LECTURE 15: COMPLETENESS

1. The Hopf-Rinow Theorem

¶ The Hopf-Rinow Theorem and consequences.

Last time we proved that on any compact Riemannian manifold, any nontrivial path-homotopy class contains a shortest curve and that curve must be a geodesics. Today we will study geodesics in a wider class of Riemannian manifolds, namely, complete Riemannian manifolds, and prove the existence of shortest geodesics in each non-trivial path-homotopy class in such manifolds.

We will first prove the existence of shortest geodesics [i.e. a geodesic of length $d(p,q)$] between any two given points p, q on any complete Riemannian manifold. This is the second part of a well-known theorem proved by Hopf and Rinow in 1931. Recall that a Riemannian manifold (M, g) is called geodesically complete if the maximal defining interval of any geodesic on M is $\mathbb{R}$. On the other hand, any Riemannian manifold (M, g) admits a Riemannian metric structure given by

$$d(p,q) = \inf\{L(\gamma) \mid \gamma \text{ is a piecewise smooth curve connecting } p \text{ to } q\},$$

and thus we can talk about the completeness of d: a metric space is complete if any Cauchy sequence in it converges.

Now we state Hopf-Rinow theorem, which contains two parts: the first part claims that for Riemannian manifolds, the two notions of completeness coincide; while the second part claims the existences of shortest geodesic on such manifolds.

Theorem 1.1 (Hopf-Rinow). *Let (M, g) be a connected Riemannian manifold.*

(Part I) *The following statements are equivalent:*
- *(1) (M, d) is a complete metric space.*
- *(2) (M, g) is geodesically complete.*
- *(3) There exists $p \in M$ so that $\exp_p$ is defined for all $X_p \in T_pM$.*
- *(4)* [Heine-Borel property] *Any bounded closed subset in M is compact.*

(Part II) *Moreover, each of the previous statements implies*
- *(5) for any $p, q \in M$, there exists a geodesic of length $d(p,q)$ connecting p and q.*

Definition 1.2. A connected Riemannian manifold (M, g) satisfying any of (1)-(4) is called a *complete Riemannian manifold*.

Remark. Property (5) is NOT enough to guarantee that (M, g) is complete. For example, the open unit ball $B_1(0)$ in $(\mathbb{R}^m, g_0)$ satisfies (5), but is not complete.

Remark. For a general metric space, condition (1) does NOT imply condition (4): any infinite dimensional Banach or Hilbert space like l^2 is a counterexample. So as metric spaces, Riemannian manifolds are special (and *nice*) metric spaces.

We list a couple immediate consequences of Hopf-Rinow theorem. Since any compact metric space is complete, we get another proof of

Corollary 1.3. *Any compact Riemannian manifold is geodesically complete.*

Since any two points can be connected by a geodesic,

Corollary 1.4. *If (M, g) is complete and connected, then for any $p \in M$, the exponential map $\exp_p : T_pM \to M$ is surjective.*

Since the Heine-Borel property is inherited by closed subsets, we have[warning: although any closed subspace of a complete metric space is complete, one cannot apply (1) here since "the induced metric on a submanifold S in the metric space (M, d)" is not the same as "the Riemannian distance generated by the induced Riemannian metric on $S \subset (M, g)$"]

Corollary 1.5. *Any closed submanifold of a complete Riemannian manifold, when endowed with the induced Riemannian metric, is complete.*

¶ Proof of "Hopf-Rinow Theorem, Part II".

We first prove Part II of Hopf-Rinow theorem. More precisely, we prove $\boxed{(2)\Rightarrow(5)}$, or equivalently, its local version, namely $\boxed{(3)\Rightarrow(5')}$, where

($5'$) for any $q \in M$, there exists a geodesic of length $d(p, q)$ connecting p and q.

Denote $r = d(p, q)$. We have already seen that there exists $0 < \delta < r$ so that the exponential map $\exp_p$ is a diffeomorphism from $B_\delta(0) \in T_pM$ to $B(p, \delta) \in M$. Note that the geodesic sphere $S(p, \delta) = \exp_p(S_\delta(0))$ is compact. Since the distance function is continuous [c.f. lecture 3], there exists $p_0 \in S(p, \delta)$ so that

$$d(p_0, q) = \inf_{p' \in S(p,\delta)} d(p', q).$$

Let γ be the normal geodesic from p to p_0. By (3), γ is defined over $\mathbb{R}$. Let

$$A = \{s \in [\delta, r] \mid d(\gamma(s), q) = r - s\}.$$

We will show $\sup A = r$, which implies $\gamma(r) = q$.

To prove this, we first notice that $\delta \in A$, since

$$r = d(p, q) = \inf_{p' \in S(p,\delta)} (d(p, p') + d(p', q)) = \delta + \inf_{p' \in S(p,\delta)} d(p', q) = \delta + d(\gamma(\delta), q).$$

So A is nonempty.

Secondly, it's easy to see that A is closed, since the function

$$f(s) = d(\gamma(s), q) - r + s$$

is continuous and $A = f^{-1}(0) \cap [\delta, r]$.

Now let $s_0 = \sup A$. Since A is nonempty and closed, $s_0 \in A$. Suppose $s_0 < r$. Then by repeating the previous argument, we know that there exists $0 < \delta' < r - s_0$ and $p_0' \in S(\gamma(s_0), \delta')$ so that

$$d(p_0', q) = \min_{p' \in S(\gamma(s_0), \delta')} d(p', q) = d(\gamma(s_0), q) - \delta'.$$

Since $s_0 \in A$, we get

$$d(p_0', q) = r - s_0 - \delta'.$$

So by the triangle inequality,

$$d(p_0', p) \geq d(p, q) - d(p_0', q) = r - (r - s_0 - \delta') = s_0 + \delta'.$$

On the other hand, the curve $\tilde{\gamma}$ by connecting p to $\gamma(s_0)$ along γ and then connecting $\gamma(s_0)$ to p_0' by the "radial" minimal geodesic has length exactly $s_0 + \delta'$. So $\tilde{\gamma}$, with the arc-length parametrization, must be a geodesic. Obviously $\tilde{\gamma}$ has to coincide with γ. In other words, $p_0' = \gamma(s_0 + \delta')$. As a consequence,

$$d(\gamma(s_0 + \delta'), q) = r - (s_0 + \delta'),$$

i.e. $s_0 + \delta' \in A$. This conflicts with the fact that $s_0 = \sup A$.

¶ Proof of "Hopf-Rinow Theorem, Part I".

Having proved (3) $\Longrightarrow$ (5′), now we prove part I of Hopf-Rinow's theorem by

$$(4) \Longrightarrow (1) \Longrightarrow (2) \Longrightarrow (3) \quad \text{and} \quad (3) + (5') \Longrightarrow (4).$$

(4)⇒(1) This is a standard result in general topology: Let p_i be any Cauchy sequence, then the set $\{p_i\}$ is contained in bounded ball B whose closure is compact by (4). It follows that p_i has a subsequence that converges to some p_0. But p_i is a Cauchy sequence, so the entire sequence $p_i \to p_0$.

(1)⇒(2) Let γ be any normal geodesic on M. By the existence and uniqueness theorem, the maximal defining interval of γ must be an open interval (a, b). If $b < \infty$, then we can take a sequence $s_i \to b-$. In particular, s_i is a Cauchy sequence in $\mathbb{R}$. But γ is a normal geodesic, so

$$d(\gamma(s_i), \gamma(s_j)) \leq |s_i - s_j|.$$

As a consequence, $\gamma(s_i)$ is a Cauchy sequence in (M, d). If follows that there exists a point $p \in M$ so that $\gamma(s_i) \to p$.

Since $\mathcal{E}$ is open and $(p, 0) \in \mathcal{E}$, there exists $\varepsilon > 0$ so that $(q, Y_q) \in \mathcal{E}$ for any q with $d(q, p) < \varepsilon$ and any $Y_q \in T_qM$ with $|Y_q| < 2\varepsilon$. So if we take i large enough so that $b - s_i < \frac{\varepsilon}{2}$ and thus $d(\gamma(s_i), p) < \frac{\varepsilon}{2}$, then $\gamma(t; \gamma(s_i), \varepsilon\dot{\gamma}(s_i))$ is defined for $t \in [0, 1]$. In other words, the geodesic $\gamma_1(t) = \gamma(t; \gamma(s_i), \dot{\gamma}(s_i))$ is well defined for $0 < t < \varepsilon$. Since γ_1 coincides with γ at s_i, they must be the same. In particular, γ can be defined for all $t < s_i + \frac{\varepsilon}{2}$, which exceeds the upper bound b, a contradiction.

Similarly by considering the "reverse geodesic" one also has $a = -\infty$. So any normal geodesic on M, and thus any geodesic on M, has defining interval $\mathbb{R}$.

(2)⇒(3) This is obvious. ((M,g) is geodesically complete $\Leftrightarrow \mathcal{E} = TM$.)

(3)+(5′)⇒(4) Let $K \subset M$ be a bounded closed set. Then there exists a constant $C > 0$ so that $d(p,k) < C$ for all $k \in K$. According to (3) and (5′), $K \subset \exp_p(\overline{B_C(0)})$, where $\overline{B_C(0)}$ is the *closed* ball of radius C in T_pM, which is compact in T_pM. Since $\exp_p$ is smooth, $\exp_p(\overline{B_C(0)})$ is also compact. Thus K, as a closed subset of a compact set, is compact.

2. Geodesics and Riemann covering map

¶ Lifting to the Riemannian covering.

Next let's turn to prove the existence of length minimizing geodesics in any path-homotopy class of curves connecting p to q. The idea is to straightforward: instead of working on piecewise smooth curves in M connecting given points p and q that lies in a given path-homotopy classes, we will move to the universal covering $\pi : \widetilde{M} \to M$ of M and work on piecewise smooth curves starting with a fixed $\widetilde{p} \in \pi^{-1}(p)$ and ends at the point $\widetilde{q} \in \pi^{-1}(q)$ so that any curve connecting $\widetilde{p}$ and $\widetilde{q}$ projects to a curve in the given path-homotopy classes, and then we can apply the second part of Hopf-Rinow theorem.

For the argument mentioned above to work, we need a couple ingredients. First, we need to lift the complete metric g on M to a complete metric on its universal covering $\widetilde{M}$. Recall

- Let M, N be connected smooth manifolds. A smooth map $f : M \to N$ is said to be a *smooth covering map* if
  1. for any $q \in N$, there is a neighborhood V of q in N and open subsets U_α of M so that $f^{-1}(V) = \cup_\alpha U_\alpha$,
  2. for each α, $f : U_\alpha \to V$ is a diffeomorphism,
  3. these U_α's are disjoint.

  As is well known, if $f : M \to N$ is a covering map, then
  - $\dim M = \dim N$ and f is surjective,
  - fix any $p_\alpha \in f^{-1}(q)$, any path (and path homotopy) starts at q in N admits a unique lifting to M that starts at p_α,
  - moreover, if N is simply connected, then f is a global diffeomorphism.
- If (M, g_M) and (N, g_N) are Riemannian manifolds, then a smooth covering map $\pi : M \to N$ is called a *Riemannian covering map* if $\pi^* g_N = g_M$. Note:
  - given any smooth covering map $\pi : M \to N$ and any Riemannian metric on N, one may pullback that metric to M to make the covering map a Riemannian covering. [c.f. PSet 1 Problem 3]
  - any Riemannian covering map is a local isometry.
- We also need some standard properties of local isometries. Let $f : (M,g) \to (N,h)$ be a local isometry, then

- M and N have "the same" Riemannian metrics at corresponding points, and thus the same Levi-Civita connection and the same Riemannian curvature at corresponding points [c.f. PSet 2 Problem 1].
- In particular, f maps geodesics into geodesics, and if f is a Riemannian covering, then the lifting of a geodesic is a geodesic.
- for any piecewise smooth curve γ in M, one has $|\dot\gamma|_{\gamma(t)} = |df_{\gamma(t)}(\dot\gamma)|_{f(\gamma(t))}$ and thus $L(\gamma) = L(f(\gamma))$.

Now we prove

Proposition 2.1. *Let (M,g) be a complete Riemannian manifold, and $\pi : \widetilde{M} \to M$ be a smooth covering map. Then $(\widetilde{M}, \pi^*g)$ is complete.*

Proof. For any $\widetilde{p} \in \widetilde{M}$ and any $\widetilde{v} \in T_{\widetilde{p}}\widetilde{M}$, we denote $p = \pi(\widetilde{p})$ and $v = d\pi_{\widetilde{p}}(\widetilde{v})$. Then by definition of completeness, there is a geodesic $\gamma : \mathbb{R} \to M$ with $\gamma(0) = p$ and $\dot\gamma(0) = v$. By the path-lifting property for covering space, there is a unique lifting $\widetilde{\gamma} : \mathbb{R} \to \widetilde{M}$ with $\widetilde{\gamma}(0) = \widetilde{p}$, which is a geodesic since it is the lifting of a geodesic.. Moreover, since $\pi : (\widetilde{M}, \pi^*g) \to (M, g)$ is a local isometry and since $\pi \circ \widetilde{\gamma} = \gamma$, we get

$$\dot{\widetilde{\gamma}}(0) = (d\pi_{\widetilde{p}})^{-1}(\dot\gamma(0)) = (d\pi_{\widetilde{p}})^{-1}(v) = \widetilde{v}.$$

So the geodesic starts at $\widetilde{p}$ in the direction $\widetilde{v}$ is defined over $\mathbb{R}$. □

¶ Length minimizing curves in given path-homotopy class.

As a consequence, we can extend Theorem 1.4 in Lecture 14 to complete Riemannian manifolds.

Theorem 2.2. *Let (M,g) be a complete 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 piecewise smooth curve and it is a geodesic.*

Proof. Consider the universal covering $\pi : \widetilde{M} \to M$. Equip $\widetilde{M}$ with the covering Riemannian metric π^*g. Given any path $\sigma : [0,1] \to M$ connecting p and q in the given homotopy class, and given any $\widetilde{p} \in \pi^{-1}(p)$, there is a unique lifting $\widetilde{\sigma} : [0,1] \to \widetilde{M}$ of σ with $\widetilde{\sigma}(0) = \widetilde{p}$. Since $(\widetilde{M}, \pi^*g)$ is complete, by Hopf-Rinow theorem, there is a minimizing geodesic $\widetilde{\gamma}$ from $\widetilde{p}$ to $\widetilde{q} := \widetilde{\sigma}(1)$. Since π is a local isometry, the projection $\gamma = \pi \circ \widetilde{\gamma}$ is a geodesic in M with $\gamma(0) = p, \gamma(1) = q$. Since $\widetilde{M}$ is simply connected, $\widetilde{\gamma}$ is path-homotopic to $\widetilde{\sigma}$ and thus γ is path-homotopic to σ.

Finally suppose σ_1 be any piecewise smooth curve in M from p to q in the given path homotopy class, then its lifting $\widetilde{\sigma}_1$ in $\widetilde{M}$ with starting point $\widetilde{\sigma}_1(0) = \widetilde{p}$ must ends at $\widetilde{q}$, and thus by our choice of $\widetilde{\gamma}$,

$$L(\gamma) = \text{Length}(\widetilde{\gamma}) \le \text{Length}(\widetilde{\sigma}_1) = L(\sigma_1).$$

So γ is the shortest curve in the given path homotopy class. □

¶ The theorem of Ambrose.

In Proposition 2.1 we start with a complete Riemannian manifold downstairs and a smooth covering map [topological information], and end with a complete Riemannian structure upstairs so that the map is a local isometry [geometric information]. It turns out that the theorem has an "inverse", i.e. given an upstairs complete Riemannian manifolds and a local isometry f [geometric information], the downstairs metric must be complete and the map is a covering map [topological information]:

Theorem 2.3 (Ambrose)**.** *Let (M, g) and (N, h) be connected Riemannian manifold, and $f : (M, g) \to (N, h)$ a local isometry. Suppose (M, g) is complete, then f is a smooth covering map, and (N, h) is complete.*

Note that "$f : (M, g) \to (N, h)$ a local isometry and (N, h) is complete" is not enough to guarantee f to be a covering map. We give an immediate consequence of Ambrose's theorem, which will be used later in studying structures of Riemannian manifolds of non-positive sectional curvature:

Corollary 2.4. *Let (M, g) be a connected Riemannian manifold, and $p \in M$. If $\exp_p : T_pM \to M$ is a local diffeomorphism everywhere, then $\exp_p$ is a covering map.*

Proof. Note that the condition implies $\exp_p$ is defined on the whole T_pM, and thus by Hopf-Rinow, (M, g) is complete. Endow T_pM with the metric $\bar{g} = (\exp_p)^*g$, then

- $\exp_p : (T_pM, \bar{g}) \to (M, g)$ is a local isometry.
- For any $v \in T_pM$, the curve $\gamma(t) = tv$ is a geodesic in $(T_pM, \bar{g})$ since its image $\exp_p(tv)$ is a geodesic in (M, g). In other words, $\exp_0 : T_0(T_pM) \to T_pM$ is define on the whole $T_0(T_pM)$, and thus by Hopf-Rinow, $(T_pM, \bar{g})$ is complete.

So by Ambrose theorem, $\exp_p$ is a covering map. □

¶ Proof of Ambrose's theorem.

We will prove this in four steps.

Step 1 "Lift" geodesics in N to geodesics in M:

Lemma 2.5. *Under the conditions of the theorem, given any geodesic $\gamma : [a, b] \to N$ and any $p \in f^{-1}(\gamma(a))$, we can "lift" γ to a geodesic $\bar{\gamma} : [a, b] \to M$ so that $\gamma(t) = f(\bar{\gamma}(t))$ and $\bar{\gamma}(a) = p$. Moreover, the lift is unique.*

Proof. Since $df_p : T_pM \to T_{\gamma(a)}N$ is a linear isometry, one can find a unique $X_p \in T_pM$ so that

$$df_p(X_p) = \dot{\gamma}(a).$$

Let $\bar{\gamma} : [a, b] \to M$ be the geodesic in M with (used completeness here)

$$\bar{\gamma}(a) = p, \dot{\bar{\gamma}}(a) = X_p.$$

Then $f \circ \bar{\gamma}$ is a geodesic in N with same initial conditions as γ, and thus $\gamma = f \circ \bar{\gamma}$. The uniqueness of lift also follows directly from the fact that $df_p : T_pM \to T_{\gamma(a)}N$ is a linear isometry. □

Step 2 (N, h) is complete.

Fix any point $p \in N$ which lies in the image of f. For any geodesic γ starting at p, by Lemma 2.5, we can lift γ to a geodesic $\bar{\gamma}$ starting at any $\bar{p} \in f^{-1}(p)$. Since (M, g) is complete, $\bar{\gamma}$ is a geodesic defined for all t. It follows that $\gamma = f \circ \bar{\gamma}$ is a geodesic defined for all t. So by Hopf-Rinow theorem, $(N, \hbar)$ is complete.

Step 3 f is surjective.

Fix any point $p \in N$ which lies in the image of f. For any $q \in N$, since (N, h) is complete, there is a minimizing geodesic γ from p to q. By Lemma 2.5, we can lift γ to a geodesic $\bar{\gamma}$ starting at any $\bar{p} \in f^{-1}(p)$. It follows that $q \in f(\bar{\gamma}) \subset \mathrm{Im}(f)$.

Step 4 Verify covering properties.

Step 4.1 Construct V and U_α: Fix any $q \in N$, we may assume $f^{-1}(q) = \{p_\alpha\}_{\alpha \in I}$. Take δ small enough so that $V = B(q, \delta)$ is a normal geodesic ball. For each α, let

$$U_\alpha = B(p_\alpha, \delta) \subset M.$$

We note that each point in U_α can be connected to p_α through a unique minimizing geodesic of length less than δ: if there exists $p' \in U_\alpha$ that can be connected to p_α by two geodesics γ, γ' of lengths less that δ starting at p_α, then $f(\gamma)$ and $f(\gamma')$ are geodesic in N from q to $f(p')$ of lengths less than δ, and thus we must have $f(\gamma) = f(\gamma') =: \tilde{\gamma}$ and thus $\dot{\gamma}(0) = (df_{p_\alpha})^{-1}(\dot{\tilde{\gamma}}(0) = \dot{\gamma}'(0)$.

Step 4.2 $f^{-1}(V) = \cup_\alpha U_\alpha$: For any $p \in f^{-1}(V)$, let $\gamma : [0, 1] \to N$ be the minimal geodesic in V connecting $f(p)$ to q, and $\bar{\gamma}$ its lift starting at p. Then $f(\bar{\gamma}(1)) = \gamma(1) = q$, so there exists α so that $\bar{\gamma}(1) = p_\alpha$. Note $L(\bar{\gamma}) = L(\gamma) < \delta$, we conclude that $p \in U_\alpha$. So $f^{-1}(V) \subset \cup_\alpha U_\alpha$.

Conversely, for any point $p \in U_\alpha$, there is a minimal geodesic $\bar{\gamma} : [0, 1] \to M$ connecting p_α to p with length $< \delta$. It follows that $\gamma = f \circ \bar{\gamma}$ is a geodesic starting from q with length $< \delta$. So $f(p) = f(\bar{\gamma}(1)) = \gamma(1) \in V$, and thus $U_\alpha \subset f^{-1}(V)$.

Step 4.3 $f : U_\alpha \to V$ is diffeomorphism: Since local isometry maps geodesics into geodesics, and geodesics are determined by initial values, we have

$$\exp_q [df_{p_\alpha}(X_\alpha)] = f(\exp_{p_\alpha}(X_\alpha))$$

for any $X_\alpha \in T_{p_\alpha} M$. Moreover, when restricted to balls of radius δ, both $\exp_q$ and $\exp_{p_\alpha}$ are diffeomorphisms. Since df is an linear isomorphism which is also a diffeomorphism on the whole $T_{p_\alpha} M$, we conclude that

$$f = \exp_q \circ df_{p_\alpha} \circ \exp_{p_\alpha}^{-1}$$

when restricted to balls of radius δ, so in particular $f : U_\alpha \to V$ is a diffeomorphism.

Step 4.4 For $\alpha \neq \beta$, $U_\alpha \cap U_\beta = \emptyset$: Suppose there exists $p \in U_\alpha \cap U_\beta$ and $\alpha \neq \beta$. Let $\bar{\gamma}_\alpha$ and $\bar{\gamma}_\beta$ be the minimal geodesic from p to p_α and p_β respectively. Then $f(\bar{\gamma}_\alpha)$ and $f(\bar{\gamma}_\beta)$ are minimal geodesics in N, both from $f(p)$ to q. It follows that $f(\bar{\gamma}_\alpha) = f(\bar{\gamma}_\beta)$, and both $\bar{\gamma}_\alpha$ and $\bar{\gamma}_\beta$ are lifts of $f(\bar{\gamma}_\alpha)$ from p. By uniqueness of lift, $\bar{\gamma}_\alpha = \bar{\gamma}_\beta$ and thus $p_\alpha = p_\beta$.