LECTURE 14: EXISTENCE OF SHORTEST GEODESICS

By using ODEs (or equivalently, vector fields on tangent or cotangent bundles), we have proved local existence of geodesics as well as the local length-minimizing property. In what follows we turn to global aspects of geodesics.

1. Length minimizing curves on compact Riemannian manifolds

¶ Length minimizing curves are geodesics.

Last time we showed that near any point p, there is a neighborhood U so that for any two points $q_1, q_2 \in U$, there is a unique normal geodesics γ in U that connects q_1 and q_2, and more over, γ is the only shortest curve connecting q_1 and q_2. As a consequence, one gets

Proposition 1.1. *Let p, q be two points on (M, g). If $\gamma : [0, l] \to M$ is a piecewise smooth curve that connects $p = \gamma(0)$ and $q = \gamma(l)$ and is parametrized by arc-length [so $l = \text{Length}(\gamma)$], and if $l = d(p, q)$, then γ is smooth and is a geodesics.*

Proof. First note that by definition of the Riemannian distance function d, there is no piecewise smooth curve connecting p and q of length less than l.

By compactness of $\gamma([0, l])$, there is $\varepsilon > 0$ so that any $t_1, t_2 \in [0, l]$ with $|t_1 - t_2| < \varepsilon$, there is a unique arc-length parametrized length-minimizing curve γ_{t_1,t_2} connecting $\gamma(t_1)$ and $\gamma(t_2)$ which is a normal geodesic. If $\gamma_{t_1,t_2} \neq \gamma|_{[t_1,t_2]}$, then γ_{t_1,t_2} is strictly shorter than $\gamma|_{[t_1,t_2]}$, and we may replace $\gamma|_{[t_1,t_2]}$ by γ_{t_1,t_2} to get a piecewise smooth curve which is shorter than γ, a contradiction. So for any $t_1, t_2 \in [0, l]$ with $|t_1 - t_2| < \varepsilon$, $\gamma|_{[t_1,t_2]}$ is smooth and satisfies the geodesic equation. It follows that γ is smooth and satisfies the geodesic equation on the whole $[0, l]$. □

So the shortest curve between any two points on a connected Riemannian manifold must be a geodesic. Here are two natural subsequent questions:

> **Question 1:** Given $p, q \in M$, does there exist a smooth curve of length $d(p, q)$ between p and q [which then becomes a shortest geodesic]?
> **Question 2:** Given $p, q \in M$, is the length-minimizing curve the only geodesics connecting p and q?

By studying very simple examples, it is quite obvious that the answers to both questions are **NO**. However, as usual a simple "no" is not a satisfied answer. In the next couple lectures we will give more in-depth answer to these two questions. For example, under which condition the answer to **Question 1** is yes? For **Question 2**, how does a geodesic change from length minimizing to non-minimizing?

¶ Length minimizing curves on compact Riemannian manifolds.

For the remaining of this lecture, we focus on **Question 1**, or more generally, on conditions for the existence of various "global shortest geodesics". So let's start with a simple counterexample to **Question 1**:

Example. Let $M = \mathbb{R}^2 \setminus \{0\}$ be the punctured plane, equipped with the standard Euclidean flat metric. Then there is no smooth curve of length 2 connecting the points $(-1, 0)$ and $(1, 0)$.

One can easily find the issue in this example: there are many curves of length $2+\varepsilon$ connecting $(-1, 0)$ and $(1, 0)$, but their "limit curve" does not exist as a curve in M because of the puncture. In fact, this is always the case for those examples that **Question 1** has answer "NO": Given any connected Riemannian manifold (M, g) and any $p, q \in M$, by definition there exist piecewise smooth curves γ_ε of length no more than $d(p, q) + \varepsilon$. By using the technique in the proof of Proposition 1.1 one can even assume these curves to be "piecewise geodesics" [i.e. piecewise smooth with each piece a geodesic]. But these curves cannot converge to a piecewise smooth curve in M: If they converge (in the uniform convergence topology) to a piecewise smooth curve in M, then by the lower semi-continuity of the length functional [c.f. Problem 1 in PSet 1], the limit curve must has length $d(p, q)$ and thus by Proposition 1.1, must be a shortest geodesic connecting p and q.

Now suppose M is compact. As one can imagine, in this case such a sequence will converge, and thus give us a YES to **Question 1**. The key observation is:

Lemma 1.2. *Let $\gamma_i : [0, 1] \to M$ be a family of piecewise smooth curve parametrized with constant speed [i.e. with $|\dot\gamma_i(t)| = \mathrm{Length}(\gamma_i)$ at all smooth points t of γ_i], such that $\mathrm{Length}(\gamma_i) < L$ for some constant L, then the family $\{\gamma_i\}$ is equicontinuous.*

Proof. For any $x_0 \in [0, 1]$ and any $\varepsilon > 0$, if we take $\delta = \frac{\varepsilon}{L}$, then for $|x - x_0| < \delta$,

$$d(\gamma_i(x), \gamma_i(x_0)) = |x - x_0|\mathrm{Length}(\gamma_i) < \varepsilon, \qquad \forall i,$$

and the conclusion follows. □

Now we prove

Theorem 1.3. *If (M, g) is a compact connected Riemannian manifold, then for any p, q on M, there is a geodesic of length $d(p, q)$ connecting p and q.*

Proof. Let γ_i be a sequence of piecewise smooth curves with

$$\gamma_i(0) = p, \qquad \gamma_i(1) = q, \qquad \mathrm{Length}(\gamma_i) < d(p, q) + \frac{1}{i},$$

parametrized with "constant speed". By Lemma 1.2, $\{\gamma_i\}$ is equicontinuous. It is also pointwise precompact since M is a compact metric space. Applying Arzela-Ascoli theorem we know γ_i has a subsequence that converges to a continuous map $\gamma : [0, 1] \to M$. To get a piecewise smooth curve out of γ, we fix $\varepsilon_0 < \mathrm{inj}(M, g)$, and fix N such that $\frac{1}{N} < \frac{\varepsilon_0}{d(p,q)+1}$. Split each γ_i into N pieces, $\gamma_i^j = \gamma_i|_{[\frac{j}{N},\frac{j+1}{N}]}$. Then γ_i^j converges to $\gamma^j = \gamma|_{[\frac{j}{N},\frac{j+1}{N}]}$. Denote $p_j = \gamma(\frac{j}{N})$. Then

$$d(p_j, p_{j+1}) \le \liminf_{i\to\infty} \mathrm{Length}(\gamma_i^j) \le \frac{d(p,q)}{N} < \varepsilon_0.$$

Let $\tilde\gamma$ be the piecewise geodesic obtained by connecting each p_j to p_{j+1} by the shortest geodesic (which exists since $\varepsilon_0 < \mathrm{inj}(M,g)$). Since by definition $p_0 = p, p_N = q$ and

$$\mathrm{Length}(\tilde\gamma) = \sum_{j=0}^{N-1} d(p_j, p_{j+1}) \le d(p,q),$$

we conclude from Proposition 1.1 that $\tilde\gamma$ is the shortest geodesic from p to q (whose length is $d(p,q)$). □

¶ Length minimizing curves in given path-homotopy class.

With a little bit more work, one can find geodesics between p and q that are not absolutely length minimizing, but only "relatively length minimizing":

Theorem 1.4. *Let (M,g) be a compact connected Riemannian manifold, and p, q are two points in M. Then in each path-homotopy class of curves γ with $\gamma(0) = p, \gamma(1) = q$, there is a length-minimizing curve and the curve is a geodesic.*

Proof. Let l_0 be the infimum of length of all piecewise smooth curves in the given path homotopy class, which is positive (at least $d(p,q)$) since $p \ne q$. Again take a sequence of piecewise smooth curves γ_i in the given path homotopy class so that $\mathrm{Length}(\gamma_i) < l_0 + \frac{1}{i}$. By Arzela-Ascoli theorem as above, γ_i has a convergent subsequence whose limit is a continuous curve γ. We take ε_0 small so that each geodesic ball $B(p, 2\varepsilon_0)$ is strongly convex (as in Whitehead theorem). Again we may divide each γ_i into N pieces, and let $p_j = \gamma(\frac{j}{N})$. Then we still have $d(p_j, p_{j+1}) < \varepsilon_0$. As a result, $\gamma|_{[\frac{j}{N},\frac{j+1}{N}]}$ is path-homotopic to the shortest geodesic connecting p_j to p_{j+1}. So if we let $\tilde\gamma$ be the piecewise geodesic obtained by connecting each p_j to p_{j+1} by shortest geodesic, then $\tilde\gamma$ is path homotopic to γ and has shortest length in the given homotopy class. Finally by Proposition 1.1, each sufficiently small part in $\tilde\gamma$ must be geodesic. So $\tilde\gamma$ is a geodesic. □

Remark. It was proved by Serre in 1951 that in any compact Riemannian manifold, there are infinitely many geodesics joining any pair of points. [Note that for the sphere, the geodesics could contain a whole great circle which repeat many times]

¶ Length minimizing closed curves in given free homotopy class.

One may also apply the same argument to the case $p = q$, i.e. i.e. consider closed curves with base point p. There are two issues:

(1) If the homotopy class is trivial, then "the shortest curve" is a single point and thus is not interesting.

(2) In each non-trivial homotopy class of curves with base point p, by the same argument one gets:

Proposition 1.5. *There is a shortest curve $\gamma : [0,1] \to M$ with $\gamma(0) = \gamma(1) = p$ in the given homotopy class with base point p, and it is a geodesic.*

However, in general γ is not smooth at the point p (i.e. $\dot\gamma(0) \neq \dot\gamma(1)$). Such a curve is called a *geodesic loop* with base point p.

Note that although a geodesic loop γ is a closed curve on M, it is not closed if we take an "upstairs" point of view: the integral curve in S^*M that corresponds to γ is not a closed curve in S^*M. In applications those geodesics that are "not only closed on M, but also closed on S^*M" are more important:

Definition 1.6. We say a geodesic $\gamma : [0,1] \to M$ is a *closed geodesic* if $\gamma(0) = \gamma(1)$ and $\dot\gamma(0) = \dot\gamma(1)$.

So closed geodesics are projections of closed integral curves in S^*M to M, and they can also be regarded as smooth maps $\gamma : S^1 \to M$ that satisfies the geodesic equation for all $t \in S^1$. Here are some simple examples:

- Any geodesic on S^m is a closed geodesic.
- On the standard cylinder $S^1 \times \mathbb{R}$, a geodesic is either a closed geodesic ("horizontal circles") or a non-self-intersecting geodesic that "goes to infinity" in both direction. [Similar for the standard torus $S^1 \times S^1$].
- For the one-sheet hyperboloid $x^2 + y^2 - z^2 = 1$, there is a unique closed geodesic (the circle with $z = 0$), and many geodesic loops based at points not on the closed geodesic, as well as many "unbounded geodesics".

As motivated by the last example, to get a closed geodesic one cannot use curves in the "homotopy class with base point p" any more. Instead, one should look at the free homotopy class of closed curves. By adjusting the proofs above, one has

Theorem 1.7. *Let (M, g) be a compact connected Riemannian manifold which is not simply connected. Then in each free homotopy class, there is a length-minimizing curve and the curve is a geodesic.*

Remark. The theorem fails for non-compact connected Riemannian manifold (even if we add completeness assumption).

Remark. For compact simply-connected Riemannian manifold, one can also prove the existence of a closed geodesic: It was proved by Birkhoff for Riemannian 2-spheres (with any Riemannian metric), and later by Lusternik-Fet for any compact simply connected Riemannian manifold. It was further proved by Gromoll-Meyer in 1971 that for simply connected closed manifolds whose cohomology ring $H^*(M; \mathbb{Q})$ is not generated by a single element, there are always infinitely many closed geodesics.