q be two arbitrary points of M^n. Let $\gamma(t)$ be a curve of length L connecting points p and q. Assume that there exists an interval $(l, l+2d]$, such that there are no geodesic loops based at p on M^n of length in this interval that provide a local minimum of the length functional on $\Omega_p M^n$. Then there exists a curve $\tilde{\gamma}(t)$ of length $\leq l+d$ connecting p and q and a path homotopy between γ and $\tilde{\gamma}$ such that the lengths of all curves in this path homotopy do not exceed $L+2d$.*

Observe that this theorem immediately implies Theorem 1.1 for $m=0$. Indeed, S^0 consists of two points, so, if $m=0$, then f is just a set of two loops. In the absence of k short geodesic loops providing local minima for the length functional each of these two loops can be shortened as in Theorem 1.5.

The statement of Theorem 1.1 can be interpreted as a parametric version of Theorem 1.5. Yet note that the curve-shortening process that will be used to prove Theorem 1.5 does not depend on γ continuously, and there is no obvious way to obtain a desired parametric version. The best that one can do is to choose a sufficiently dense finite set of loops L_i in $f(S^m)$ and to shorten them as in Theorem 1.5. Indeed, we will do that in the course of proving Theorem 1.1. But this will leave us with all the other loops in between that still remain long. The further idea can be very vaguely described as follows. In the process of shortening loops L_i we will create continuous 1-dimensional families of paths of controlled length (*"rails"*) that connect p with all points on L_i. The image of $g(S^m)$ in M^n will be the union of the image of $f(S^m)$ in M^n and the constructed "rails". (Recall that each point of $f(S^m)$ is a loop in M^n; here we are talking about the union of all these loops.) The image of $f(S^m)$ in M^n will be cut into very short arcs starting and ending on curves L_i. These short arcs form an m-dimensional family A. Each of these arcs from family A will be included into a loop from the family $g(S^m)$. Every loop from $g(S^m)$ will contain only a controlled number of arcs from A, so their total contribution to the length of the loop is negligibly small. Besides arcs from A each loop from $g(S^m)$ will also contain a controlled number of "rails" and arcs of the curves L_i, so that its total length will be under control.

Now we would like to give a brief review of some existing results related to Theorem 1.2. The first curvature-free upper bounds for the length of a shortest non-trivial geodesic loop on a closed Riemannian manifold in terms of either diameter or the volume of the manifold were proven by S. Sabourau in [S]. However, Sabourau considered the situation when the minimization of the length of the geodesic loop was performed also over all possible choices of the base point of the loop. R. Rotman ([R]) demonstrated that for every point p on every closed n-dimensional manifold of diameter d the length of the shortest geodesic loop based at p does not exceed $2nd$. (It is easy to see that if the base point is prescribed, then there is no upper bound for the length of the shortest geodesic loop in terms of the volume of the manifold, even if the manifold is diffeomorphic to the 2-sphere.) Note also that a shortest non-trivial geodesic loop based at p is the second shortest geodesic starting and ending at p. In [NR1] it was proven that the same estimate $2nd$ holds for the length of the second shortest geodesic between two arbitrary points p and q of an arbitrary n-dimensional Riemannian manifold of diameter d. This is the best known upper bound in the case when $k = 2$ (for every simply-connected manifold M^n).

If $n = 2$, one can also produce better estimates than the estimates provided by Theorem 1.2. In [NR2] we proved that if $n = 2$ and M^n is diffeomorphic to the 2-sphere, then two arbitrary points can be connected by at least k distinct geodesics of length $\leq (4k^2 - 2k - 1)d$. (In the same paper we have also shown that this estimate can be improved to $4k^2 - 6k + 2$ if these points coincide). Making an almost obvious observation that the cycles corresponding to powers of u from the proof of A. Schwarz do not "hang" on local minima of the length functional and, therefore, geodesics corresponding to cup powers of u are different from the geodesics that are local minima of the length functional, we can immediately improve these upper bounds for $k > 2$ to $(k^2+3k+3)d$ in the case of geodesics connecting two distinct points of M^2 and $(k^2+k)d$ in the case of geodesic loops based at any prescribed points of M^2. One of these k geodesic loops can be trivial. Also, one has linear in k bounds $20(k-1)d$ for the lengths of k shortest geodesic loops based at a prescribed point of M^2 ([NR4]), and $(36k-35)d$ for the lengths of k shortest geodesics connecting two points $p, q \in M^2$, where $p \neq q$ ([NR5]). Obviously, these linear upper bounds are better for all sufficiently large values of k. The simple argument used above to reduce Theorem 1.2 in the non-simply conected case to its simply-connected version (for the univeral covering) can be combined with these estimates for S^2 to deduce the same upper bounds for the lengths of k shortest geodesic loops (or, more generally, geodesic arcs connecting a prescribed pair of points) in the case when M^2 is diffeomorphic to RP^2. Note that Theorem 1.4 yields even better upper bounds (namely, kd) when a closed two-dimensional Riemannian manifold M^2 is not diffeomorphic to S^2 or RP^2. Thus, Theorem 1.2 should only be used in the case when $n, k \geq 3$ (and $|\pi_1(M^n)|$ is finite and "small").

In section 7 we will discuss generalizations of Theorems 1.1, 1.5 and 5.3 that involve the notion of the depth of local minima of the length functional. The formal definition of the depth will be given in section 7. Informally, the depth of a non-trivial local minimum γ of the length functional on $\Omega_p M^n$ is the difference between the maximal length of a loop during an "optimal" path homotopy contracting γ and the length of γ.

First, we observe that Theorem 1.5 remains valid if instead of assuming that there are no local minima of the length functional with length in the interval $(l, l+2d]$ we choose any positive S and assume that there are no local minima of depth $\geq S$ with length in this interval. As a corollary, one can strengthen Theorem 1.1 as well as Theorem 5.3 by requiring in the first case that the geodesic loops with lengths in the intervals $(2(i-1)d, 2id]$ are not only local minima of the length functional, but local minima of depth $\geq S$. The "price" is a corresponding increase of the lengths of the loops that must appear in the second case that is proportional to S.

We finish section 7 by observing that for a specific sufficiently large value of S the first case in Theorem 1.1 cannot occur already for $k = 1$, and the generalized form of the second case holds unconditionally.

As the result, we obtain a different proof of a well-known theorem first proven by M. Gromov (see section 1.4 in [Gr0] or ch. 7 in [Gr]) that asserts that for every simply-connected closed Riemannian manifold M^n there exists a constant C such that for every m the inclusion $\Omega_p^{Cm}M^n \subset \Omega_p M^n$ induces surjective homomorphisms of homotopy groups in all dimensions up to m. Our proof yields a good estimate for the constant C that seems to be better than the value that one can extract from the proof by Gromov. The comparison between our results and the results by M. Gromov is done in the last section of the paper.

2 A simple lemma and its multidimensional generalization.

The proof of 1.5 uses the following known lemma. To state this lemma we are going to introduce the following notation that we will be widely using further below in this paper. Let $\tau(t)$ be a path in M^n. We are going to denote the "same" path travelled in the opposite direction as $\bar{\tau}$. If a is a path from x to y, and b is a path from y to z, then we will denote by $a * b$ the join of a and b, that is, the path from x to z that first follows a from x to y, and then b from y to z. Observe that if e_1, e_2 are two paths from p to q, then $e_1 * \bar{e}_2$ is a loop based at p.

Lemma 2.1 *Let e_1, e_2 be two paths starting at q_1 and ending at q_2 on a complete Riemannian manifold M^n. Denote the length of $e_i, i = 1, 2$, by l_i.*

If the loop $\alpha_0 = e_1 * \bar{e}_2$ can be connected to a (possibly trivial) loop $\alpha = \alpha_1$, (see Fig. 1 (a)), by a path homotopy that passes via loops α_τ, $\tau \in [0, 1]$, of length $\leq l_1 + l_2$, then there is a path homotopy $h_\tau(t), \tau \in [0, 1]$, such that $h_0(t) = e_1(t), h_1(t) = \alpha * e_2(t)$ and the length of the paths during this homotopy is bounded above by $l_1 + 2l_2$.

Proof. For the proof see Fig. 1. Note that e_1 is path homotopic to $e_1 * \bar{e}_2 * e_2$ along the curves of length $\leq l_1 + 2l_2$; see Fig. 1 (b,c). (We just insert longer and longer segments of $\bar{e}_2$ travelled twice in the opposite directions.) Now observe that as $e_1 * \bar{e}_2$ is path homotopic to α via the curves α_τ of length $\leq l_1 + l_2$, the path $e_1 * \bar{e}_2 * e_2$ is path homotopic to $\alpha * e_2$ along the curves $\alpha_\tau * e_2$ of length at most $l_1 + 2l_2$; see Fig. 1 (d,e). □

Note that the above lemma has the following higher-dimensional generalization: Let $f : S^m \longrightarrow \Omega^L_{q_1,q_2} M^n, i = 1, 2,\ m \geq 1$, be a continuous map from the m-sphere into a space of (piecewise differentiable) paths on a complete Riemannian manifold M^n between points $q_1, q_2 \in M^n$ of length at most L. Let $L_0 = \min_{s \in S^m}$ length$(f(s))$ and $s_0 \in S^m$ be such that

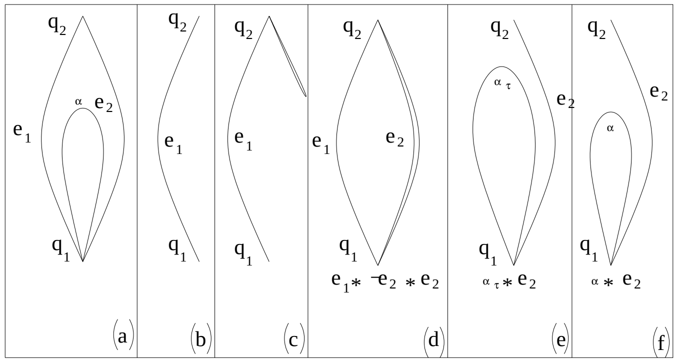

Figure 1: Illustration of the proof of Lemma 2.1.

length$(f(s_0)) = L_0$. One can define a new map $F : S^m \longrightarrow \Omega^{L+L_0}_{q_1} M^n$ by the formula $F(s) = f(s) * \bar{f}(s_0)$. Assume that there exists a homotopy $F_t : S^m \longrightarrow \Omega^{L+L_0}_{q_1} M^n$ contracting F. (Here $t \in [0,1], F_0 = F$, and F_1 is the constant map to the trivial loop based at q_1.) Then we have the following simple lemma:

Lemma 2.2 *There exists a homotopy* $f_t : S^m \longrightarrow \Omega^{L+2L_0}_{q_1 q_2} M^n$, $t \in [0,1]$, *between* $f = f_0$ *and the constant map* f_1 *of* S^m *identically equal to* $f(s_0)$.

Proof. First note that the main point of the lemma is that for every t, f_t takes values in the space of paths of length $\leq L + 2L_0$ connecting q_1 and q_2. The desired homotopy is constructed in two stages. During the first stage we connect f with $f_{\frac{1}{2}}$ defined by the formula $f_{\frac{1}{2}}(s) = f(s) * \bar{f}(s_0) * f(s_0) = F(s) * f(s_0)$. At this stage we join $f(s)$ with longer and longer segments of $\bar{f}(s_0)$ travelled twice with opposite orientations.

During the second stage we contract F using the homotopy F_t, $t \in [0,1]$ leaving intact $f(s_0)$ at the end of each loop $F_t(s) * f(s_0)$. □

In the next section we will present a proof of Theorem 1.5.

3 Curve-shortening process in the absence of geodesic loops.

3.1. Proof of Theorem 1.5.

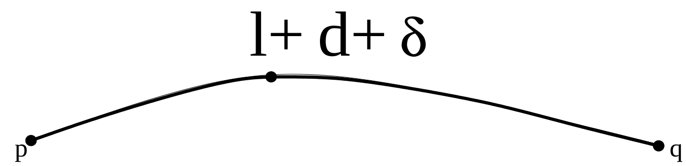

Figure 2: Curve shortening process (i).

Proof of Theorem 1.5.

Assume that there are no geodesic loops based at p with the length in the interval $(l, l+2d]$. Using a standard compactness argument we observe that there exists a positive δ_0 such that there are no geodesic loops based at p with length in the interval $(l, l+2d+\delta_0]$. Obviously, one can choose the value of δ_0 to be arbitrarily small, if desired. Without any loss of generality we can assume that the length L of the curve $\gamma : [0,\ L] \longrightarrow M^n$ parametrized by its arclength is greater than $l+d$. Let $\delta = \delta_0$, if $L \geq l+d+\delta_0$, and $\delta = L-l-d$, if $\delta \in (l+d, l+d+\delta_0)$. Consider the segment $\gamma|_{[0,l+d+\delta]}$ of γ, which we will denote $\gamma_{11}(t)$, and the segment $\gamma|_{[l+d+\delta,L]}$ denoted $\gamma_{12}(t)$ (see fig. 2).

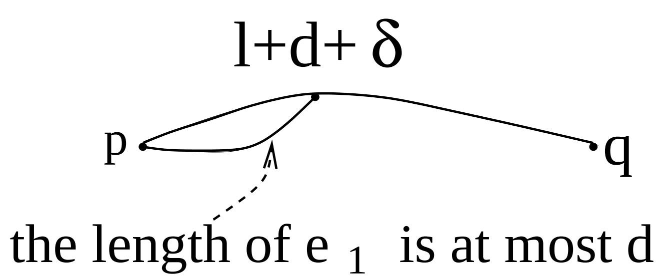

Figure 3: Curve shortening process (ii).

Let us connect points p and $\gamma(l+d+\delta)$ by a minimal geodesic, e_1 (of length $\leq d$), (see fig. 3). Then curves γ_{11} and e_1 form a loop $\gamma_{11} * \bar{e}_1$ of length $\leq l+2d+\delta$ based at p.

Consider a (possibly trivial) shortest loop α_1 that can be connected with $\gamma_{11} * \bar{e}_1$ by a length non-increasing path homotopy. (Its existence follows from the Ascoli-Arzela theorem.) Obviously, α_1 is a geodesic loop based at p that provides a local minimum for the length functional on $\Omega_p M^n$. Therefore our assumptions imply that the length of α_1 is at most l.

By Lemma 2.1, γ_{11} is path homotopic to the curve $\alpha_1 * e_1 = \tilde{\gamma}_{11}$ along

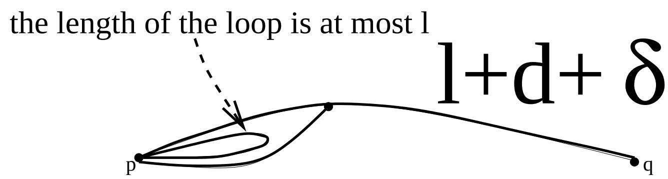

Figure 4: Curve shortening process (iii).

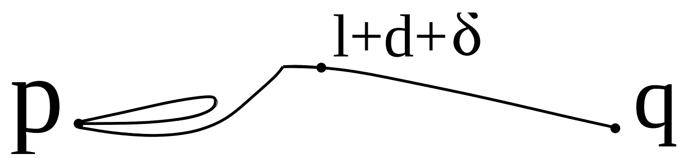

Figure 5: Curve shortening process (iv).

the curves of length at most $l + 3d + \delta$ and, thus, the original curve γ is homotopic to a new curve $\tilde{\gamma}_{11} * \gamma_{12}$ along the curves of length at most $L + 2d$, (see Fig. 5).

Note that the length L_1 of this new curve $\gamma_1 = \tilde{\gamma}_{11} * \gamma_{12}$ is at most $L - \delta$. Assuming that L_1 is still greater than $l + d$, we repeat the process again: We parametrize γ_1 by its arclength. Now, let $\gamma_{21} = \gamma_1|_{[0,l+d+\delta]}$ and $\gamma_{22} = \gamma_1|_{[l+d+\delta,L_1]}$. (Here, as before, if $L_1 < l + d + \delta$, then we use $L_1 - l - d$ as the new value of δ, but otherwise keep the old value of $\delta = \delta_0$.)

Connect the points p and $\gamma_1(l + d + \delta)$ by a minimal geodesic segment e_2, (see Fig. 6). Then γ_{21} and e_2 form a geodesic loop $\gamma_{21} * \bar{e}_2$ based at p of length at most $l + 2d + \delta$.

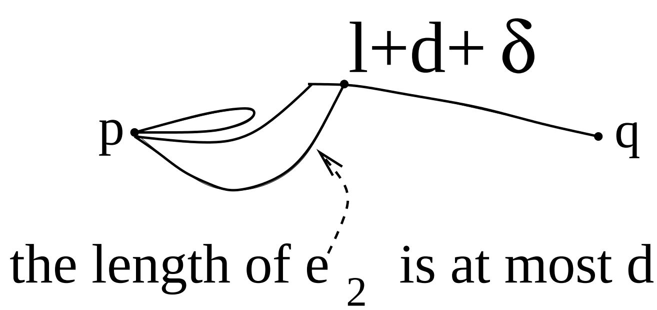

Figure 6: Curve shortening process (v).

Again, we try to connect this loop with a shortest possible loop α_2 by

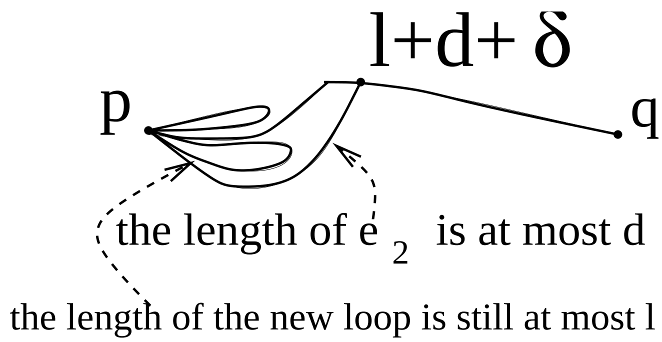

Figure 7: Curve shortening process (vi).

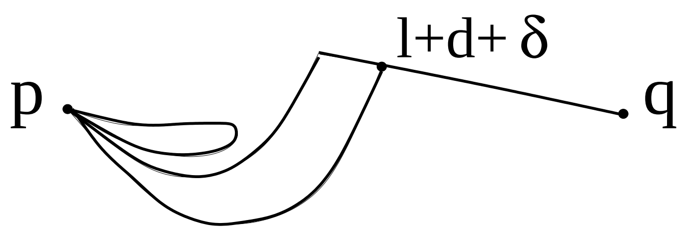

Figure 8: Curve shortening process (vii).

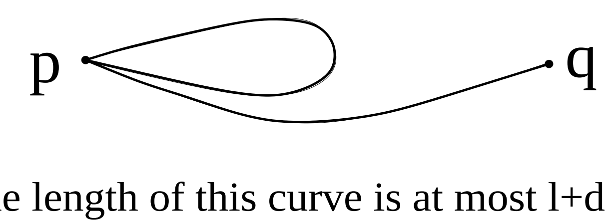

Figure 9: Curve shortening process (viii)

means of a length non-increasing path homotopy. The existence of a minimizer α_2 follows from an easy compactness argument, and α_2 is a geodesic loop providing a local minimum of the length functional on $\Omega_p M^n$, (see Fig. 7). Therefore, the length of α_2 is at most l.

Thus, γ_{21} is path-homotopic to $\tilde{\gamma}_{21} = \alpha_2 * \sigma_2$ along the curves of length at most $l + 3d + \delta$. It follows that γ_1 is homotopic to $\gamma_2 = \tilde{\gamma}_{21} * \gamma_{22}$ along the curves of length at most $L + 2d$, (see Fig. 8).

We will continue this process in the same manner. It is easy to see that this process will terminate in a finite number of steps with a curve of length $\leq l + d$, and that the number of steps will not exceed $\lfloor \frac{L-l-d}{\delta_0} \rfloor + 1$. □

3.2. Proof of Theorem 1.5: an additional remark. Note that we have proven a stronger statement. We have shown, assuming the hypothesis of the theorem above, that for each path $\gamma(t)$ connecting p and q there exists an increasing but not necessary strongly increasing function $\tau(s)$ and a 1-parameter family of curves C_s^γ of length $\leq l+3d+\delta$ continuously depending on a parameter s such that:

A. For every s, C_s^γ connects p with $\gamma(\tau(s))$. In other words, $C_s^\gamma(0) = p$, $C_s^\gamma(1) = \gamma(\tau(s))$.

B. There exist two partitions: $P^\gamma = \{0 = t_0^\gamma < t_1^\gamma < t_2^\gamma < ... < t_{k^\gamma}^\gamma = 1\}$ and $Q^\gamma = \{0 = s_0^\gamma < s_1^\gamma < ... < s_{2k^\gamma}^\gamma = 1\}$, such that

(1) $C_s^\gamma(1) = \gamma(t_i^\gamma)$ for $s \in [s_{2i-1}^\gamma, s_{2i}^\gamma]$. In particular, the endpoint of c_s^γ remains constant for $s \in [s_{2i-1}^\gamma,\ s_{2i}^\gamma]$.

(2) For every $s \in [s_{2i}^\gamma,\ s_{2i+1}^\gamma]$ $C_s^\gamma(1) = \gamma(t)$ for some $t \in [t_i^\gamma, t_{i+1}^\gamma]$. Moreover, τ strictly increases on $[s_{2i},\ s_{2i+1}]$, and $\tau([s_{2i},\ s_{2i+1}]) = [t_i,\ t_{i+1}]$.

(3) For all i the length of $C_{s_{2i}^\gamma}^\gamma$ does not exceed $l + d$.

(4) The curve $C_{s_{2k^\gamma}^\gamma}^\gamma = C_1^\gamma$ is the final result of the application of the curve-shortening process described in the proof of Theorem 1.5 to γ.

The curves C_s^γ are depicted on Fig. 6-9 as the curves connecting p with a variable point that moves from p to q along γ. Note also that the partition P^γ can be chosen as fine as desired.

Further, notice that the constructed path homotopy H_γ between our original path, γ, and C_1^γ can be described as follows: At each moment of time t $H_\gamma(t)$ is the path that first goes along C_t^γ, and then runs along γ from $C_t^\gamma(1)$ to $\gamma(1)$. Below we will be calling curves $H_\gamma(t)$ *partial shortenings* of γ.

Our final observation is that when we apply the shortening procedure from the proof of Theorem 1.5 to an initial segment of γ, $\gamma_{[0,\lambda]}$ for some

$\lambda \in [0,1]$ we are going to obtain a subfamily of the 1-parametric family of curves C^γ (up to the first of these curves that connects p with $\gamma(\lambda)$).

3.3. Case of $m = 1$ in Theorem 1.1. Beginning of the proof. Next we will prove the following theorem, which together with Theorem 1.5 immediately implies Theorem 1.1 in the case of $m = 1$.

Theorem 3.1 *Let M^n be a closed Riemannian manifold of diameter d, p a point of M^n, and k a positive integer number. Assume that there exists a positive integer $j \leq k$ such that the length of every geodesic loop that provides a local minimum for the length functional on $\Omega_p M^n$ is not in the interval $(2(j-1)d,\ 2jd]$. Consider a continuous map $f : [0,1] \longrightarrow \Omega_p M^n$ such that the lengths of both $f(0)$ and $f(1)$ do not exceed $2(j-1)d$. Then f is path homotopic to $\tilde{f} : [0,\ 1] \longrightarrow \Omega_p^{(6j-1)d+o(1)} M^n \subset \Omega_p^{(6k-1)d+o(1)} M^n$. Moreover, assume that for some L the image of f is contained in $\Omega_p^L M^n$. Then one can choose a path homotopy between f and $\tilde{f}$ so that its image is contained in $\Omega_p^{L+(4j+2)d+o(1)} M^n$.*

Proof. Without any loss of generality we can assume that f is Lipschitz. (If not, we can make f Lipschitz by performing an arbitrarily small deformation.) Choose a partition of $t_0 = 0 < t_1 < t_2 < ... < t_N = 1$ of the interval $[0,1]$, so that $\max_i \max_{\tau \in [0,1]} \text{length}(f(t)|_{t \in [t_{i-1}, t_i]}(\tau)) \leq \varepsilon$ for a small ε that will eventually approach 0, (see Fig. 10 which depicts a situation, when $f(0)$ and $f(1)$ are both constant loops, but the general case is completely analogous. Note that $f(t)|_{t \in [t_i,\ t_{i+1}]}(\tau)$ are the short "vertical" curves connecting "horizontal" curves $f(t_i)$.)

Let us denote loops $f(t_i)$ by $\gamma_i(r)$.We can use Theorem 1.5 to replace all $\gamma_i(r)$ of length greater than $(2j-1)d$ by the loops β_i of length $\leq (2j-1)d$. The loops that were originally shorter than $(2j-1)d$ will remain as they were. Note that in this last case we will also have a family of indexed paths C of controlled length connecting p with all points on the loop as at the end of the proof of Theorem 1.2: namely, the initial segments of the loop.

We will now construct path homotopies between each pair β_{i-1}, β_i. These homotopies will pass through loops of length $\leq (6j-1)d + o(1)$. (Here and below $o(1)$ denotes terms that are bounded by a linear function of the parameters of our shortening process, namely, δ_0 (see the proof of Theorem 1.5) and ε that can be made arbitrarily small.)

Consider such a pair of consecutive curves $\sigma = \beta_{i-1}$ and $\alpha = \beta_i$. Recall that these curves were obtained from γ_{i-1} and γ_i, respectively, as the final

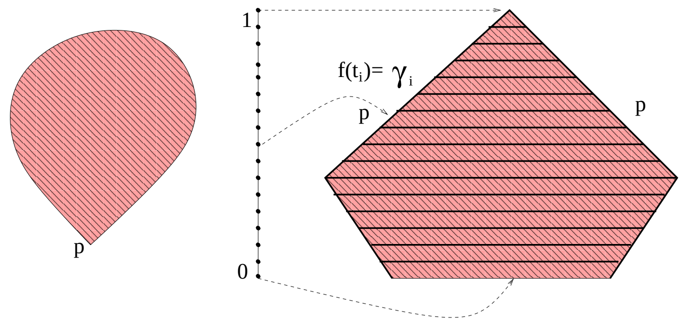

Figure 10: Partition of the map $f : S^1 \longrightarrow M^n$ into "small" intervals

result of the application of the curve shortening process. These two applications of the curve shortening process also result in two 1-parameter families of curves C_s^{σ}, C_s^{α} that have properties described after the proof of Theorem 1.5 (for $l = (2j-2)d$).

Recall that these curves connect p with points on γ_{i-1} and γ_i, that $C_1^{\gamma_{i-1}}$ coincides with σ, and that $C_1^{\gamma_i}$ coincides with α.

We will construct a path between σ and α in two steps. In the first step we will consider a loop that is a join of σ and $\bar{\alpha}$, namely, $\sigma * \bar{\alpha}$. We will construct a homotopy that contracts this loop to a point via loops of length $\leq 4jd + o(1)$ (when $\delta_0 + \varepsilon \longrightarrow 0$). The second step will be merely an application of Lemma 2.1. (On the second step the summand of $2(j-1)d + d = (2j-1)d$ will be added to the previous upper bound $4jd + o(1)$ for the length of loops during the contracting homotopy.)

So, we need only to describe the first step to finish our construction, namely, a homotopy that contracts $\sigma * \bar{\alpha}$ via loops of controlled lengths. Note that γ_{i-1} and γ_i are very close to each other, and are connected by the *continuous* family of very short curves $f(t)|_{t\in[t_{i-1},t_i]}(\tau)$, $\tau \in [0,1]$. We will write the desired homotopy as a homotopy through loops of the form $C_{s_1}^{\gamma_{i-1}} * f(t)|_{t\in[t_{i-1},t_i]}(\tau) * \bar{C}_{s_2}^{\gamma_i}$. Note that in order for this curve to be defined we must have $C_{s_1}^{\gamma_{i-1}}(1) = f(t_{i-1})(\tau)$ and $C_{s_2}^{\gamma_i}(1) = f(t_i)(\tau)$, or, equivalently, $\tau_1(s_1) = \tau_2(s_2) = \tau$. Here $\tau_1(s)$ and $\tau_2(s)$ denote the increasing functions from $[0,1]$ to $[0,1]$ for the curves γ_{i-1} and γ_i. For every z, the inverse image $\tau_j^{-1}(z)$ is either a point or a closed interval, and it is only an interval when

z is a boundary point for one of the partitions $P^{\gamma_{i-1}}$ or P^{γ_i}. Without loss of generality, we may assume that these partitions are disjoint, so that the loops of the form $C^{\gamma_{i-1}}_{s_1} * f(t)|_{t\in[t_{i-1},t_i]}(\tau) * \bar{C}^{\gamma_i}_{s_2}$ form a 1-parametric family. This family is the desired homotopy.

Furthermore, if $\tau_1(s_1)$ is not a boundary point of $P^{\gamma_{i-1}}$, then length of $C^{\gamma_{i-1}}_{s_1} \leq 2(j-1)d + d + \delta_0$. Since $P^{\gamma_{i-1}}$ and P^{γ_i} are disjoint, if $\tau_1(s_1) = \tau = \tau_2(s_2)$, then the length of $C^{\gamma_{i-1}}_{s_1}$ does not exceed $(2j-1)d + \delta_0$ and/or length of $C^{\gamma_i}_{s_2} \leq (2j-1)d + \delta_0$. Thus each curve in the homotopy consists of a curve of length at most $l + 3d + \delta = (2j+1)d + \delta_0$, an ε-short "vertical" curve, and a curve of length at most $(2j-1)d + \delta_0$.

3.3.A. Here is a more concrete description of the resulting one-parametric family of loops that also takes into account some details of the construction of C^{γ}_s in the proof of Theorem 1.5 above.

Let $\varepsilon_\tau = f(t)(\tau)$, where τ is fixed and t varies in the interval $[t_{i-1}, t_i]$. Recall that we can ensure that the length of ε_τ is arbitrarily small for all $\tau \in [0,1]$ by choosing $t_i - t_{i-1}$ to be sufficiently small.

Let us begin with the loop $\sigma * \bar{\alpha} = C^{\sigma}_1 * \bar{C}^{\alpha}_1$ that is based at the point p. Corresponding to C^{σ}_s and C^{α}_s consider two pairs of partitions: $\{P^{\sigma}, Q^{\sigma}\}$ and $\{P^{\alpha}, Q^{\alpha}\}$. Let $P^{\sigma} = \{0 = r^{\sigma}_0 < r^{\sigma}_1 < ... < r^{\sigma}_{k_\sigma - 1} < r^{\sigma}_{k_\sigma} = 1\}$ and $P^{\alpha} = \{0 = r^{\alpha}_0 < r^{\alpha}_1 < ... < r^{\alpha}_{k_\alpha - 1} < r^{\alpha}_{k_\alpha}\}$. Also let $P = P^{\sigma} \cup P^{\alpha}$. Without any loss of generality, we can assume that $P = \{0 = r^{\sigma}_0 = r^{\alpha}_0 < r^{\alpha}_1 < r^{\sigma}_1 < r^{\alpha}_2 < r^{\sigma}_2 < ... < r^{\alpha}_{k_\alpha - 1} < r^{\sigma}_{k_\sigma - 1} < r^{\alpha}_{k_\alpha} = r^{\sigma}_{k_\sigma} = 1\}$.

We will now present a homotopy that contracts $\sigma * \bar{\alpha}$ to p as a loop over short loops.
(a) $C^{\sigma}_1 * \bar{C}^{\alpha}_1$ is homotopic to $C^{\sigma}_{s^{\sigma}_{2k_\sigma - 1}} * C^{\alpha}_1$ over loops of length at most $4jd$; see Fig. 11.
(b) $C^{\sigma}_{s^{\sigma}_{2k_\sigma - 1}} * C^{\alpha}_1$ is homotopic $C^{\sigma}_{s^{\sigma}_{2k_\sigma - 1}} * C^{\alpha}_{s^{\alpha}_{2k_\alpha - 1}}$ over the loops of length $4jd$, (see Fig. 12).
(c)$C^{\sigma}_{s^{\sigma}_{2k_\sigma - 1}} * C^{\alpha}_{s^{\alpha}_{2k_\alpha - 1}}$ is homotopic to $C^{\sigma}_{s^{\sigma}_{2k_\sigma - 2}} * \bar{\varepsilon}_{r^{\sigma}_{k-1}} * \bar{C}^{\alpha}_{s^{\alpha}}$ for $s^{\alpha} \in [s^{\alpha}_{2k-2}, s^{\alpha}_{2k-1}]$ over the curves of length at most $(4j-2)d + 2\delta + \varepsilon$, (see Fig. 13).
(d) $C^{\sigma}_{s^{\sigma}_{2k_\sigma - 2}} * \bar{\varepsilon}_{r^{\sigma}_{k-1}} * \bar{C}^{\alpha}_{s^{\alpha}}$ is homotopic to $C^{\sigma}_{s^{\sigma}_{2k_\sigma - 3}} * \bar{\varepsilon}_{r^{\sigma}_{k-1}} * \bar{C}^{\alpha}_{s^{\alpha}}$ over the curves of length $4jd + 2\delta + \varepsilon$, (see Fig. 14).
(e) $C^{\sigma}_{s^{\sigma}_{2k_\sigma - 3}} * \bar{\varepsilon}_{r^{\sigma}_{k-1}} * \bar{C}^{\alpha}_{s^{\alpha}}$ is homotopic to $C^{\sigma}_{s^{\sigma}} * \bar{\varepsilon}_{r^{\alpha}_{k-1}} * \bar{C}^{\alpha}_{s^{\alpha}_{2k_\alpha - 2}}$ for $s^{\sigma} \in [s^{\sigma}_{2k_\sigma - 3}, s^{\sigma}_{2k_\sigma - 4}]$ over the curves of length at most $(4j-2)d + 2\delta + \varepsilon$, (see Fig. 15).

Proceeding in the above manner we will contract the loop to p over curves of length at most $4jd + o(1)$.

3.4. Synchronization. As before, let $\gamma_i = f(t_i)$ and $\gamma_{i+1} = f(t_{i+1})$

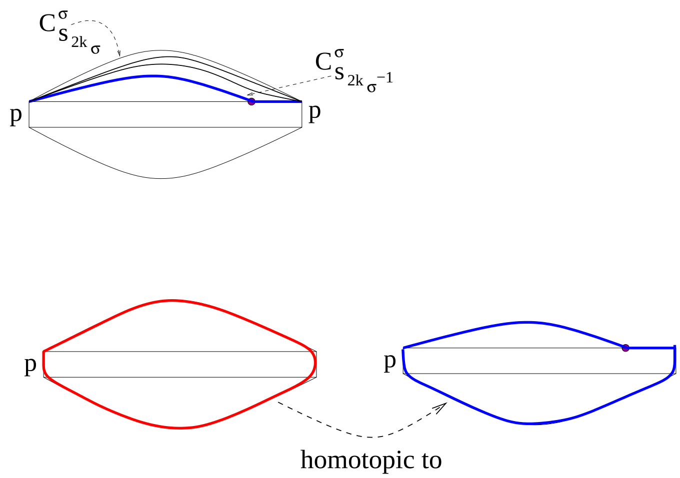

Figure 11: Contracting $\sigma * \bar{\alpha}$ as a loop (i).

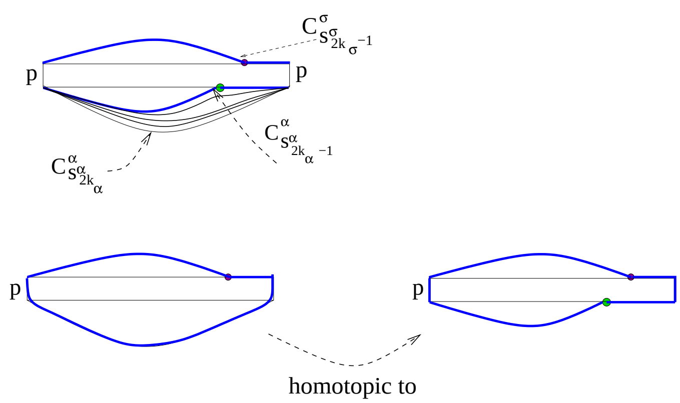

Figure 12: Contracting $\sigma * \bar{\alpha}$ as a loop (ii).

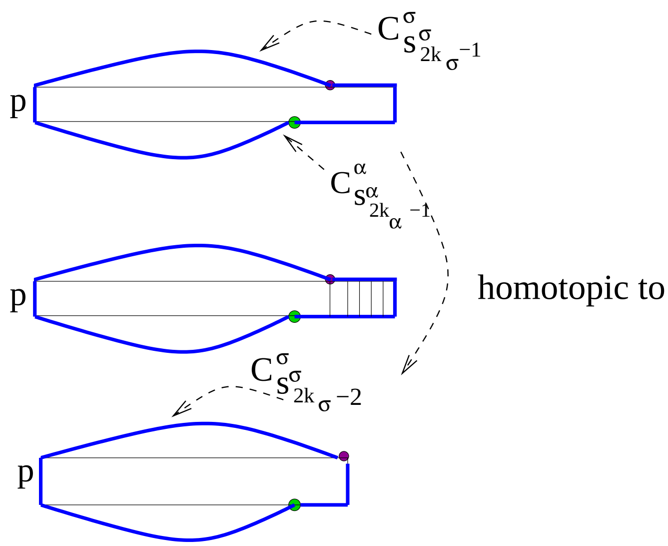

Figure 13: Contracting $\sigma * \bar{\alpha}$ as a loop (iii).

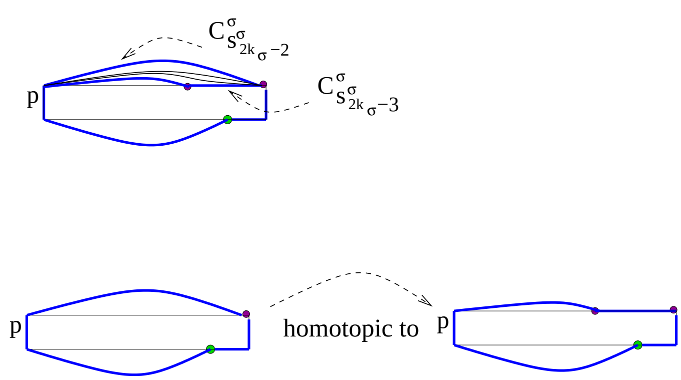

Figure 14: Contracting $\sigma * \bar{\alpha}$ as a loop (iv).

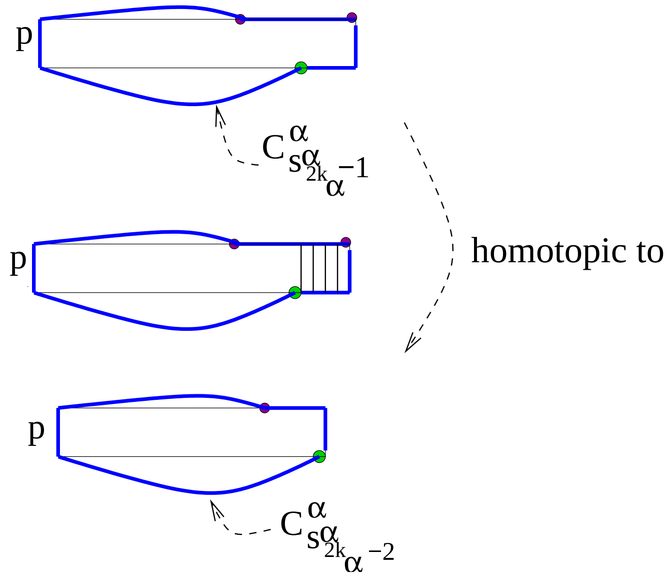

Figure 15: Contracting $\sigma * \bar{\alpha}$ as a loop (v).

denote two close loops. In our proof we used and will be using loops formed by a curve from the family C^{γ_i} that goes to $f(t_i)(\tau)$ for some τ, a short "vertical" segment in the image of f connecting $f(t_i)(\tau)$ with $f(t_{i+1})(\tau)$ and a curve from the family $C^{\gamma_{i+1}}$ that goes to $f(t_{i+1})(\tau)$. It would be convenient for us to change the dependence of *all* curves $C^{\gamma_i}_s$ on s (for all values of i) to ensure that we could write these loops as $C^{\gamma_i}_r * f(t)|_{t\in[t_i,t_{i+1}]}(s(r)) * \bar{C}^{\gamma_{i+1}}_r$ (for a new indexing of the family of curves C^{γ_i} by r and some increasing (but not necessarily strictly increasing) function $s(r)$). Equivalently, this just means that for every i the endpoint of $C^{\gamma_i}_r$ is $\gamma_i(s(r))$. More formally, a synchronization is a collection of strictly increasing surjective functions $\psi_i : [0,1] \longrightarrow [0,1]$ such that there exists an increasing (but not necessarily strictly increasing) surjective function $s : [0,1] \longrightarrow [0,1]$ with the following property: For each i and $r \in [0,1]$ $\gamma_i(s(r)) = C^{\gamma_i}_{\psi_i(r)}(1)$. Of course, finding such functions would be very easy if for every i and t there was exactly one value of r such that $C^{\gamma_i}_r(1) = \gamma_i(t)$. In this case we could assign to the curve from the family C^{γ_i} that ends at $\gamma_i(r)$ the index r. (In other words, we would define $\psi_i(t) = r$ for all i, t, and then just take $s(r) = r$.) However, this situation is not typical. In the general case we can assume without any loss of generality that for each value of r there exists at most one value of i such that there exists an interval $I(r)$ of non-zero length such that curves $C^{\gamma_i}(z)$ end at the same point $\gamma_i(r)$ for all values of $z \in I(r)$. (One can ensure this property by introducing arbitrarily small changes in the curve-shortening process described in sections 3.1, 3.2. Namely, one needs to introduce arbitrarily small generic perturbations in the choice of points $t_i^{\gamma_j}$ described in section 3.2, condition B (1).) Further, there are only finitely many values of r such that there exists such an interval $I(r)$ (for at least one value of i). Denote this value of i (if it exists) by $i(r)$. We are going to cut $[0,1]$ at all such values of r and insert there closed intervals of length equal to the length of $I(r)$. In our new parametrization all curves γ_i will be constant on all newly inserted intervals. (In other words, the function $s(r)$ will be constant on each of these intervals - or, more precisely, on the image of each of these intervals after a reparametrization of the considered long interval back to $[0,1]$ that will be described below.) Also, the curve of the family C^{γ_i} connecting p with $\gamma_i(r)$ will be initially indexed by r providing that r is not one of the points where we have cut $[0,1]$ (and, thus, there is only one curve from the family C^{γ_i} that ends at $\gamma_i(r)$). If r is one of the points where we have cut $[0,1]$, then there exists a whole interval I of values of z (in the old parametrization) such that $C_z^{\gamma_i}(1)=\gamma_i(r)$. This happens only if $i=i(r)$. In this case there is an obvious isometry between $I=I(r)$ and the interval that we have inserted at r, and we will initially replace the old indices $z \in I$ with the matching numbers in the newly inserted interval. (More formally, we have just described how to define the functions ψ_i, which are, however, now defined on a very long interval instead of $[0,1]$.) Now the only remaining problem is that r runs over some very long interval. This problem can be resolved by the rescaling the domain of r linearly back to $[0,1]$. We are going to call the resulting reindexing of all families $C_r^{\gamma_i}$ a *synchronization*. The function $s(r)$ will be called a *synchronization function*.

3.5. End of the proof of Theorem 1.1 for $m=1$. It remains to prove that the constructed path $\tilde{f}$ in Ω_pM^n is path homotopic to f. Here is the construction of a path homotopy G between $\tilde{f}$ and f: $G(1)=\tilde{f},\ G(0)=f$. Now define $G(\lambda)$ for each $\lambda \in (0,1)$. The basic idea is that at the moment of time λ we do not shorten $f(t_i)$ (for all i) all the way using the construction in the proof of Theorem 1.5, as we did above. Instead we use the partial shortening $H_i(\lambda) := H_{f(t_i)}(\lambda)$ defined in section 3.2. Recall that for every $\lambda \in [0,1]$ the path $H_i(\lambda)$ consists of two arcs. The first arc is the path $C_\lambda^{\gamma_i}$. The second arc is the arc of γ_i that starts at $C_\lambda^{\gamma_i}(1)$ and ends at $\gamma_i(1)$. To construct the desired path homotopy G between f and $\tilde{f}$ in Ω_pM^n at the moment λ we replace all long curves γ_i not by

$\beta_i = H_i(1)$ but by $H_i(\lambda)$. In other words, $G(\lambda)(t_i) = H_i(\lambda)$. Assuming that we have already synchronized the parametrizations of different one-parametric families C (as in section 3.4), and $s(r)$ denotes the synchronization function, $C_r^{\gamma_i}(1)$ and $C_r^{\gamma_{i-1}}(1)$ can be connected by a (very short) arc $f(t)|_{t\in[t_{i-1},t_i]}(s(r))$. Now we can form loops $C_r^{\gamma_i} * f(t)|_{t\in[t_{i-1},t_i]}(s(r)) * \bar{C_r}^{\gamma_{i-1}}$ for all $r \leq \lambda$, as before. These loops provide a contracting homotopy for the loop $C_\lambda^{\gamma_i} * f(t)|_{t\in[t_{i-1},t_i]})(s(\lambda)) * \bar{C_\lambda}^{\gamma_{i-1}}$. Let τ denote $s(\lambda)$, f_1^r denote the arc $f(t)|_{t\in[t_{i-1},(t_{i-1}+t_i)/2]}(s(r))$, f_2^r denote $f(t)|_{t\in[(t_{i-1}+t_i)/2,t_i]}(s(r))$. Using the proof of Lemma 2.1 we can transform this homotopy into a path homotopy between $C_\lambda^{\gamma_{i-1}} * f_1^\lambda$ and $C_\lambda^{\gamma_i} * \bar{f}_2^\lambda$. (The first of these paths plays the role of e_1 in the terminology of Lemma 2.1, and the second plays the role of e_2.) Joining all paths in this path homotopy with the arc $f(\frac{t_{i-1}+t_i}{2})(r)|_{r\in[\tau,1]}$ that connects m_i^λ and p we are going to get a 1-parametric family of paths that starts at $C_\lambda^{\gamma_{i-1}} * f_1^\lambda * f(\frac{t_{i-1}+t_i}{2})(r)|_{r\in[\tau,1]}$ and ends at $C_\lambda^{\gamma_i} * \bar{f}_2^\lambda * f(\frac{t_{i-1}+t_i}{2})(r)|_{r\in[\tau,1]}$. We will define $G(\lambda)$ on a segment $[\frac{t_{i-1}+t_i}{2} - \lambda\frac{t_i-t_{i-1}}{2}, \frac{t_{i-1}+t_i}{2} + \lambda\frac{t_i-t_{i-1}}{2}]$ of the segment $[t_{i-1}, t_i]$using this 1-parametric family of paths. (These paths will be the values of $G(\lambda)(t)$ for t in the considered interval.) In particular, $G(\lambda)(\frac{t_{i-1}+t_i}{2} - \lambda\frac{t_i-t_{i-1}}{2}) = C_\lambda^{\gamma_{i-1}} * f_1^\lambda * f(\frac{t_{i-1}+t_i}{2})(r)|_{r\in[\tau,1]}$ and $G(\lambda)(\frac{t_{i-1}+t_i}{2} + \lambda\frac{t_i-t_{i-1}}{2}) = C_\lambda^{\gamma_i} * \bar{f}_2^\lambda * f(\frac{t_{i-1}+t_i}{2})(r)|_{r\in[\tau,1]}$.

For reasons that will be explained below we are going to parametrize $G(\lambda)$ as follows. The arcs in the 1-parametric family of paths that starts at $C_\lambda^{\gamma_{i-1}} * f_1^\lambda$ and ends at $C_\lambda^{\gamma_i} * \bar{f}_2^\lambda$ will be parametrized by $[0, \lambda + \lambda(1-\lambda)]$, and the arcs $f(\frac{t_{i-1}+t_i}{2})(r)|_{r\in[\tau,1]}$ will be parametrized by $[\lambda + \lambda(1-\lambda), 1]$. (The choice of the function $\lambda(1-\lambda)$ here is motivated by that fact that $\lambda(1-\lambda) \longrightarrow 0$ as either $\lambda \longrightarrow 0$, or $\lambda \longrightarrow 1$. Also, if $\lambda \in (0,1)$, then $\lambda + \lambda(1-\lambda) = \lambda(2-\lambda) < 1$. Each time when we need to reparametrize a path p defined on $[0,1]$ by an interval $[a,b]$, we replace $p(t)$ by $p_{new}(t) = p(\frac{t-a}{b-a})$. It is not difficult to verify that the considered reparametrizations do not lead to discontinuities as λ approaches one of the endpoints of $[0,1]$.)

It remains to define $G(\lambda)$ on two intervals $[t_{i-1}, \frac{t_{i-1}+t_i}{2} - \lambda\frac{t_i-t_{i-1}}{2}]$ and $[\frac{t_{i-1}+t_i}{2} + \lambda\frac{t_i-t_{i-1}}{2}, t_i]$ that can be regarded as a neighborhood of the boundary of the interval $[t_{i-1}, t_i]$. The "thickness" of this neighborhood is variable, and tends to 0, as $\lambda \longrightarrow 1$. When $t \in [t_{i-1}, \frac{t_{i-1}+t_i}{2} - \lambda\frac{t_i-t_{i-1}}{2}]$, all paths start from $C_\lambda^{\gamma_{i-1}}$. Then they follow a variable path that consists of the arc $f(r)|_{r\in[t_{i-1}, t_{i-1}+\frac{1}{1-\lambda}(t-t_{i-1}))]}(s(\lambda))$ followed by $f(t_{i-1} + \frac{1}{1-\lambda}(t - t_{i-1}))(s(r))|_{r\in[\lambda,1])}$. The parametrization of the join of these three paths by $[0,1]$ is the following: We use $[0,\lambda]$ to parametrize $C_\lambda^{\gamma_{i-1}}$, $[\lambda, \lambda + \lambda(1-\lambda)]$ to parametrize the second arc, and $[\lambda + \lambda(1-\lambda), 1]$ to parametrize the third arc. It is not difficult to see that no discontinuity arises, as $\lambda \longrightarrow 1$ (despite the term $\frac{1}{1-\lambda}$ in the definition of $G(\lambda)$). Of course, it is crucial here that $f(t)(1)$ does not depend on t (and is equal to p in our situation, and to a possibly diffent point q in a more general situation considered below). This property of f ensures that all paths in the image of the restriction of $G(\lambda)$ to the considered interval become close to $C_\lambda^{\gamma_{i-1}}$ as $\lambda \longrightarrow 1$.

One can define the restriction of $G(\lambda)$ on $[\frac{t_{i-1}+t_i}{2} + \lambda\frac{t_i - t_{i-1}}{2}, t_i]$ in a completely analogous way. In particular, all paths start as $C_\lambda^{\gamma_i}$ followed first by a longer and longer arc of f_2^λ, and then by the image under $f(t)$ of the segment $[s(\lambda), 1]$, where a fixed for each vertical segment value of t varies between $\frac{t_{i-1}+t_i}{2}$ and t_i.

To find an upper bound for lengths of loops in the image of G we can just add L, $o(1)$ and the maximal length of the curves in the path homotopy between $C_\lambda^{\gamma_{i-1}} * f_1^\lambda$ and $C_\lambda^{\gamma_i} * \bar{f}_2^\lambda$. This last length can be naturally majorized by $(6j+1)d + o(1)$, see the proof of Lemma 2.1. But we can do somewhat better, if we note that $C_\lambda^{\gamma_{i-1}}$ replaces and shortens a segment of γ_{i-1} whose length was counted as a part of the term L. Therefore, we can subtract $(2j-1)d$ from our upper bound obtaining the upper bound $L + (4j+2)d + o(1)$.

We would like to provide the following less formal explanation (or reinterpretation) of the construction of the path homotopy between $H_{i-1}(\lambda)$ and $H_i(\lambda)$. These two paths consist of curves $C_\lambda^{\gamma_{i-1}}$ and $C_\lambda^{\gamma_i}$ joined with nearly identical "tails" that are the arcs of $\gamma_{i-1}(\tau)$ and $\gamma_i(\tau)$ between $\tau = s(\lambda)$ and $\tau = 1$. The "tails" form a part of one parametric family of "tails" $f(\varrho, s), s \in [\tau, 1]$, where the parameter ϱ ranges in $[t_{i-1}, t_i]$. The idea was to "fill" the "digon" formed $C_\lambda^{\gamma_{i-1}}$ and $C_\lambda^{\gamma_i}$ in exactly the same way as we filled the digon formed by $\beta_{i-1} = C_1^{\gamma_{i-1}}$ and $\beta_i = C_1^{\gamma_i}$ by a one-parametric family of paths of controlled length, and to attach to each of these paths the corresponding "tail" $f(\varrho, s), s \in [\tau, 1]$, for an appropriate value of ϱ. Of course, this idea required some minor corrections as $C_\lambda^{\gamma_{i-1}}$ and $C_\lambda^{\gamma_i}$ end at very close but still different points, and were made to form a digon only after we attached to them (very short) arcs inside $f(t)|_{t\in[t_{i-1},t_i]}(\tau)$. After the endpoints of these curves were made identical, we were able to add to all these curves identical "tails". As we will see below, this idea directly generalizes to the situation of maps of higher dimensional spheres to $\Omega_p M^n$ (or to $\Omega_{pq} M^n$). $\square$

4 Small spheres in the loop space.

In this section we will demonstrate that in the absence of a great number of short geodesic loops, every homotopy class of $\Omega_p M^n$ can be represented by a sphere that passes through short loops (thus proving Theorem 1.1). We assume that there exists k such that there are no geodesic loops based at p on M^n with the length in the interval $(2(k-1)d, 2kd]$ providing a local minimum for the length functional. We are going to prove that f is homotopic to a map $\tilde{f}$ with the image in $\Omega_p^{\tilde{L}+o(1)} M^n$, where $\tilde{L} = ((4k+2)m + (2k-3))d$, and the homotopy can be chosen so that it has its image in $\Omega_p^{\bar{L}+o(1)}$, where $\bar{L} = L + \tilde{L} - (2k-1)d$.

The proof of the theorem is done using a recursion with respect to m. The case $m = 1$ of this theorem had been proven in the previous section. Before presenting the proof of the general case, we will explain the proof of the theorem in the case $m = 2$. (We will explain the proof in the case when f is a map of S^2. It will be clear from our explanations that the case when f is a map of the pair $(D^2, \partial D^2)$ can be treated exactly the same.) Note, that the general case is completely analogous to the case $m = 2$. Yet we decided to include a detailed explanation of the case $m = 2$ in this paper, as one can better visualize and explain the geometric ideas of the proof.

4.1. From $m = 1$ to $m = 2$. A general plan. Let $I = [0, 1]$. Consider a map $f : I \times I \longrightarrow \Omega_p M^n$, where $\partial(I \times I)$ is mapped to p. Without any loss of generality we can assume that all paths in the image of f are parametrized proportionally to their arclengths. Let us subdivide $I \times I$ into squares of a very small size. (The maximal length of the image of one of these sides under f will contribute towards $o(1)$ term in the conditions of the theorem.) Denote these squares by R_{ij}, and their vertices by $(x_{i-1}, y_{j-1}), (x_{i-1}, y_j), (x_i, y_{j-1}), (x_i, y_j)$. We can consider the subdivision of $I \times I$ into squares R_{ij} as a cell subdivision. The vertices (x_i, y_j) will be 0-cells, edges of squares R_{ij} will be 1-cells, and their interiors will be 2-cells. Note that f maps each of 0-cells to a loop in $\Omega_p M^n$. Each of those loops that is too long, (i.e. of length greater than $(2k-1)d$) will be replaced by a shorter one as in Theorem 1.5 via a path homotopy described in the proof of Theorem 1.5. This yields the desired map $\tilde{f}$ on the 0-skeleton of the cell subdivision.

Now we perform the synchronization (as in section 3.4) of all families $C_\lambda^{f(v)}$, where v runs over the set of all vertices of all squares R_{ij}. Recall that the purpose of the synchronization is to find (simultaneous) parametrizations

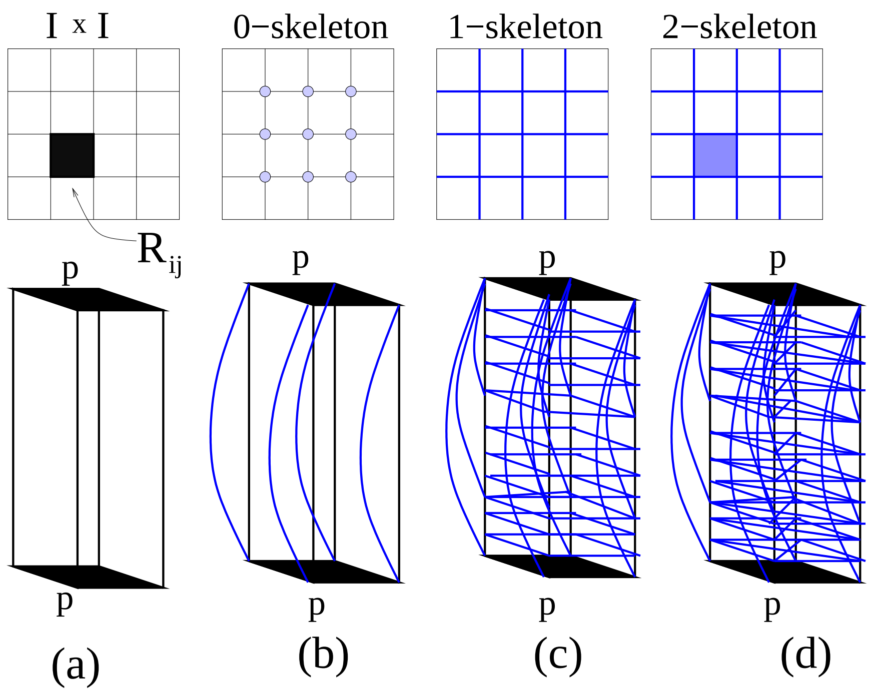

Figure 16: Replacing the map

of all these families by a parameter $r \in [0,1]$ so that for each r and v the endpoint $C_r^{f(v)}$ coincides with $f(v)(s(r))$, where $s(r) \in [0,1]$ does not depend on v.

Every time we apply Lemma 2.1 (directly or as a part of our proof of Theorem 3.1) to a pair of arcs starting at p and ending at a common point naturally associated with some two vertices of R_{ij}, we need to consider one of these arcs as e_1 and the other as e_2 (in terminology of Lemma 2.1), and there is is an asymmetry involved in this choice. We would like to ensure that the choice is done in a canonical way. For this purpose we enumerate all 0-cells by numbers $1, 2, \ldots$ and agree always to choose the path associated with the vertex with a smaller number as e_1, and the path associated with the vertex with a larger number as e_2.

Afterwards, we will replace the restrictions of f to the edges that connect these vertices as in the proof of Theorem 3.1 in the previous section (see Fig. 16 (a)-(c)). We obtain a new map $\tilde{f}$ from the 1-skeleton of the cell subdivision to $\Omega_p M^n$. All loops in the image of $\tilde{f}$ have a controlled length (more precisely, the length will not exceed $(6k-1)d + o(1)$). Now our goal is to "fill" the interior of every square R_{ij}, (Fig. 16 (d)), that is, to extend $\tilde{f}$ from ∂R_{ij} to R_{ij} for all i, j. (After this is done, we will only need to verify that $\tilde{f}$ and f are homotopic and that this homotopy can be chosen so that all loops in its range are not very long.)

The restriction of $\tilde{f}$ to the boundary of a small square R_{ij} can be regarded as a map $\tilde{F}$ of a 2-sphere to M^n, where $\tilde{F}$ can be described as follows. First, consider the 2-sphere (that is being mapped by $\tilde{F}$) as a cell complex with the cell structure of the boundary of a parallelipiped in which two opposite (say, the top and the bottom) faces have been collapsed to a point mapped by $\tilde{F}$ into p. (Note that the two copies of p at the beginning and the end of considered loops will be sometimes depicted on our figures as two different points. This convention will make our figures easier to draw and comprehend, and will also make clearer the fact that our proof can be easily adapted to prove the generalization of the theorem where $\Omega_p M^n$ in the conclusion of the theorem is replaced by $\Omega_{pq} M^n$ for two arbitrary points p, q.) To describe $\tilde{F}$ on one of the four "large" cells of the sphere that corresponds to an edge e of R_{ij} with vertices v_1, v_2 consider loops $C_r^{f(v_1)} * f(t)|_{t\in e}(s(r)) * \bar{C}_r^{f(v_2)}$, $r \in [0,1]$, where $s(r)$ is the synchronization function. The lengths of these loops do not exceed $4kd + o(1)$. Applying Lemma 2.1 to this homotopy one obtains a 1-parametric family of paths starting at $C_1^{f(v_1)}$ and ending at $C_1^{f(v_2)}$. These paths are the images of vertical segments (perpendicular to e) of the considered face of S^2 under $\tilde{F}$.

The filling will be done in three steps.

Step 1. For each t denote the the center of the square $R_{ij}\times\{s(t)\}$ by c_t, and $F(c_t)$ by q_t. (Of course, both c_t and q_t depend on $i,\ j$, but we regard i and j as fixed and suppress the dependence on i, j in our notations.) We partition R_{ij} into a 1-parametric family r_t of boundaries of squares that are concentric with R_{ij}, have sides parallel to sides of R_{ij}, and have side length equal to $t\times$ the side length of R_{ij}. In particular, $r_1 = \partial R_{ij}$, and r_0 is the center of R_{ij}. On the first step, we are going to construct a continuous 1-parametric family of maps S^2_t from r_t to $\Omega^{(6k+1)d+o(1)}_{pq_t} M^n$, such that $S^2_1 = \tilde{f}|_{\partial R_{ij}}$, and S^2_0 is the constant map to the constant path p. Here t varies in $[0, 1]$, and a somewhat unusual notation for the maps S^2_t reflects the fact that they correspond to the (maps of) 2-spheres in M^n. (In fact, we prefer to visualize these maps as 2-spheres in M^n.) Note that for each t r_t has four edges that correspond to four edges of R_{ij}. The construction of S^2_t is done separately for each of these four edges (but so that that the restrictions to the endpoints of the adjacent edges match). We are going to describe this construction in the next subsection.

The images of points of r_t (under S^2_t) will be called *vertical curves*. For each value of t the corresponding vertical curves will connect p and q_t.

Note that here we changed terminology that we used in section 3.1. The vertical curves connecting p and q_t are analogs of horizontal curves introduced in section 3.1 (especially for $t = 1$, when $q_t = p$). This change of terminology is due to the fact that these curves look as vertical lines on Figures 16, 17, 18 (which otherwise would be less convenient to draw). The short vertical curves from the previous section are analogous to small "horizontal" squares depicted on Fig. 16 (c), (d), 17.

Summarizing, at the end of this step we obtain a continuous 1-parametric family of maps $S^2_t : S^1 = r_t \longrightarrow \Omega^{(6k+1)d+o(1)}_{pq_t} M^n$ representing a homotopy between $\tilde{f}|_{\partial R_{ij}}$ and a constant map into the constant loop based at p. These maps can be also regarded as 2-spheres M^n sliced into "short" vertical curves connecting p and q_t, where q_t runs over a loop based at p.

Step 2. Each of the vertical curves connecting p and q_t can be joined with a fixed path of length $\leq (2k+1)d + o(1)$ from q_t to p to obtain a sweep-out of each sphere S^2_t by loops of length $\leq (8k+2)d + o(1)$ based at p (or, more presisely, a map from r_t to $\Omega^{(8k+2)d+o(1)}_p M^n$). This fixed path is the image of one of the vertices v of $r_1 = \partial R_{ij}$ under the map constructed on Step 1 but taken with the opposite orientation. More specifically, let F be defined by the formula $F(x, y, t) = f(x, y)(t)$ for all $(x, y) \in R_{ij}, t \in I$. The image of v under S_t^2 is the join of $C_t^{f(v)}$ (from the proof of Theorem 1.5) with the image under F of a very short straight line segment in $R_{ij} \times \{s(t)\}$ that connects the endpoint of $C_t^{f(v)}$ $(= f(v)(s(t)))$ with the center c_τ of $R_{ij} \times \{s(t)\}$. Of course, the image under F of this segment has length $o(1)$. (In other words, it can be made arbitrarily small by choosing a sufficiently fine subdivision of the unit square into squares R_{ij}.) For the reasons of continuity one needs to consider the same vertex v for all values of t. We define v as the vertex of R_{ij} with the largest number in the chosen enumeration of all vertices of all squares R_{ij}. As the result we obtain a continuous 1-parametric family of maps $\tilde{S}_t^2$ from $S^1 = r_t$ to $\Omega_p^{(8k+2)d+o(1)}M^n$ (parametrized by t). Note that all loops corresponding to a fixed value of t pass through q_t.

Step 3. We apply Lemma 2.2 to obtain a 3-disc filling the original 2-sphere, S_1^2, so that this 3-disc is swept-out by paths connecting p and $q_1 = p$ of length at most $(8k+2)d + (2k-1)d + o(1) = (10k+1)d + o(1)$. In other words, this lemma will produce an extension of S_1^2 to a map of a 2-disc to $\Omega_p^{(10k+1)d+o(1)}M^n$ (that can be also regarded as a map $\tilde{F}$ of a 3-disc into M^n). More specifically, Lemma 2.2 will be applied to the map S_1^2 of r_1 regarded as f in the terminology of Lemma 2.2, and to v regarded as s_0. The homotopy F_t in terminology of Lemma 2.2 is the homotopy $\tilde{S}_t^2$ supplied by Step 2. The application of Lemma 2.2 yields the desired contraction of S_1^2 via maps of S^1 into the space of loops based at p that have length $\leq (10k+1)d + o(1)$. (Here we would like to review the geometry behind the application of Lemma 2.2. First, one attaches $\bar{C}_1^{f(v)} * C_1^{f(v)}$ to all vertical paths forming S_1^2. Of course, one does this by means of a homotopy gradually attaching longer and longer arcs of $C_1^{f(v)}$ travelled in both directions. At the end of this stage of the homotopy one obtains $\tilde{S}_1^2 * C_1^{f(v)}$. On the second stage of the homotopy one contracts $\tilde{S}_1^2$ through $\tilde{S}_t^2$ to a point keeping the arc $C_1^{f(v)}$ attached at the end of all curves in the images of $\tilde{S}_1^2$ intact. At the end of this stage all loops in the image of S^1 are mapped into the same loop $C_1^{f(v)}$, and our construction ends by mapping the center of R_{ij} $(=r_0)$ into $C_1^{f(v)}$.)

This completes the construction of the filling $\tilde{f}$ of $\partial R_{ij} \times [0, 1]$ for one small square R_{ij}. Combining these fillings for all values of i, j we obtain a desired map $\tilde{f} : S^2 \longrightarrow \Omega_p M^n$ with the image in $\Omega_p^{(10k+1)d+o(1)}M^n$ (or, equivalently, a map $\tilde{F} : S^3 \longrightarrow M^n$ with a vertical sweep-out (by paths connecting p and $q_1 = p$, that is, by loops based at p) of a controlled length).

After the completion of these three steps we will need to prove that the constructed map $\tilde{f}$ is homotopic to f.

4.2. From $m = 1$ to $m = 2$. A detailed description of Step 1. The restriction of the original map f to R_{ij} induces a map $F : R_{ij} \times [0, 1] \longrightarrow M^n$ defined by the formula $F(x, y, t) = f(x, y)(t)$.

Consider a slicing of $R_{ij} \times I$ into squares $R_{ij} \times \{t\},\ t \in [0, 1]$. We want to ensure that the length of the image under F of each straight line segment of $R_{ij} \times \{t\}$ is much smaller than some $\varepsilon > 0$, which will eventually go to 0, (see Fig. 17(a)). This can be achieved by making the original subdivision into the squares R_{ij} sufficiently fine. Each of the considered slices can be swept-out by short straight line segments as in Fig. 17 (b) in a continuous canonical way. Namely, we connect each vertex v with all points w on the two straight line segments connecting the midpoints of the opposite sides of $R_{ij} \times \{t\}$ such that the distance between v and w is not greater than the distance between w and some other vertex of $R_{ij} \times \{t\}$.

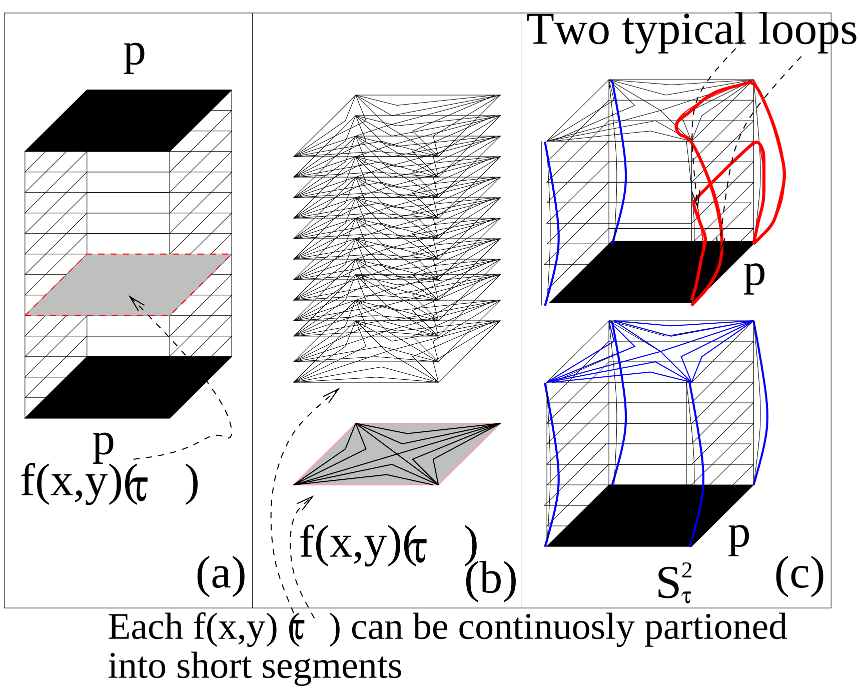

Figure 17: Slicing.

Recall that in the course of the proof of Theorem 3.1 we have replaced the curves corresponding to vertices of R_{ij} by short curves (of length $\leq (2k-$

$1)d + o(1))$, and that we have then constructed path homotopies between the pairs corresponding to the edges of R_{ij}. These homotopies correspond to the edges of R_{ij} and generate 2-discs in M^n. For each edge $e = [v_1 v_2]$ of R_{ij} the corresponding homotopy was produced by an application of Lemma 2.1 to a certain homotopy that contracted the loop formed by joining two paths (in fact, loops) obtained as the "shortening" of images of the vertices. (One of these two paths was taken with the opposite orientation.) This homotopy consisted of the loops $g^t = C_t^{f(v_1)} * F(e,\tau) * \bar{C}_t^{f(v_2)}$ of length $\leq (2k-1)d + (2k+1)d + o(1) = 4kd + o(1)$, where τ denotes $s(t)$, $s(t)$ is the synchronization function, and the arc in the middle is the image under F of the product of e with $\{s(t)\}$ in $R_{ij} \times [0,1]$.

Observe that for every value of $t \in [0,1]$ we can restrict these homotopies, so that they end at g^t instead of g^1 (and connect g^t with the constant loop g^0). So, fix a value of t and consider the restricted homotopy between g^0 and g^t. Now extend this path homotopy by a new stage, when we change only the small middle segement $F(e,\tau)$ of g^t (where τ denotes $s(t)$): We gradually replace it by images under F of broken geodesics in $R_{ij} \times \{\tau\}$ that connect two endpoints of $e \times \{\tau\}$ through varying points $M_{\bar{t}}(\tau)$, $\bar{t} \in [0,1]$, of the perpendicular to $e \times \{\tau\}$ in $R_{ij} \times \tau$ that starts at the midpoint of $e \times \tau$ and ends at the center c_τ of the square $R_{ij} \times \{\tau\}$. Each of those new loops consists of two arcs connecting p and $F(M_{\bar{t}}(\tau))$. Each of these two arcs corresponds to one of the two endpoints v_1, v_2 of e, and starts as $C_t^{f(v_i)}$ followed by the image under F of the straight line segment in $R_{ij} \times \{\tau\}$ that connects (v_i, τ) with $M_{\bar{t}}(\tau)$. Denote these arcs by $A_{v_l}^{\bar{t}}$, $l = 1, 2$. When $\bar{t} = 1$, we obtain $A_{v_1}^1 = C_t^{f(v_1)} * F([(v_1, \tau)\ c_t])$ and $A_{v_2}^1 = C_t^{f(v_2)} * F([(v_2, \tau])\ c_t])$. Joining the first of these arcs with the second arc taken with the opposite orientation, we obtain a loop $\alpha_t = A_{v_1}^1 * \bar{A}_{v_2}^1$. We can proceed similarly for all values of $\bar{t}$ forming loops $A_{v_1}^{\bar{t}} * \bar{A}_{v_2}^{\bar{t}}$. This family of loops constitutes a homotopy between α_t and g^t. We can combine this homotopy with the homotopy between g^t and g^0. The result will be a homotopy between α_t and g^0, that is, between α_t and the point p. Figure 17 (c) depict typical loops that arise during this homotopy. Now we can apply Lemma 2.1 to this homotopy. As the result, we will obtain a path homotopy between $A_{v_1}^1$ and $A_{v_2}^1$ that passes through paths connecting p and q_τ of length less than $(6k+1)d + o(1)$. The map S_t^2 that we are constructing will map the endpoints of the edge e in the boundary of R_{ij} into $A_{v_l}^1$, $l = 1, 2$, and other points of edge e into individual paths in the just constructed path homotopy between $A_{v_1}^1$ and $A_{v_2}^1$. Repeating this process for all four edges of ∂R_{ij} and gluing the resulting maps of the edges into the space of paths connecting p and q_τ, we will obtain a map of the boundary of a square to $\Omega_{pq_\tau}M^n$, which will be the desired S^2_t. Finally, we are going to interpret the domain of the constructed map not as $r_1 = \partial R_{ij}$ but as r_t. This completes the construction of $\tilde{f}$.

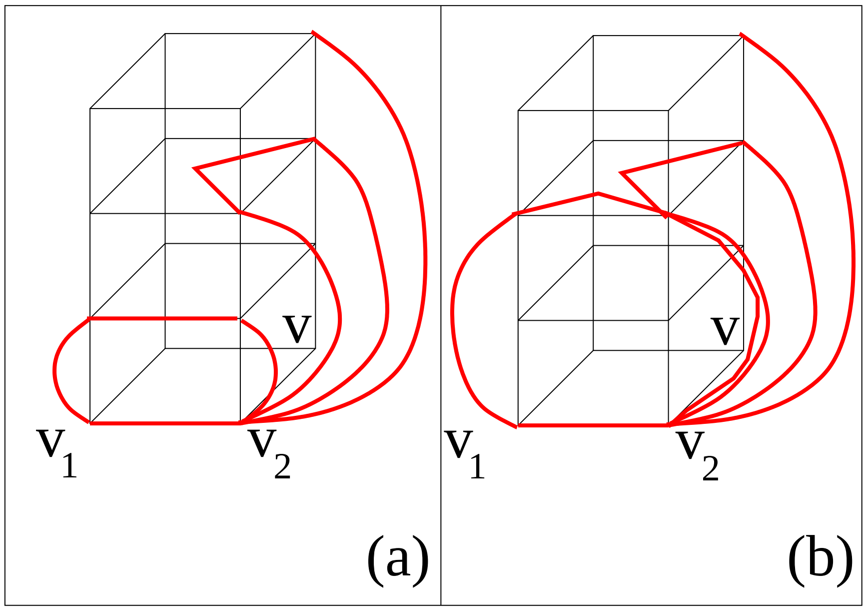

Figure 18: Typical loops.

4.2.A. A description of two typical loops in the image of $\tilde{f}$. Such a loop, ω, must correspond to one of the squares R_{ij} in the chosen fine partition of $[0,1]\times[0,1]$, and, more specifically, to one of its edges $e = [v_1v_2]$. Without any loss of generality assume that v_1 has a smaller number in the chosen enumeration of the set of all vertices of the partition than v_2. Denote the vertex of R_{ij} with the largest number among all vertices of R_{ij} by v. In principle, this vertex might coincide with v_2. We will consider two types of typical loops. These two types incorporate two different types of loops arising during the homotopy contracting α_t described in the previous section. (These loops were depicted on Fig. 17 (c).) Loops ω of both types depend on two parameters: $t_1 \in [0,1]$ and t_2. For loops ω of type A we will be assuming that $t_2 \in [0,t_1]$, for type B $t_2 \in [0,1]$.

Type A. The described loop starts as $C^{f(v_1)}_{t_2}$ followed by the (very short) image of $e\times s(t_2)$ under F. Then ω follows $\bar{C}^{f(v_2)}_{t_2}$, that is, $C^{f(v_2)}_{t_2}$ but travelled in the opposite direction towards its initial point p. Further, ω follows $C_{t_1}^{f(v_2)}$ to $f(v_2)(s(t_1))$. Afterwards, ω passes through the image under F of the half-diagonal of the square $R_{ij} \times \{s(t_1)\}$ that starts at $(v_2, s(t_1))$ and ends at the center of the square, followed by the image under F of the half-diagonal that starts at the center of the square and ends at $(v, s(t_1))$. From there the loop goes along $C_{t_1}^{f(v)}$ but in the opposite direction towards its initial point p. The final arc of ω goes all the way along $C_1^{f(v)}$ to $C_1^{f(v)}(1) = p$.

Type B. The loop starts as $C_{t_1}^{f(v_1)}$ followed by the very short arc that will be denoted $a_1(i, j, e, t_1, t_2)$: This short arc is the image under F of the straight line segment connecting $(v_1, s(t_1))$ with a point $M_{t_2}(s(t_1))$ on the segment perpendicular to $e \times \{s(t_1)\}$ in the plane $R_{ij} \times \{s(t_1)\}$ that connects the midpoint $M_0(s(t_1))$ of $e \times \{s(t_1)\}$ with the center $c_{t_1} = M_1(s(t_1))$ of the square $R_{ij} \times \{s(t_1)\}$. This very short arc is then followed by another very short arc that is the image under F of the straight line segment connecting $M_{t_2}(s(t_1))$ and $(v_2, s(t_1))$. This second short arc will be denoted $a_2(i, j, e, t_1, t_2)$. Then ω follows $\bar{C}_{t_1}^{f(v_2)}$, that is, $C_{t_1}^{f(v_2)}$ but travelled in the opposite direction towards its initial point p. Now ω returns back along $C_{t_1}^{f(v_2)}$ to $f(v_2)(s(t_1))$. Afterwards, ω passes through the image under F of the half-diagonal of the square $R_{ij} \times \{s(t_1)\}$ that starts at $(v_2, s(t_1))$ and ends at the center of the square, followed by the image under F of the half-diagonal that starts at the center of the square and ends at $(v, s(t_1)$. From there the loop goes along $C_{t_1}^{f(v)}$ but in the opposite direction towards its initial point p. The final arc of ω goes all the way along $C_1^{f(v)}$ to $C_1^{f(v)}(1) = p$.

We would like to attract the attention of the reader to two very short arcs $a_1(i, j, e, t_1, t_2)$ and $a_2(i, j, e, t_1, t_2)$ inside the just described loop (of type B). The lengths of these arcs can be made arbitrarily small, as the partion of $[0, 1]^2$ into small squares becomes finer and finer. Yet, the arcs $a_l(i, j, e, t_1, t_2)$ sweep-out the whole image $F([0, 1]^3)$, as i, j, l, e, t_1, t_2 vary over their respective domains. This was precisely the goal of our construction. We wanted to divide the domain of F into arbitrarily small pieces and to distribute them among loops that form a continuous 2-parametric family, so that the lengths of the remaining parts of the loops remain controlled. Below we are going to generalize this construction for an arbitrary m. In this case the image of the constructed family of loops will contain the image of F (which is, in general, $(m + 1)$-dimensional) thanks to the images of similar very short arcs which, however, will be contributing only $o(1)$ summands to lengths of the individual loops in the image of $\tilde{f}$. On the other hand, the union over the whole family of loops of their remaining ("service") parts (made out of curves $C_t^{f(v)}$ for various vertices v) will have image of dimension ≤ 2 in M^n. This contrasts with the fact that for each of these loops the length of the complement to the "service" part will be negligibly small.

The typical loops of Types A and B are shown on Figure 18.

4.3. From $m = 1$ to $m = 2$. Construction of the homotopy between f and $\tilde{f}$. Denote the desired homotopy by G and its parameter by λ. We construct G first over the vertices and the edges of the considered cell subdivision of $[0,1] \times [0,1]$ into 2-cells R_{ij}. This construction is done exactly as in section 3.5. Now it remains to extend G from the boundary of each square R_{ij} to its interior (or, more precisely, from $\partial R_{ij} \times [0,1]$ to $R_{ij} \times [0,1]$). We are going to assume that i, j are fixed, and will describe $G(\lambda)|_{R_{ij}} : R_{ij} \longrightarrow \Omega_p M^n$ for each value of λ.

The informal idea of our construction is that at each moment of time λ we would like to define a map $\tilde{F}_\lambda$ on $R_{ij} \times [0, s(\lambda)]$ exactly as we constructed the map $\tilde{F}$ of $R_{ij} \times [0,1]$. (Recall that $\tilde{F}$ coincides with $\tilde{f}$ when it is regarded as a map from $R_{ij} \times [0,1]$ into M^n, i.e. $\tilde{F}(x,y,t) = \tilde{f}(x,y)(t)$.) In particular, we want $\tilde{F}_1$ to coincide with $\tilde{F}$. On the other hand, we would like the restriction of $\tilde{F}_\lambda$ on $R_{ij} \times [s(\lambda), 1]$ to coincide with F. (So, $\tilde{F}_0 = F$.) The map $G(\lambda)$ will then coincide with $\tilde{F}_\lambda$, when $\tilde{F}_\lambda$ is regarded as a map from R_{ij} to $\Omega_p M^n$.

Yet this plan encounters an obvious (minor) technical problem due to the fact that $F|_{R_{ij} \times \{s(\lambda)\}}$ is not a constant map if $s(\lambda) \neq 1$. However, it is not difficult to make the necessary alterations of this idea using short segments in $F|_{R_{ij} \times \{s(\lambda)\}}$ similarly to how it was done in section 3.5. Here is one possible idea. Recall that for each edge of R_{ij} and each moment of time r the homotopy between the restrictions of f and $\tilde{f}$ to this edge at the moment r consists of a "bottom" part formed by the images under $\tilde{F}$ of arcs of curves in $\partial R_{ij} \times [0, s(\lambda)]$ and "tails" that are curves $f(\varrho, \tau)|_{\tau \in [s(\lambda), 1]}$ for some $\varrho \in \partial R_{ij}$. The extension of the homotopy between f and $\tilde{f}$ to the interior of R_{ij} is done in two stages. The first stage consitutes an extension of the homotopy from ∂R_{ij} to the collar of ∂R_{ij} in R_{ij}; the second stage extends the homotopy to the remaining (inner) part of R_{ij}. The thickness of the collar depends on λ. The collar becomes very thin as $\lambda \longrightarrow 1$ and becomes almost the whole square R_{ij} as $\lambda \longrightarrow 0$. During the first stage we gradually move ϱ towards the center of the square R_{ij} until it reaches the center. We also connect the endpoints of paths in the "bottom" part with the new initial points of "tails" by (short) horizontal curves in $F(R_{ij}, s(\lambda))$.

(One can call this stage "squeezing the tails".) At the "end" of this stage all tails become identical, and all curves from the "bottom" part meet at the same point. Therefore the mentioned technical difficulty disappears, and we can perform the second stage of the homotopy by contracting the "bottom" 2-sphere in M^n (or, equivalently, a circle in $\Omega_p M^n$) proceeding exactly as it was described in sections 4.1, 4.2 but for S^2_λ instead of S^2_1 (and keeping the "tails" fixed).

To more formally describe this construction let λ be in $(0,1)$, Divide R_{ij} into nine rectangles by four straight line segments inside R_{ij} parallel to four sides of R_{ij}. Each of these four segments is located at the distance equal to $\frac{1-\lambda}{2}\times$ the side length of R_{ij} from the corresponding side. Define $G(\lambda)$ on the boundary of the middle square S_{ij} as $\tilde{S}^2_\lambda$ (defined on Step 2; see section 4.1). As it had been already mentioned, in order to define $G(\lambda)$ on the interior of S_{ij} we proceed exactly as in sections 4.1, 4.2 but for S^2_λ instead of S^2_1. As the result, we will obtain a map w of S_{ij} into $\Omega_{p,q_\lambda} M^n$, and then we will form a join of every path in the image of w with $F(c_\lambda,\varrho)|_{\varrho\in[s(\lambda),1]}$, thus obtaining the desired extension of $G(\lambda)$ to S_{ij}. When forming this join we parametrize a path in the image of this map by $[0,\lambda(2-\lambda)]$ and $F(c_\lambda,\varrho)|_{\varrho\in[s(\lambda),1]}$ by $[\lambda(2-\lambda),1]$ (as we have done in a similar situation described in section 3.5).

To explain "proceed exactly as in sections 4.1, 4.2 but for S^2_λ instead of S^2_1" note that we can use the spheres S^2_t, $t\in[0,\lambda]$, as Step 1 of the construction. (Of course, one would need to perform a reparametrization from $[0,\lambda]$ to $[0,1]$). The paths in each of these spheres end at q_t. Again, let v be the vertex of R_{ij} with the maximal number. We are going to use paths $S^2_t(v)$ on Step 2 as before. (We attach $S^2_t(v)$ with the opposite orientation at the end of each path in the image of S^2_t.) Step 3 involves an application of Lemma 2.2 for $s_0=v$ and f_0 (in terminolgy of Lemma 2.2) equal to S^2_λ. After performing all three steps we will obtain a map from S_{ij} to $\Omega_{pq_\lambda} M^n$. (Recall, that one needs to attach $F(c_\lambda,\varrho)|_{\varrho\in[s(\lambda),1]}$ to all paths in the image of this map to obtain a restriction of $G(\lambda)$ to S_{ij}.)

It remains to define $G(\lambda)$ on $R_{ij}\setminus S_{ij}$ that can be viewed as a collar of ∂R_{ij}. Recall that this collar consists of 4 squares adjacent to the vertices of R_{ij} and 4 rectangles adjacent to sides of R_{ij}. We are going to first explain how to define $G(\lambda)$ on a square adjacent to one of the vertices, say v_1, of the square R_{ij}, and then how to define it on a square adjacent to the middle segment of some edge $e_1=[v_1v_2]$. The construction for other vertices/edges will be completely analogous.

Let T_{ij} be a small square adjacent to v_1. Denote another edge of R_{ij}

adjacent to v_1 by e_2. Two sides of the square T_{ij} are segments of the edges e_1, e_2 that are adjacent to v_1. Denote the square with one of the vertices at v_1, and sides that are the halves of the sides of R_{ij} adjacent to v_1 by U_{ij}. Consider the rescaling map $\Phi : T_{ij} \longrightarrow U_{ij}$ such that $\Phi(v_1) = v_1$. We are going to define $G(\lambda)$ at a point $x \in T_{ij}$ as follows: It will be the path that first follows $C_\lambda^{f(v_1)}$, then follows the image under $F(*, s(\lambda))$ of the straight line segment in $U_{ij} \subset R_{ij}$ between v_1 and $\Phi(x)$, and, finally, follows the arc $F(\Phi(x), t)|_{t\in[s(\lambda),1]}$. One can verify that this slice of the homotopy G coinsides with the slices of similar homotopies constructed in section 3.5 for two sides of T_{ij} adjacent to v_1. It is not difficult to verify that the value of $G(\lambda)$ at the vertex of T_{ij} opposite to v_1 coincides with the previously defined value of $G(\lambda)$ at this vertex regarded as a vertex of the "middle" square S_{ij}.

Let W_{ij} denote the rectangle in the considered subdivision of R_{ij} that is adjacent to e_1. We can assume that e_1 is parallel to the X-axis of the plane of R_{ij}. (The case when e_1 is parallel to the Y-axis can be treated in exactly the same way.) Denote the side of W_{ij} which lies on e_1 by e_0. In section 3.5 we have defined $G(\lambda)$ on e_1, and therefore on e_0. On the other hand $G(\lambda)$ has already been defined on the side of W_{ij} opposite to e_0, as it shares this side with S_{ij}. Finally, it had already been defined on two sides perpendicular to e_0 as W_{ij} shares these sides with T_{ij} and a similar square adjacent to v_2. We are going to describe an extension of $G(\lambda)$ to W_{ij} compatible with the already defined $G(\lambda)|_{\partial W_{ij}}$.

Let w denote a point of W_{ij}, $x(w), y(w)$ denote its coordinates, Ψ_0 denote a rescaling map of W_{ij} into a rectangle $\tilde{W}_{ij}$ that shares side e_0 with W_{ij}, has the second side of length equal to the half of the side length of R_{ij}, and is located inside R_{ij}. We also require that the restriction of Ψ_0 on e_0 is the identical map. Further, let Pr denote the orthogonal projection of $\tilde{W}_{ij}$ onto Y-axis, and Ψ denotes the composition of Pr with Ψ_0.

For each y, we will define $G(\lambda)$ on the segment $S_y \subset W_{ij}$ with Y-coordinate y. Denote the midpoint of this segment by m_y and its endpoints by s_1, s_2. Here we are denoting the endpoint that is closer to v_1 than to v_2 by s_1, and the other one by s_2. Observe that we have already defined $G(\lambda)$ at the endpoints of this segment. Both these values are loops that start from certain (different) arcs connecting p with $F(x(m_y), \Psi(m_y), s(\lambda))$. We are going to call these two arcs *initial arcs*. In both cases the corresponding initial arc is then followed by the image under F of the vertical curve $(x(m_y), \Psi(m_y), t)$, where t varies beteen $s(\lambda)$ and 1. Consider loop γ_y obtained as the join of the initial arcs of $G(\lambda)(s_1)$ and $G(\lambda)(s_2)$. In the next paragraph we are going to describe a specific path homotopy of this loop to the constant loop. Then we will apply Lemma 2.1 to construct a path homotopy between the initial arc of $G(\lambda)(s_1)$ and the initial arc of $G(\lambda)(s_2)$. (Here, as elsewhere, we use the numbers of vertices in the numeration of all vertices of all squares to determine which of these arcs is e_1, and which is e_2 in terminology of Lemma 2.1. The initial arc of $G(\lambda)(s_1)$ corresponds to the vertex v_1, and $G(\lambda)(s_2)$ to v_2.) Then we attach the arc $F(x(m_y), \Psi(m_y), t)|_{t\in[s(\lambda),1]}$ to all paths in this path homotopy. The resulting 1-parametric family of paths will constitute the set of values of $G(\lambda)$ on S_y.

It remains to describe the path homotopy contracting γ_y. It consists of two stages. On the first stage we construct a homotopy between γ_y and $C_\lambda^{f(v_1)} * f(t)|_{t\in e_1}(s(\lambda)) * \bar{C}_\lambda^{f(v_2)}$ through loops that also start as $C_\lambda^{f(v_1)}$ and end as $C_\lambda^{f(v_2)}$ taken with the opposite orientation. The middle sections of each intermediate loop in this homotopy is the image under f of the broken line in $R_{ij} \times \{s(\lambda)\}$ made of two straight line segments. These two straight line segments connect $(v_1, s(\lambda))$ and $(v_2, s(\lambda))$ with a variable point moving along the straight line segment connecting the center of $R_{ij} \times \{s(\lambda)\}$ with the midpoint of its edge $e_1 \times \{s(\lambda)\}$. On the second stage we contract $C_\lambda^{f(v_1)} * f(x, y_{e_1})|_{x\in[x(v_1),x(v_2)]}(s(\lambda)) * \bar{C}_\lambda^{f(v_2)}$ through loops $C_r^{f(v_1)} * f(x, y_{e_1})|_{x\in[x(v_1),x(v_2)]}(s(r)) * \bar{C}_r^{f(v_2)}$, where r decreases from λ to 0, and y_{e_1} denotes the common value of the Y-coordinate for all points of $e \subset R_{ij}$. (Recall, that we assumed that e_1 is parallel to the X-axis.)

As in section 3.5, the lengths of the loops in the image of this homotopy can be bounded by $L + (10k + 3)d + o(1) - (2k - 1)d$. Here we subtract $(2k-1)d$, because one of the segments of curves C shortens a segment whose length is counted as a part of the term L.

4.4. Proof in the general case (m is arbitrary): Construction of $\tilde{f}$.

Proof. We will present a proof only in the case when f is a map of S^m. The proof in the case when f is a map of the pair $(D^m, \partial D^m)$ is completely analogous.

Let $f : I^m \longrightarrow \Omega_p M^n$ be a continuous map such that ∂I^m is mapped to p. In this section we are going to construct $\tilde{f} : I^m \longrightarrow \Omega_p^{((4k+2)m+(2k-3))d+o(1)} M^n$. In the next section we will prove that $\tilde{f}$ and f are homotopic.

We are going to consider a map $F : I^{m+1} \longrightarrow M^n$ defined as $f(u)(t)$ for each $u \in I^m$, where I^m corresponds to the first m coordinates of I^{m+1}. We can partition I^m into m-cubes $R_{i_1,\ldots,i_m}$, so that for each t the image of each straight line in $R_{i_1,\ldots,i_m} \times \{t\}$ under F has an arbitrarily small length. We enumerate the set of all vertices of all cubes $R_{i_1,\ldots i_m}$ by consecutive natural numbers. We are going to consider shortenings and partial shortenings of all paths $f(v)$, where v runs over the set of all vertices of all cubes $R_{i_1,\ldots,i_m}$. We are going to perform the synchronization of all these paths as in section 3.4, and for each vertex v define $\tilde{f}(v)$ as $C_1^{f(v)}$. Now we are going to extend $\tilde{f}$ to cells of higher and higher dimension of the considered cell subdivision of I^m. We are going to describe this construction for faces of one cube $R_{i_1,\ldots,i_m}$. Of course, the extension on each face E of this cube should be independent on a particular choice of the ambient cube $R_{i_1,\ldots,i_m}$ that contains E as its face.

We will need the following definitions. Let E be a face of $R_{i_1,\ldots,i_m}$, $t \in [0,1]$. We are going to call $E \times \{s(t)\} \subset I^{m+1}$ the *$t - E$-slice* of I^{m+1}. (Here $s(t)$ denotes the synchronization function.) Given E and t the set of points $x \in R_{i_1,\ldots i_m} \times \{s(t)\}$ such that all the distances from x to different vertices of the $t - E$-slice are equal is called *the median* of the $t - E$-slices. Equivalently, a point of $R_{i_1,\ldots,i_m} \times \{s(t)\}$ is in the median of the $t - E$-slice if and only if it projects to the center of $t - E$-slice under the orthogonal projection on the $t - E$-slice. A *reduced median*, M_{tE}, of the $t - E$ slice is, by definition, the subset of the median of the $t - E$-slice that consists of all points x with the following property:The distance from x to any vertex of the $t - E$-slice is less than or equal to the distance from x to any other vertex of $R_{ij} \times \{t\}$. For each face E of $R_{i_1,\ldots,i_m}$, $t \in [0,1]$, each point $o \in M_{tE}$ and each vertex v of E *the $v - t - o$-vertical curve* is formed as a join of $C_t^{f(v)}$ and the image under F of the straight line segment connecting $(v, s(t)) \in R_{i_1,\ldots,i_m} \times \{s(t)\}$ and o. It starts at $p = F(v,0)$ and ends at $F(o)$. The collection of $v - t - o$-vertical curves for all vertices v of E (and some fixed t and $o \in M_{tE}$) is called *the $E - t - o$-umbrella*. (We call these objects umbrellas because $v - t - o$-vertical curves remind us of umbrella ribs. Further, we will later define the process of “filling of an umbrella”. The main step of this process reminds us of attaching of an umbrella canopy to its frame.) In this definition we are assuming that all $v-t-o$-vertical curves forming the $E - t - o$-umbrella are indexed by the corresponding vertices of E. Equivalently, each $E - t - o$-umbrella u is, by definition, endowed with a map Z_u from the set $V(E)$ of all vertices of E to the set of all vertical curves that form the umbrella, where Z_u maps each vertex to the corresponding vertical curve. If $E_1 \subset E$ is a subface of E, we can consider the restriction of Z_u to $V(E_1) \subset V(E)$. We will call the resulting umbrella a *face* of u.

We denote the union of the sets of all $E-t-o$-umbrellas over all $t \in [0,1]$ and $o \in M_{tE}$ by U_E. Note that for every E, t and each point $o \in M_{tE}$ there is a unique $E-t-o$-umbrella. So, we can introduce the map $Y = Y_E$ from the set P_E of all pairs (t,o), where $t \in [0,1]$, $o \in M_{tE}$ to U_E. One can endow U_E by a C^0-topology. In this case Y becomes a continuous map.

Further, we define *the canonical path* from a point $(t,o) \in P_E$ as the broken line that consists of two straight line segments: The value of t is constant on the first straight line segment. The second coordinate varies along the straight line segment that connects o with the center of the $t-E$-slice. If o is the center of the $t-E$-slice then the first straight line segment will be reduced to just one point. The second straight line segment starts at the center of the $t-E$-slice and passes through the centers of all $r-E$-slices, where r decreases from t to 0, and ends at the center of the $0-E$-slice. The composition of Y and the canonical path is called *the canonical homotopy* of E-umbrellas. For each $E, t, o \in M_{tE}$ the canonical homotopy is a path in U_E that starts at the $E-t-o$-umbrella $Y(t,o)$ and ends at the constant E-umbrella, where all verical curves are constant curves equal to p.

Let o be a point in M_E. A *filling* of $E-t-o$-umbrella $Y(t,o)$ is a continuous map from E to $\Omega_{po}M^n$ that coincides with Z_u on $V(E) \subset E$. Now we are going to define the *canonical filling* of all $E-t-o$-umbrellas for all faces E of $R_{i_1,\ldots,i_m}$ (including $R_{i_1,\ldots,i_m}$), all t and all $o \in M_E$. Our goal is to define a filling that for every E will depend continuously on umbrellas in U_E, and will have the following *coherence* property: If E_1 is a subface of E, then the restriction of the canonical filling defined for E to E_1 must coincide with the canonical filling defined for E_1. Here is a more detailed explanation of coherence. One can regard a canonical filling as a system of functions defined for all faces E of $R_{i_1,\ldots,i_m}$. For each E the corresponding function maps each triple $x,\ t,\ o$, where $x \in E,\ t \in [0,1],\ o \in M_{tE}$ to a path connecting p and o in M^n. The coherence means that if E_1 is a subface of E_2, then the restriction of the function defined for E_2 to the set of triples $\{(x,t,o)|x \in E_1,\ t \in [0,1], o \in M_{tE_2}(\subset M_{tE_1})\}$ coincides with the restriction of the function defined for E_1 to ther same set of triples.

Our construction of the canonical filling will be recursive with respect to the dimension of the considered faces. After defining the canonical filling for all faces with the next value of dimension we will need to verify the continuity and coherence properties of the constructed canonical filling. If $dim\ E = 0$ (i.e. E consists of just one vertex of $R_{i_1,\ldots,i_m}$), then the canonical filling of an $E-t-o$-umbrella coincides with this $E-t-o$-umbrella. If

$dim\ E = 1$, and $E = [v_1, v_2]$ for some two vertices v_1, v_2 of $R_{i_1,\dots i_m}$, and u is a $E - t - o$-umbrella, then the canonical filling is defined as follows: Consider the canonical homotopy between u and the trivial umbrella. All the umbrellas in the image of this homotopy consist of two vertical curves, and one can consider this homotopy as a contraction of the loop formed by two vertical curves forming u. One can use Lemma 2.1 to obtain a path homotopy between two vertical curves forming u. As before, we denote the vertical curve corresponding to one of the vertices v_1, v_2 that has a smaller number is the chosen enumeration of all vertices by e_1 (in terminology of Lemma 1.2), and the other by e_2. The filling of u provided by this path homotopy is, by definition, the canonical filling of u. The coherence of the canonical filling can be immediately seen from the construction. The continuity follows from the continuity of the dependance on t of all paths $C_t^{f(v_i)}, i = 1, 2$.

Assume that the canonical filling has been already defined for all faces up to a dimension $l - 1$, where $l \geq 2$. We are going to define it for a l-dimensional face E. Let u be a $E - t - o$-umbrella, where $o \in M_{tE}$ for some $t \in [0, 1]$. Consider the canonical homotopy u_r of u, where $r \in [0, 1]$, $u_0 = u$, u_1 is the trivial umbrella. For each value of r consider the canonical fillings of all $(l - 1)$-dimensional faces of u_r. The result will be a collection of maps of all $(l - 1)$-dimensional faces of E into $\Omega_{po_r} M^n$, where o_r is the common endpoint of all vertical curves forming u_r. The coherence property for the canonical fillings of all $(l - 1)$-faces implies that the constructed maps of the $(l - 1)$-faces match on the intersections of these faces and can be considered as a map S_r^l of the boundary of a l-cube (or, topologically, of a $(l - 1)$-dimensional sphere). The unusual notation reflects the fact that these maps can be also regarded as maps of l-dimensional spheres into M^n. A reader is invited to check that the just described part of the construction corresponds to "Step 1" introduced in section 4.1 and described in section 4.2. Now we are going to proceed similarly to Steps 2 and 3 described in section 4.1. Namely, we first choose the vertex v of $R_{i_1,\dots,i_m}$ with the maximal number in the choosen enumeration of all vertices of all cubes and attach the vertical curve corresponding to v in $Y_E(r, o_r)$ at the end of all paths in the image of S_r^l. We attach this vertical curve with the opposite orientation. Now all curves in the image of this new 1-parametric family $\tilde{S}_r^l$ of maps of $(l - 1)$-dimensional spheres end not at o_r but at p, and become loops based at p. This is the analog of Step 2. Finally, we apply Lemma 2.2 (as on Step 3 described in section 4.2). The family $\tilde{S}_r^l,\ r \in [0, 1]$ plays the role of the homotopy F_t in the terminology of Lemma 2.2, and the vertical curve in u corresponding to the vertex v with the largest number plays the role of $f(s_0)$. Lemma 2.2 will yield the desired canonical filling of u. The continuity and the coherence of the just constructed canonical fillings can be easily seen from the construction.

To estimate the length of paths in the canonical filling note that when one disregards $o(1)$ terms, the increase of the length happens on "Steps 2 and 3" of the just described construction. Each time we add not more than $(2k+1)d$ to our estimate for the length resulting in the extra $(4k+2)d$ when we pass from the dimension $(l-1)$ to the dimension l (of the face E). (Recall, that our estimate for lengths of vertical curves $C_t^{f(v)}$ is $(2k+1)d$. In some cases this estimate improves to $(2k-1)d$, for example, if $t=1$.) Thus, it seems that the total length of paths in the image of the canonical filling of u is bounded by $(2k+1)d+l(4k+2)d+o(1)$. However, we can slightly improve this estimate, if we notice that as the result of the synchronization the lengths of all but one vertical curves forming u are bounded by $(2k-1)d+o(1)$, and the length of the remaining vertical curve is bounded by $(2k+1)d+o(1)$. As a corollary, in the worst case scenario our upper bound becomes $(2k-1)d+l(4k+2)d+o(1)$. A further improvement can be achieved if u is $E-1-o$-umbrella (that is, $t=1$). In this case we can shave an extra $2d$ from our estimate on Step 3 resulting in the upper bound $(4k+2)ld+(2k-3)d+o(1)$ that appears (for $l=m$) in the text of Theorem 1.1.

Now we can finally define $\tilde{f}$ on $R_{i_1,\dots,m}$. We define $\tilde{f}$ as the canonical filling of the $E-t-o$-umbrella u_0, where $t=1$, $E=R_{i_1,\dots i_m}$, and o is the center of $R_{i_1,\dots i_m}\times\{1\}$.

Observe that our construction of $\tilde{f}$ for $m=1$ coincides with the construction described in section 3.3, and for $m=2$ coincides with the construction described in sections 4.1, 4.2.

4.5. Proof in the general case (m is arbitrary): Construction of a homotopy between f and $\tilde{f}$.

Denote the desired homotopy by G. For each cube $R_{i_1,\dots,i_m}$ the construction will be inductive with respect to the dimension of the considered faces of $R_{i_1,\dots,i_m}$. It can be defined at all vertices v as partial shortenings $H_{f(v)}(t)$ introduced in section 3.2. At each induction step corresponding to a value of the dimension l between 1 and m we construct G over all l-dimensional faces of all cubes $R_{i_1,\dots,i_m}$ assuming that it had already been constructed over all $(l-1)$-faces. For every l-face $E=R_{i_1,\dots i_l}$ G is supposed to map $R_{i_1,\dots,i_l}\times[0,1]$ to Ω_pM^n. Denote the restriction of G to $R_{i_1,\dots,i_l}\times\{\lambda\}$ by

$G(\lambda)$ for each $\lambda \in [0,1]$. We need $G(0)$ to coincide with the restriction of f to $E = R_{i_1,\dots,i_l}$ and $G(1)$ to coincide with the restriction of $\tilde{f}$. We will need to define $G(\lambda)$ for every $\lambda \in (0,1)$.

The basic idea will be the same as the idea previously used in sections 3.2, 3.5, 4.3: We would like to perform a partial shortening of all paths in the image of f up to parameter λ, and then attach the unchanged remaining part of the path for t between λ and 1. As before, this plan encounters a difficulty: We can shorten only a disc in the space of paths that end at the same point. Yet if we cut all paths in $f(E)$ at some value of $t \neq 1$, they will end at sufficiently close points but not at the same point. Therefore we will first need to "squeeze" the "tails" a bit, so that they would end at the same point. We use a collar of the boundary of E to do that. The width of this collar decreases to 0 as $\lambda \longrightarrow 1$. Then we define a map of the square at the center of E into $\Omega_p M^n$ exactly as we have defined $\tilde{f}$, only for the initial arcs of f (followed by some short arcs to make them end at the same point). The restriction of $G(\lambda)$ to this central square will then be defined by attaching to all paths in the image of this map the segments identical to the "tail" defined as the arc of the image of the center of E under F for $t \in [s(\lambda), 1]$.

Here are the details. Divide E by $2l$ $(l-1)$-dimensional planes parallel to faces of E into 3^l parallelipipeds. Each of these new hyperplanes corresponds to exactly one of $2l$ $(l-1)$-dimensional faces of E and is located at the distance equal to $\frac{1-\lambda}{2}\times$ the side length of E from the corresponding $(l-1)$-dimensional face. Let S denote the central l-dimensional cube of this partition of E, T denote one of the corner cubes adjacent to a vertex v_1 of E, and W denotes one of the parallelipipeds forming the collar that is not adjacent to a vertex. We are going to explain the construction of the restrictions of $G(\lambda)$ to S, T and W.

To define $G(\lambda)$ on S we observe that each vertex s is the closest to exactly one vertex, v_s of E. Consider a vertical curve $C_\lambda^{f(v_s)}$ followed by the image under F of the straight line segment in $E \times \{s(\lambda)\}$ that connects the endpoint of $C_\lambda^{f(v_s)}$ with the center c_λ of $E \times \{s(\lambda)\}$. When we combine all these vertical curves for all vertices of S, we obtain a $E-\lambda-c_\lambda$-umbrella. Consider the canonical filling of this umbrella. Reparametrize all paths in this canonical filling by the interval $[0, \lambda(2-\lambda)]$ and attach at the end the arc $f(c_\lambda)(t)|_{t\in[s(\lambda),1]}$ reparametrized by the interval $[\lambda(2-\lambda), 1]$. (By definition, if a path p is defined for $t \in [0,1]$, then $p(\frac{t-a}{b-a})$ is regarded as a reparametrization of p by $[a,b]$.) We define $G(\lambda)|_S$ as the rescaling of the resulting map from E to $\Omega_p M^n$ to S. (In other words, we take the composition of this map with the affine map $S \longrightarrow E$ that sends each vertex s of S to v_s.)

To define $G(\lambda)$ on T we proceed as follows. First, define an affine map Φ that rescales T into a larger cube with the same vertex v_1 and the side length equal to the half of the side length of E. Let b be a point of T. The curve $G(\lambda)(b)$ will consist of three arcs. The first arc will be $C_\lambda^{f(v_1)}$ reparametrized by $[0,\lambda]$. The second arc will be the image under F of the straight line segment connecting $(v_1, s(\lambda))$ with $(\Phi(b), s(\lambda))$. This arc will be parametrized by $[\lambda, \lambda(2-\lambda)]$. Finally, we attach $F(\Phi(b),t)|_{t\in[s(\lambda),1]}$ that will be reparametrized by $[\lambda(2-\lambda),1]$.

It remains to define $G(\lambda)$ on W. For each of l cartesian cooordinates of E consider the projection of W onto one of sides of E parallel to the corresponding coordinate axis. Denote this side of E by s_i. The result can be a segment of length equal to $\frac{1-\lambda}{2}\times$ the side length of E adjacent to one of the vertices of s_i. Collect all such coordinates in a linear space P_1. Or, alternatively, the result will be a segment of length $\lambda\times$ the side length of $R_{i_1,\ldots,i_l}$ in the interior of s_i. Collect all such coordinates in a linear space P_2. The direct sum of P_1 and P_2 will be the whole space R^l parallel to E. Let $O_i,\ i=1,2$ denote the orthogonal projections of R^l onto P_i. We can slice W into isometric copies of $O_2(W)$, namely $O_2(W)+p_1$, where p_1 varies over $O_1(W)$. We are going to define $G(\lambda)$ on each of these copies of $O_2(W)$ separately (but so that the resulting map $G(\lambda)$ on W will also depend continuously on p_1). Note that $A=O_2(W)$ is a face of $E=R_{i_1,\ldots,i_l}$. Denote the center of the cube A by m_A. Consider the affine rescaling Ψ of W that preserves P_2-coordinates, fixes $W\bigcap\partial R_{i_1,\ldots,i_l}$ and maps the $(l-1)$-face of W opposite to $W\bigcap\partial R_{i_1,\ldots,i_l}$ to the affine subspace of R^l, where the values of all P_1-coordinates are equal to those of the center of the cube $R_{i_1,\ldots,i_l}$. Define $m_\lambda(p_1)$ as $m_A+\Psi(p_1)$, and $o_\lambda(p_1)$ as $(m_\lambda(p_1), s(\lambda))$. Consider the $(A-\lambda-o_\lambda(p_1))$-umbrella $u=Y_A(\lambda, o_\lambda(p_1))$. Now consider the canonical filling of u. Reparametrize each curve in the image of the canonical filling of u by $[0,\lambda(2-\lambda)]$ and then extend it by the image under F of the straight line segment in $R_{i_1,\ldots,i_l}\times[0,1]$ connecting $o_\lambda(p_1)$ and $(m_\lambda(p_1),1)$. This last arc should be reparametrized by $[\lambda(2-\lambda),1]$. Take the composition of the resulting map $A\longrightarrow\Omega_pM^n$ with the translation $A+p_1\longrightarrow A$. Now define $G(\lambda)$ on $A+p_1$ as the resulting map.

This completes our construction of G. $\square$

5 Short geodesic segments connecting pairs of points

In this section we will prove that for each pair of points on a closed Riemannian manifold there exist "many" "short" geodesic segments that join the points. This fact follows directly from the following lemmas, which are restatements of the similar lemmas for geodesic loops.

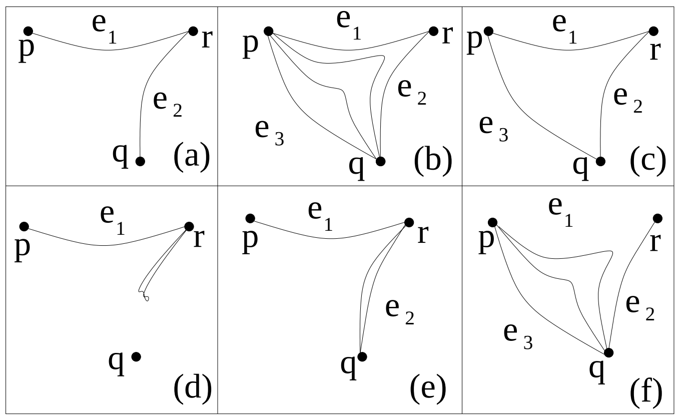

Figure 19: Short path homotopy

Lemma 5.1 *Let M^n be a closed Riemannian manifold of diameter d. Let e_1, e_2 be segments of lengths l_1, l_2 respectively, where e_1 connects a point p with a point r and e_2 connects the point r to a point q, (see Fig. 19 (a)). Consider the join $e_1 * e_2$. Assume that it is path homotopic to a path e_3 of length $l_3 \leq l_1 + l_2$ via a length non-increasing path homotopy, (see Fig. 19 (b)).*

Then there exists a path homotopy between e_1 and $e_3 * \bar{e_2}$, (Fig. 19 (c)) that passes through curves of length at most $l_1 + 2l_2$.

Proof. The proof is essentially demonstrated by Fig. 19 (d)-(f). We begin with e_1, (Fig. 19 (d)), which is homotopic to $e_1 * e_2 * \bar{e_2}$ over the curves of length $l_1 + 2l_2$, (Fig. 19 (e)). Since $e_1 * e_2$ is path homotopic to e_3, and the path homotopy does not increase the length, we can attach $\bar{e_2}$ to all paths in this path homotopy to obtain a homotopy between $e_1 * e_2 * \bar{e_2}$ and $e_3 * \bar{e_2}$, (see Fig. 19 (f)). $\square$

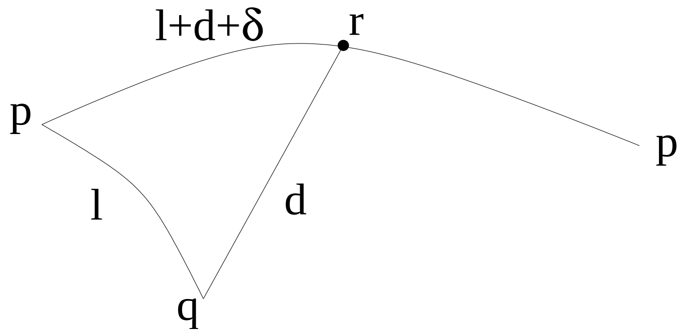

Figure 20: Modified length shortening process

Theorem 5.2 *Let M^n be a closed Riemannian manifold, p, q, x points of M^n. Let $\gamma(t)$ be a curve of length L starting at p and ending at x. Then, if there exists an interval $(l, l + 2d]$, such that there is no geodesic connecting p and q on M^n of length in this interval providing a local minimum of the length functional on $\Omega_{pq}M^n$, then there is a path-homotopy between $\gamma(t)$ and a path $\tilde{\gamma}(t)$ of length at most $l + d$ passing through curves of length at most $L + 2d$.*

Proof. The proof relies on the previous lemma, but is otherwise analogous to the proof of Theorem 1.5, (see Fig. 20).

For example, here is an adaptation of the first step of the curve shortening process described in our proof of Theorem 1.5. By compactness there exists a small δ, such that there are no short geodesics connecting p and q of length in the interval $(l, l + 2d + \delta]$. Consider a segment e_1 of the original curve of length $l + d + \delta$ connecting p with some point r. Let us denote a minimizing geodesic connecting the point r with the point q by e_2. The curve $e_1 * e_2$ is path homotopic to e_3 of length at most l. (Here we define e_3 as a shortest path which is path homotopic to $e_1 * e_2$ via a length non-increasing homotopy.) Therefore, by the previous lemma there is a path homotopy between e_1 and $e_3 * \bar{e_2}$ of length at most $l + d$ over the curves of length at most $l + 2d$.

□

The above result has the following corollaries.

Theorem 5.3 *Let M^n be a closed Riemannian manifold of diameter d, p, q, x be points of M^n. Assume that there exists $k \in N$ such that* *there is no geodesic of length in the interval in the interval* $(dist(p,q) + (2k-2)d, dist(p,q) + 2kd]$ *joining* p *and* q *that is a local minimum of the length functional on* $\Omega_{pq}M^n$*. Then for every positive integer* m *every map* $f : S^m \longrightarrow \Omega_{px}M^n$ *is homotopic to a map* $\tilde{f} : S^m \longrightarrow \Omega_{px}^{L+o(1)}M^n$*, where* $L = ((4k+2)m + (2k-3))d + (2m+1)dist(p,q)$*. Furthermore, every map* $f : (D^m, \partial D^m) \longrightarrow (\Omega_{px}M^n, \Omega_{px}^L M^n)$ *is homotopic to a map* $\tilde{f} : (D^m, \partial D^m) \longrightarrow \Omega_{px}^{L+o(1)}M^n$ *relative to* ∂D^m*. In addition, if for some* R *the image of* f *is contained in* $\Omega_{px}^R M^n$*, then the homotopy between* f *and* $\tilde{f}$ *can be chosen so that its image is contained in* $\Omega_{px}^{R+(4k+2)md+(2m+1)dist(p,q)+o(1)}M^n$*. Also, in this case for every* $R > 0$ *every map* $f : S^0 \longrightarrow \Omega_{px}^R M^n$ *is homotopic to a map* $\tilde{f} : S^0 \longrightarrow \Omega_{px}^{(2k-1)d+dist(p,q)}$ *by means of a homotopy with the image inside* $\Omega_{px}^{R+2d}M^n$*.*

This theorem can be proven exactly as Theorem 1.1 but using the previous theorem instead of Theorem 1.5 on the very first step of induction (from $m = 1$ to $m = 2$). As before, $o(1)$ denotes an arbitrarily small positive summand, and all $o(1)$ terms can be omitted, if M^n is an analytic Riemannian manifold.

Corollary 5.4 *Let* M^n *be a closed* $(m-1)$*-connected Riemannian manifold of diameter* d *with a non-trivial* m*th homotopy group. Then if for some pair of points* p, q *there exists* $k \in N$*, such that no geodesic connecting* p *and* q *has the length in the interval* $(2(k-1)d + dist(p,q), 2kd + dist(p,q)]$ *and is a local minimum of the length functional on* $\Omega_{pq}M^n$*, then the length of a shortest non-trivial periodic geodesic on* M^n *is at most* $((4k+2)m + (2k-3))d + (2m+1)dist(p,q) \leq (4km + 2k + 4m - 2)d$*.*

Proof. By Theorem 5.3 we can construct a non-contractible sphere of dimension m in the space of loops based at p that is swept-out by closed curves of length at most $((4k+2)m+(2k-3))d+(2m+1)dist(p,q)+o(1)$. Now the standard proof of the Lyusternik-Fet theorem establishing the existence of a non-trivial periodic geodesic on every closed manifold (cf. [Kl]) will yield the desired upper bound for the length of a shortest periodic geodesic. □

This corollary vastly generalizes previous results by F. Balacheff ([B]) and R. Rotman ([R2]) in some directions. These previous results correspond to the cases $k = 1$, $p = q$ and $m = 1$ ([B]) and $k = 1$, $p = q$ and $m = 2$ ([R2]). Yet the upper bound for the length of a shortest non-trivial periodic geodesic provided by the corollary in these two cases is somewhat worse than the corresponding bounds in the quoted papers ($5d$ versus $4d$ in [B], and $11d$ versus $6d$ in [R2]).

6 Non-simply connected case

Here we prove Theorem 1.3, which is required to complete the proof of Theorem 1.2 in the non-simply connected case, as well as its generalization Theorem 1.4.

First, recall a result of Gromov ([Gr]) asserting that for every closed Riemannian manifold M of diameter d and every point $p \in M$ there exists a finite presentation of $\pi_1(M^n)$ such that all its generators can be realized by geodesic loops of length $\leq 2d$ based at p. This result will be repeatedly used in this section.

Definition 6.1 *Let G be a finitely presented group. A word in generators of G and their inverses is called minimal if the element of G presented by this word cannot be presented by a word of smaller length. The complexity of an element of G with respect to the considered finite presentation is the length of a minimal word representing this element. The complexity of the trivial element is, by definition, zero.*

Proposition 6.2 *Let G be a finitely presented group. Assume that there exists an element $h \in G$ of complexity $m \geq 1$. Then G has at least $2m$ elements: the trivial element e, h, and at least two elements of complexity i for every $i = 1, 2, \dots, m-1$.*

This proposition has the following immediate corollary:

Corollary 6.3 *Assume that G is a finitely presented finite group of order l. Then the complexities of elements of G do not exceed $\frac{l}{2}$. If there exists an element of complexity $\frac{l}{2}$, then it is unique.*

Proof of Proposition 6.2. We will start from the following observation that will be repeatedly used in our proof: Any subword of a minimal word is minimal.

Now assume that h can be represented by a minimal word starting with a positive power of a generator a. Among all minimal words representing h and starting with a^j for some j choose a word w for which j is maximal possible.

Case 1. If $w = a^m$, then for every i between 1 and $m-1$, the words a^i and a^{-i} are minimal and represent different words of complexity i. (Indeed, if $a^r = a^{-s}$ for $r, s \in \{-(m-1), \ldots, m-1\}$, $r \neq -s$, then $a^{r+s} = a^{-(r+s)} = a^{|r+s|} = e$. As $|r+s| \leq 2m-2$, a^m can be represented by a shorter word $a^{m-|r+s|}$, which contradicts the minimality of a^m.)
Case 2. $w = a^k b l_1 \ldots l_{m-k-1}$, where b, l_1, $\ldots, l_{m-k-1}$ are generators of G or their inverses. Moreover, we can assume that b is not equal to a power of a.

Now we can consider $2k$ words a^i, a^{-i} for $i = 1, \ldots, k-1$. We have two distinct words of complexity i for each considered value of i in this set.

For every value of $i \in \{k, \ldots m-1\}$ consider the initial subword of w of length i, and the subword of w of length i starting from the second letter. For example, for $i = k$ we will be considering a^k and $a^{k-1}b$, for $i = k+1$ $a^k b$ and $a^{k-1} b l_1$. These words are minimal and represent elements of G of complexity i. We need only to verify that they are not equal to each other. But if $a^{k-1} b l_1 \ldots l_{i-k} = a^k b l_i \ldots l_{i-k-1}$, then we can replace the subword $a^{k-1} b l_1 \ldots l_{i-k}$ by $a^k b l_i \ldots l_{i-k-1}$ in w and obtain the word $a^{k+1} b l_1 \ldots l_{i-k-1} l_{i-k+1} \ldots l_{m-k-1}$ of length m representing h but starting from a higher power of a than w. This contradicts the definition of w. $\square$

Proof of Theorem 1.3. Let $\tilde{p}, \tilde{q}$ be two points of the universal covering $\tilde{M}$ of M. We know that $\tilde{M}$ can be tiled by isometric connected fundamental domains of radius d centered at points $\tilde{p}_i$ that project to the same point $p \in M$ as $\tilde{p}$. The interiors of these domains are Voronoi cells of $\tilde{p}_i$, i.e. sets of points x of $\tilde{M}$ for which $\tilde{p}_i$ is the (unique) closest point to x in the inverse image of p.

Observe that every point s in a connected fundamental domain $S \subset \tilde{M}$ is within distance d of the boundary of S. Indeed, let y be a point in $P(\partial S)$, where P denotes the universal covering map $\tilde{M} \longrightarrow M$. There exists a curve γ of length $\leq d$ connecting $P(s)$ and y in M. This curve lifts to a curve $\tilde{\gamma}$ between s and a lift $\tilde{y}$ of y. If $\tilde{y} \in S$, then $\tilde{y} \in \partial S$, and we are done. Otherwise, $\tilde{y} \notin S$, and so $\tilde{\gamma}$ must at some point cross the boundary of S. This point is a point of ∂S within the distance d of s. (This proof of the observation stated at the beginning of this paragraph was suggested to us by an anonymous referee of this paper. We would like to thank the referee for this suggestion.)

We can assume that $\tilde{p} = \tilde{p}_1$ is the base point of $\tilde{M}$. The number of these fundamental domains is equal to the cardinality C of $\pi_1(M)$. Correspondingly, each of these fundamental domains contains a point $\tilde{q}_i$ that projects to the same point of M as $\tilde{q}$. Assume that $\tilde{q} = \tilde{q}_j$ for some $j \in \{1, \dots C\}$.

Note that for every j $\tilde{p}_j$ corresponds to an element g_j of $\pi_1(M)$. Vice versa for every element $g \in G$ we can define the corresponding $\tilde{p}_j$ by lifting a loop representing g; $\tilde{p}_j$ will be the endpoint of the lifted loop and will not depend on the choice of a loop representing g. If a finite presentation of $\pi_1(M)$ is chosen, all of its generators are represented by loops in M of length $\leq 2d$ based at p, and g_j is represented by a word v of length l in the generators of M and their inverses, then the distance between $\tilde{p}$ and $\tilde{p}_j$ cannot exceed $2dl$, as $\tilde{p}_j$ is the endpoint of the result of lifting to $\tilde{M}$ of a join of l loops in M representing generators of $\pi_1(M)$ and their inverses combined exactly as the corresponding letters in v. Let u be the maximal complexity of an element of $\pi_1(M)$ with respect to the chosen finite presentation of $\pi_1(M)$. Corollary 6.3 implies that either $u < \frac{C}{2}$, and, therefore, $u \leq \frac{C-1}{2}$, or $u = \frac{C}{2}$, but there is only one element of this complexity, and the complexity of the other elements does not exceed $\frac{C}{2} - 1$. In the first case, $dist(\tilde{p}_1, \tilde{p}_j) \leq 2d(\frac{C-1}{2}) \leq d(C-1)$, and $dist(\tilde{p}, \tilde{q}) = dist(\tilde{p}_1, \tilde{q}_j) \leq dist(\tilde{p}_1, \tilde{p}_j) + dist(\tilde{p}_j, \tilde{q}_j) \leq d(C-1) + d = Cd$. In the second case, $dist(\tilde{p}, \tilde{q}) \leq Cd$ by the same argument unless the element of $\pi_1(M^n)$ corresponding to $\tilde{p}_j$ has complexity $\frac{C}{2}$.

In this last case, we are first going to recall that the distance from a point z in a fundamental domain to the boundary of this domain is at most d. Now denote one of the points closest to $\tilde{q}_j$ in the boundary of its fundamental domain by ϱ. The distance between $\tilde{q}_j$ and ϱ does not exceed d. The point ϱ must be in the closure of another fundamental domain centered at $\tilde{p}_m$ for some $m \neq j$. Now we can write

$$dist(\tilde{p},\ \tilde{q}) = dist(\tilde{p}_1,\ \tilde{q}_j) \leq dist(\tilde{p}_1, \tilde{p}_m) + dist(\tilde{p}_m, \varrho) + dist(\varrho, \tilde{q}_j) \leq$$

$$\leq 2d(\frac{C}{2} - 1) + d + d = Cd.$$

□

Corollary 6.4 *Let G be a (finite or infinite) finitely presented group, and k an integer number greater than 2. Assume that G has at least k elements. Then either*
(1) There exist at least k elements of G with complexity strictly less than $\frac{k}{2}$; or
(2) The number k is even. There is at least one element of complexity $\frac{k}{2}$, and there exist exactly $k-1$ elements of complexity $\leq \frac{k}{2} - 1$. Moreover, in this case G is isomorphic to one of the following groups: $\mathbf{Z}$, $\mathbf{Z}_N$ *for some*

$N \geq k$, $\mathbf{Z}_2 * \mathbf{Z}_2$, or $\mathbf{Z}_2 * \mathbf{Z}_2/\{(ab)^N\}$ for some $N \geq \frac{k}{2}$, where a, b are the non-trivial elements in the two copies of $\mathbf{Z}_2$.

Proof. If not all of the elements of G have complexity $< \frac{k}{2}$, then there exists an element of complexity $\frac{k}{2}$, if k is even, or $\frac{k+1}{2}$, if k is odd. Arguing as in the proof of Proposition 6.2 we see that in the second case G has k elements of complexity $\leq \frac{k-1}{2}$. Assume now that k is even, and there exists an element of complexity $\frac{k}{2}$. Again, arguing as in the proof of Proposition 6.2 we see that there are at least $k-1$ elements of complexity $\leq \frac{k}{2} - 1$. Assume that all of the elements of G of complexity $\leq \frac{k}{2} - 1$ are among these $k-1$ elements constructed in the proof of Proposition 6.2. Then we see that either (a) the element of complexity $\frac{k}{2}$ is a power of a generator a, each other generator of G is equal to a or a^{-1}, and, therefore G is cyclic; or (b) the element of complexity $\frac{k}{2}$ is represented by a word of the form $a^i b \ldots$, where $b \neq a$ or a^{-1}. In case (b) every other generator c must be equal to a, b or their inverses. Furthermore, as there are exactly two elements of complexity one a^{-1} must be equal to a, and b^{-1} to b. Therefore G is isomorphic either $\mathbf{Z}_2 * \mathbf{Z}_2$ or to some its quotient. It is easy to see that if G is a quotient of $\mathbf{Z}_2 * \mathbf{Z}_2$, then this quotient must be isomorphic to $< \mathbf{Z}_2 * \mathbf{Z}_2 | (ab)^N = e >$ for $N \geq \frac{k}{2}$. □

Proof of Theorem 1.4. Let $\tilde{p}$ be a fixed lifting of p to the universal covering $\tilde{M}$ of M. Tile M by connected fundamental domains such that their interiors are the Voronoi cells of points in the inverse image of p under the covering map. These fundamental domains have radius $\leq d$. All of them correspond to different elements of G that act as a group of covering transformations. If $\tilde{p}_i$ corresponds to an element of G of a complexity l, then $dist(\tilde{p}, \tilde{p}_i) \leq 2dl$, and if $\tilde{q}_i$ is the lifting of q that lies in the fundamental domain centered at $\tilde{p}_i$, then $dist(\tilde{p}, \tilde{q}_i) \leq dist(\tilde{p}, \tilde{p}_i) + dist(\tilde{p}_i, \tilde{q}_i) \leq 2ld + d = (2l+1)d$. Assume that there exist k elements of G of complexity $\leq \frac{k-1}{2}$. Then we can connect $\tilde{p}$ with k liftings of q into $\tilde{M}$ at distances $\leq kd$ from $\tilde{p}$ by geodesics. Projecting these geodesics to M we will obtain k distinct geodesics between p and q of length $\leq kd$, which are not even pairwise path homotopic.

Corollary 6.4 implies that it only remains to consider the cases when G is a cyclic group of infinite order or of order $\geq k$, or when G is either $\mathbf{Z}_2 * \mathbf{Z}_2$ or its quotient $< a, b | a^2 = e, b^2 = e, (ab)^N = e >$, $N \geq \frac{k}{2}$.

The proof of Proposition 2 in [NR0] implies the existence of k pairwise non path-homotopic geodesics of length $\leq kd$ connecting p and q in the case when G is $\mathbf{Z}$ or $\mathbf{Z}_N$, $N > k$. Theorem 1.3 implies the desired assertion when $G = \mathbf{Z}_k$ or, when $G =< a, b | a^2 = b^2 = (ab)^{\frac{k}{2}} = e >$. It remains to consider the cases when $G = \mathbf{Z}_2 * \mathbf{Z}_2$ or $\mathbf{Z}_2 * \mathbf{Z}_2/\{(ab)^N\}$, $N > \frac{k}{2}$. This can be done using the method of the proof of Proposition 2 in [NR0] as follows:

To complete the proof in these remaining cases first note that there are exactly $k-1$ elements of G of complexity $\leq \frac{k}{2} - 1$, namely $e, a, b, ab, ba, \ldots, (ab)^{\frac{k}{2}-1}a, (ba)^{\frac{k}{2}-1}b$. If we consider the liftings of q into the corresponding fundamental domains, connect them with $\tilde{p}$ by a minimal geodesics, and project these geodesics back to M, the result will be $k-1$ distinct geodesics in M between p and q of length $\leq (\frac{k}{2}-1)(2d)+d = (k-1)d$. It remains to construct one more geodesic between p and q of length $\leq kd$. We will prove that $\tilde{p}$ and the lifting of q in either the fundamental domain corresponding to $(ab)^{\frac{k}{2}-1}a$ or to $(ba)^{\frac{k}{2}-1}b$ can be connected by a geodesic of length $\leq kd$. Note, that we are immediately guaranteed a geodesic between these points (for either of these two domains) of length $\leq \frac{k}{2}(2d)+d = (k+1)d$, but we want to improve one of these two upper bounds by d. Of course, we will immediately obtain the desired improvement if there exists a geodesic between $\tilde{p}$ and a lifting of q in one of the fundamental domains corresponding to elements of G of complexity between 2 and $\frac{k}{2} - 1$ of length $\leq 2d$. Indeed, in this case we can connect $\tilde{p}$ and the center of this fundamental domain by a path of length $\leq 3d$ (instead of at least $4d$), and at least one of these two upper bounds improves by d, as desired.

Therefore, we assume that the distances from $\tilde{p}$ to all liftings of q to fundamental domains corresponding to elements of G of complexity between 2 and $\frac{k}{2}-1$ are greater than $2d$. Now realize a and b by geodesic loops l_a and l_b of length $\leq 2d$ based at p. Denote the midpoint of l_a by A. Connect A and q by a minimizing geodesic γ_a. Denote two halves of l regarded as paths between p and A by l_{1a}, l_{2a}. Consider paths $l_{1a} * \gamma_a$ and $l_{2a} * \gamma_a$ between p and q. Apply a length non-increasing curve shortening process to both of these paths. At the end we will obtain two geodesics between p and q of length $\leq 2d$. The liftings of these geodesics to $\tilde{M}$ can connect $\tilde{p}$ only with liftings of q in the fundamental domains corresponding to e, a or b, as our assumption implies that they are too short to reach liftings of q to other fundamental domains. As the join of $l_{1a} * \gamma_a$ and $\bar{\gamma}_a * \bar{l}_{2a}$ is a loop homotopic to a, the lifting of one of these two paths connects $\tilde{p}$ with the lifting of q to the fundamental domain centered at $\tilde{p}$ and corresponding to e, and the other, say $l_{2a} * \gamma_a$ connects $\tilde{p}$ with the lifting of q into the fundamental domain corresponding to a. Now consider the path $(l_a * l_b)^{\frac{k}{2}-1} * l_{2a} * \gamma_a$ between p and

q. Its length does not exceed $(2d)(\frac{k}{2}-1)+d+d=kd$. This path lifts to a geodesic between $\tilde{p}$ and the lifting of q corresponding to $(ab)^{\frac{k}{2}-1}a$. Applying to this path a curve-shortening process we will obtain the desired geodesic between $\tilde{p}$ and the lifting of q to the fundamental domain corresponding to $(ab)^{\frac{k}{2}-1}a$ of length $\leq kd$. $\square$

7 Depth of local minima

We will start from the following definition:

Definition 7.1 *Let γ be a path connecting two (not necessarily distinct) points p and q in M^n. Assume that γ is a local minimum of the length functional on $\Omega_{pq}M^n$. Assume, further, that γ is NOT a global minimum of the length functional on the connected component of $\Omega_{pq}M^n$ that contains γ (and all paths path homotopic to γ). For every path homotopy $F:[0,1]\longrightarrow \Omega_{pq}M^n$ between γ and a path of length that is strictly smaller than the length of γ define the level of F as the maximum of lengths of paths $F(t)$ for $t\in[0,1]$. Define the level of γ as the infimum of levels of all path homotopies between γ and a path of a smaller length. Define the depth of γ as the difference between its level and length.*

If γ is a global minimum of the length functional on its connected component of $\Omega_{pq}M^n$, then we say that the level and the depth of γ are infinite.

We are going to present the following generalizations of Theorems 1.5, 1.1, 5.3:

Theorem 7.2 *Let M^n be a closed Riemannian manifold of diameter d. Let p and q be points in M^n, and S a non-negative real number. Let $\gamma(t)$ be a curve of length L connecting points p and q. Assume that there exists an interval $(l,l+2d]$, such that there is no geodesic loop based at p on M^n of length in this interval that provides a local minimum of the length functional on Ω_pM^n of depth $>S$. Then there exists a curve $\tilde{\gamma}(t)$ of length $\leq l+d$ connecting p and q and a path homotopy between γ and $\tilde{\gamma}$ such that the lengths of all curves in this path homotopy do not exceed $L+(S+2d)$.*

Proof:. The proof is essentially the same as the proof of Theorem 1.5 with the following modification: If we get stuck at a geodesic loop of length in the interval $(l,l+2d+\delta]$ which is a local minimum of the length functional on

$\Omega_p M^n$ of depth $\leq S$, we contract this loop to a point or to a geodesic loop of length $\leq l$ by a path homotopy paying the price that the length of curves during the general path homotopy increases by a summand $\leq S + 2d$. $\square$

Using Theorem 7.2 in the proof of Theorem 1.1 instead of Theorem 1.5 we obtain:

Theorem 7.3 *Let M^n be a closed Riemannian manifold of diameter d, p, q, x be points of M^n, $S \geq 0$ a real number. Assume that there exists $k \in N$ such that there is no geodesic of length in the interval $(dist(p,q) + (2k-2)d, dist(p,q) + 2kd]$ joining p and q which is a local minimum of the length functional on $\Omega_{pq}M^n$ of depth $> S$. Then for every positive integer m every map $f : S^m \longrightarrow \Omega_{px}M^n$ is homotopic to a map $\tilde{f} : S^m \longrightarrow \Omega_{px}^{L+o(1)}M^n$, where $L = ((4k+2)m + (2k-3))d + (2m+1)dist(p,q) + (2m-1)S$. Furthermore, every map $f : (D^m, \partial D^m) \longrightarrow (\Omega_{px}M^n, \Omega_{px}^L M^n)$ is homotopic to a map $\tilde{f} : (D^m, \partial D^m) \longrightarrow \Omega_{px}^{L+o(1)}M^n$ relative to ∂D^m. In addition, if for some R the image of f is contained in $\Omega_{px}^R M^n$, then one can choose the homotopy between f and $\tilde{f}$ so that its image is contained in $\Omega_{px}^{R+2mS+(4k+2)md+(2m+1)dist(p,q)+o(1)}M^n$. Also, in this case for every R every map $f : S^0 \longrightarrow \Omega_{px}^R M^n$ is homotopic to a map $\tilde{f} : S^0 \longrightarrow \Omega_{px}^{(2k-1)d+dist(p,q)}$ via a homotopy passing through curves of length $\leq R + S + 2d + o(1)$ connecting p and x.*

As before, $o(1)$ denotes a summand that can be replaced by an arbitrarily small positive number ν. If M^n is an analytic Riemannian manifold, then all these $o(1)$ summands are not necessary. Also, observe that if $x = q = p$, then all the path spaces in Theorem 7.3 become the spaces of loops based at p.

Definition 7.4 *Let M^n be a closed simply-connected Riemannian manifold, and $S_p(M^n)$ denote the maximal depth of a non-trivial local minimum of the length functional on $\Omega_p^{2d}M^n$. (The maximum exists as the set of all loops of length $\leq 2d$ on M^n parametrized by the arclength is compact.) Equivalently, we can define $S_p(M^n)$ as the infimum of S such that each loop λ based at p of length $\leq 2d$ is contractible via a path homotopy passing through loops of length $\leq$ length$(\lambda) + S$. We will call $S_p(M^n)$ the depth of (M^n, p).*

Apply Theorem 7.3 to $S = S_p(M^n)$, $x = q = p$ and $k = 1$. By definition of $S_p(M^n)$ there are no local minima of the length functional on $\Omega_p M^n$ with length in the interval $(0, 2d]$ and depth $> S_p(M^n)$. Therefore,

Theorem 7.5 *Let M^n be a closed simply-connected Riemannian manifold with diameter d, p a point of M^n, S the depth of (M^n, p), and m a positive integer number. Then every map $f : S^m \longrightarrow \Omega_p M^n$ is homotopic to a map $\tilde{f} : S^m \longrightarrow \Omega_p^{(6m-1)d+(2m-1)S+o(1)} M^n$. Furthermore, every map $f : (D^m, \partial D^m) \longrightarrow (\Omega_p M^n, \Omega_p^{(6m-1)d+(2m-1)S} M^n)$ is homotopic to a map $\tilde{f} : (D^m, \partial D^m) \longrightarrow \Omega_p^{(6m-1)d+(2m-1)S+o(1)} M^n$ relative to ∂D^m. In addition, if for some R the image of f is contained in $\Omega_p^R M^n$, then one can choose the homotopy between f and $\tilde{f}$ so that its image is contained in $\Omega_p^{R+2mS+6md+o(1)} M^n$. Also, for every $R > 0$ every map $f : S^0 \longrightarrow \Omega_p^R M^n$ is contractible by a homotopy with the image inside $\Omega_p^{R+S+2d+o(1)} M^n$.*

8 Quantitative Morse theory on loop spaces.

The quantitative Morse theory on loop spaces was initiated in [Gr0] (see also ch. 7 of [Gr]). It studies injectivity and surjectivity properties of homomorphisms in homology induced by the inclusions of sublevel sets of the length functional on a loop space into the loop space. Here is the main result which is a part of Theorem 7.3 in [Gr]:

Theorem 8.1 (M. Gromov) *For every closed simply-connected Riemannian manifold M^n and a point $p \in M^n$ there exists a constant C such that for every positive integer m the inclusion $\Omega_p^{Cm} M^n$ into $\Omega_p M^n$ induces surjective homomorphisms $H_i(\Omega_p^{Cm} M^n) \longrightarrow H_i(\Omega_p M^n)$ for all $i \in \{0, 1, \ldots, m\}$.*

In other words, for every $m \geq 1$ all m-dimensional homology classes of $\Omega_p M^n$ can be realized by cycles "made" out of loops of length $\leq Cm$ based at p. To prove this theorem Gromov demonstrated that there exists C such that for every m there exists an explicit finite dimensional CW-subcomplex $X_m \subset \Omega_p^{Cm} \subset \Omega_p M^n$ such that every map of every m-dimensional CW-complex Y into $\Omega_p M^n$ is homotopic to a map of Y into X_m.

He did not estimate C in his proof. Yet it is easy to see that his proof yields an upper bound for C in terms of the following quantity that we will denote $W_p(M^n)$: This quantity is defined as the infimum of w such that every loop of length $\leq 2d$ based at p can be contracted to p by a path homotopy H such that the length of the trajectory $H(x,t), t \in [0,1]$, of every point $x \in \gamma$ during H does not exceed w.

In other words, denote the infimum of all values of T such that the inclusion homomorphisms $\pi_i(\Omega_p^T M^n) \longrightarrow \pi_i(\Omega_p M^n)$ are surjective for all

$i \in \{0, 1, \dots, m\}$ by $T_{M^n,p}(m)$. (Whether to use the homology or the homotopy groups in this definition is a matter of taste; the resulting values of $T_{M^n,p}(m)$ will be the same.) Then the original proof of Gromov implies that $T_{M^n,p}(m) \le Cm$, and one can use his proof to get an explicit upper bound for C that linearly depends on $W_p(M^n)$. (Gromov also notes that $T_{M^n,p}(m) \ge cm$ for some $c > 0$.)

On the other hand our Theorem 7.5 has the following immediate corollary:

Theorem 8.2 *Let M^n be a closed simply-connected Riemannian manifold of diameter d, p a point of M^n, and S the depth of (M^n, p) (see Definition 7.4). Then*
A. *For every positive integer m the inclusion homomorphisms $\pi_i(\Omega_p^{(6m-1)d+(2m-1)S+o(1)} M^n) \longrightarrow \pi_i(\Omega_p M^n)$ are surjective for all $i \in \{0, 1, \dots, m\}$. Equivalently, for every (arbitrarily small) positive ν every map of a m-dimensional polyhedron X to $\Omega_p M^n$ is homotopic to a map of X into $\Omega_p^{(6m-1)d+(2m-1)S+\nu} M^n$.*
B. *Let m be any positive integer number, $R > 0$ a real number. If a map $f : S^m \longrightarrow \Omega_p^R M^n$ is contractible, then it can be contracted within $\Omega_p^{\max\{R,5d+S\}+2mS+6md+o(1)} M^n \subset \Omega_p^{R+(6m+5)d+(2m+1)S} M^n$.*

Proof:. Part A can be proven by a straightforward application of Theorem 7.3. To prove part B we first apply Theorem 7.3 to homotop f to a (contractible) map $\tilde{f}$ of S^m into $\Omega_p^{(6m-1)d+(2m-1)S+o(1)} M^n$ inside $\Omega_p^{R+2mS+6md+o(1)} M^n$.If $R \le (6m-1)d + (2m-1)S$,we just take $\tilde{f} = f$.

Then we consider a homotopy F that contracts $\tilde{f}$. We regard F as a map of $(D^{m+1}, \partial D^{m+1}) \longrightarrow (\Omega_p M^n, \Omega_p^{(6m-1)d+(2m-1)S+o(1)} M^n)$.Now we again apply Theorem 7.3 to replace F by a homotopy $\tilde{F}$ with the image inside $\Omega_p^{(6(m+1)-1)d+(2(m+1)-1)S+o(1)} M^n = \Omega_p^{(6m+5)d+(2m+1)S+o(1)} M^n$. Now we see that the combination of the homotopies from f to $\tilde{f}$ and the contracting homotopy $\tilde{F}$ takes place in $\Omega_p^{\tilde{R}+o(1)} M^n$, where $\tilde{R} = \max\{R + 2mS + 6md, (6m+5)d + (2m+1)S\} = \max\{R, 5d + S\} + 2mS + 6md$. □

Note, that for every $m \ge 1$ we have $(6m-1)d + (2m-1)S \le 2m(S + 3d)$. Therefore, Theorem 8.2 implies that $T_{M^n,p}(m) \le 2(S_p(M^n) + 3d)m$. Thus, we obtain the following corollary:

Theorem 8.3

$$T_{M^n,p}(m) \le 2(S_p(M^n) + 3d)m.$$

To compare our upper bound for $T_{M^n,p}(m)$ with the upper bound that follows from the original proof of Theorem 8.1 given by Gromov (see ch. 7 of [Gr]) observe, that a trick from [NR] can be used to show that $S_p(M^n) \leq 2W_p(M^n) + 2d$, and, as $W_p(M^n)$ is, obviously, greater than or equal to $\frac{d}{2}$, $2S_p(M^n) + 6d \leq 24W_p(M^n)$. (The inequality $S_p(M^n) \leq 2W_p(M^n) + 2d$ immediately follows from the fact that any homotopy contracting a curve γ of length L to a point p, such that the length of the trajectory of every point does not exceed W, can be converted into a homotopy where the length of curves does not exceed $2W + l$. The idea is very simple: One first moves only a very small interval of γ, so that only its central part reaches p. Then we gradually expand the "tooth". At every stage only a very short interval of γ is being homotoped towards p.) On the other hand, it is not difficult to construct examples that demonstrate that $W_p(M^n)$ cannot be majorized by any function of $S_p(M^n)$. All known upper bounds for $W_p(M^n)$ in terms of $S_p(M^n)$ involve also the injectivity radius of M^n (or the contractibility radius, or, at least, the simply connectedness radius of M^n) - and are also exponential in $\frac{S_p(M^n)}{inj(M^n)}$ (see [NR]). Also, although we did not check the details, the examples constructed in the proof of Theorem 1.2 of [P] seem to demonstrate that $W_p(M^n)$ can, indeed, be exponentially larger than $S_p(M^n)$ even in situations when the simply-connectedness radius is ~ 1. Thus, our upper bound $2S_p(M^n)+6d$ for $\sup_m \frac{T_{M^n,p}(m)}{m}$ in Theorem 8.3 seems to be qualitatively better than an upper bound following from the original proof.

Acknowledgments. This paper was partially written during both authors visit of the Max-Planck Institute at Bonn in Summer 2009. The authors would like to thank the Max-Planck Institute for its kind hospitality. The first author gratefully acknowledges partial support by his NSERC Discovery Grant and NSF Grant DMS-0405954 during his work on the present paper. The second author gratefully acknowledges the partial support by her Natural Sciences and Engineering Research Council (NSERC) University Faculty Award and Discovery Grant and the partial support by the NSF Grant DMS-0604113 during her work on the present paper.

We are very grateful to an anonymous referee for numerous suggestions that greatly helped to improve the exposition.

References

[B] F. Balacheff, Sur des problemes de la géométrie systolique, Séminaire de Théorie Spectrale et Géométrie. Vol. 22. Année 2003-2004, 71-82, Sémin. Théor. Spectr. Géom., 22, Univ. Grenoble I, Saint-Martin-d'Hères, 2004, MR2136136.

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

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

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

[FK] S. Frankel, M. Katz, The Morse landscape of a Riemannian disc, Ann. Inst. Fourier (Grenoble) 43(2)(1993), 503-507.

[Gr] M. Gromov, Metric structures for Riemannian and non-Riemannian spaces, Birkhäuser, Boston, 2007.

[Gr0] M. Gromov, Homotopical effects of dilatation, J. Differential Geom. 13(3)(1978), 303-310.

[Kl] W. Klingenberg, Lectures on closed geodesics, Springer-Verlag, NY, 1978.

[Li] Y. Liokumovich, "Spheres of small diameter with long sweep-outs", Proc. Amer. Math. Soc. 141(2013), 309-312.

[NR] A. Nabutovsky, R. Rotman, Upper bounds on the length of a shortest closed geodesic and quantitative Hurewicz theorem, J. Eur. Math. Soc. (JEMS) 5(3)(2003), 203-244.

[NR0] A. Nabutovsky, R. Rotman, Length of geodesics between two points on a Riemannian manifold, Electron. Res. Announc. Amer. Math. Soc. 13(2007), 13-20.

[NR1] A. Nabutovsky, R. Rotman, The length of the second shortest geodesic, Comment. Math. Helv. 84(2009), 747-755.

[NR2] A. Nabutovsky, R. Rotman, Lengths of geodesics on a two-dimensional sphere, Amer. J. Math. 131(2009), 545-569.

[NR3] A. Nabutovsky, R. Rotman, Short geodesic segments between two points on a closed Riemannian manifold, Geom. Funct. Anal. (GAFA) 19(2009), 498-519.

[NR4] A. Nabutovsky, R. Rotman, "Linear bounds for lengths of geodesic loops on Riemannian 2-spheres", J. of Differential Geom. 89(2011), 217-232.

[NR5] A. Nabutovsky, R. Rotman, "Linear bounds for lengths of geodesics between a pair of points on Riemannian 2-spheres", preprint.

[P] P. Papasoglu, Isodiametric and isoperimetric inequalities for complexes and groups, J. London Math. Soc. 62(2000), 97-106.

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

[R2] R. Rotman, Geodesic loops and periodic geodesics on a Riemannian manifold diffeomorphic to S^3, Math. Z. 257(2)(2007), 427-437.

[S] S. Sabourau, Global and local volume bounds and the shortest geodesic loops, Commun. Anal. Geom. 12 (2004), 1039-1053.

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

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

A. Nabutovsky
Department of Mathematics
40 St. George st.
University of Toronto
Toronto, Ontario M5S 2E4
Canada
e-mail: alex@math.toronto.edu

R. Rotman
Department of Mathematics
40 St. George st.
University of Toronto
Toronto, Ontario M5S 2E4
Canada
rina@math.toronto.edu