LECTURE 13: THE EXPONENTIAL MAP

1. The Exponential Map

Let (M, g) be a Riemannian manifold. Recall that for any $p \in M$ and any $X_p \in T_pM$, there exists a unique geodesic $\gamma(t) = \gamma(t; p, X_p)$ such that

$$\gamma(0) = p, \dot{\gamma}(0) = X_p.$$

Moreover, the geodesic $\gamma(t)$ depends smoothly on p and X_p. We let

$$\mathcal{E} = \{(p, X_p) \mid \gamma(t; p, X_p) \text{ is defined on an interval containing } [0, 1]\}.$$

By definition $\mathcal{E} = TM$ if and only if (M, g) is geodesically complete, i.e. each geodesic can be defined on $\mathbb{R}$.

Note that a linear reparametrization of a geodesic is again a geodesic. So for any geodesic $\gamma(t; p, X_p)$ and any $\lambda > 0$, the curve

$$\tilde{\gamma}(t) = \gamma(\lambda t; p, X_p)$$

is the geodesic with $\tilde{\gamma}(0) = p, \dot{\tilde{\gamma}}(0) = \lambda X_p$. This fact implies

- If $(p, X_p) \in \mathcal{E}$, then for any $0 < \lambda < 1$, $(p, \lambda X_p) \in \mathcal{E}$.
- If $(p, X_p) \notin \mathcal{E}$, then one can find $\varepsilon > 0$ so that $(p, \varepsilon X_p) \in \mathcal{E}$.

On the other hand side, the maximal existence time of geodesics is continuous with respect to p and is lower semi-continuous with respect to X_p. It follows that for each $p \in M$, $\mathcal{E} \cap T_pM$ is star-shaped around $0 \in T_pM$ and contains a disc centered at 0 in T_pM. In particular, $\mathcal{E}$ contains a neighborhood of the zero section M in TM.

Definition 1.1. The *exponential map* is defined to be

$$\exp : \mathcal{E} \to M, \quad (p, X_p) \mapsto \exp_p(X_p) := \gamma(1; p, X_p).$$

By definition the point $\exp_p(X_p)$ is the end point of the geodesic segment that starts at p in the direction of X_p whose length equals $|X_p|$.

Example. For $(\mathbb{R}^n, g_0)$, we can identify each $T_p\mathbb{R}^n$ with $\mathbb{R}^n$. Then

$$\exp_p(X_p) = p + X_p.$$

Example. For $(S^1, d\theta \otimes d\theta)$, we can identify T_eS^1 with $\mathbb{R}^1$. Then

$$\exp_e(X_p) = e^{iX_p}.$$

Remark. Let $M = G$ be a Lie group. If we take the Riemannian metric on G to be the bi-invariant metric, then $\exp_e$ coincides with *the* exponential map

$$\exp : \mathfrak{g} \to G$$

in Lie theory. In particular, if G is a matrix Lie group, then

$$\exp_e(A) = I + A + \frac{A^2}{2!} + \cdots + \frac{A^k}{k!} + \cdots .$$

According to the smooth dependence in ODE theory, the exponential map is smooth. In particular, for each $p \in M$, the map

$$\exp_p : T_pM \cap \mathcal{E} \to M$$

is smooth. By definition $\exp_p$ maps $0 \in T_pM$ to $p \in M$. The following lemma will be very useful:

Lemma 1.2. *For any $p \in M$, if we identify $T_0(T_pM)$ with T_pM, then*

$$(d\exp_p)_0 = Id|_{T_pM} : T_pM \to T_pM.$$

Proof. for any $X_p \in T_0(T_pM) = T_pM$,

$$(d\exp_p)_0(X_p) = \frac{d}{dt}\Big|_{t=0} \exp(tX_p) = \frac{d}{dt}\Big|_{t=0} \gamma(1; p, tX_p) = \frac{d}{dt}\Big|_{t=0} \gamma(t; p, X_p) = X_p.$$

□

As a consequence of the inverse function theorem, we immediately get

Corollary 1.3. *For any $p \in M$, there exists a neighborhood V of 0 in T_pM and a neighborhood U of p in M so that $\exp_p : V \to U$ is a diffeomorphism.*

In general $\exp_p$ is not a global diffeomorphism, even if it may be defined everywhere in T_pM. For example, on S^n, $\exp_p$ is a diffeomorphism from any ball $B_r(0) \subset T_pM$ of radius $r < \pi$ to an open region in S^n, but it fails to be injective for the disk $B_\pi(0)$.

Definition 1.4. For each $p \in M$, the *injectivity radius* of (M, g) at p is

$$\mathrm{inj}_p(M, g) = \sup\{r : \exp_p \text{ is a diffeomorphism on } B_r(0) \subset T_pM\},$$

and the *injectivity radius* of (M, g) is

$$\mathrm{inj}(M, g) = \inf\{\mathrm{inj}_p(M, g) \mid p \in M\}.$$

Example. $\mathrm{inj}(S^n, g_{S^n}) = \pi$.

Remark. If M is compact, then of course

$$0 < \mathrm{inj}(M, g) \le \mathrm{diam}(M, g),$$

where $\mathrm{diam}(M,g)$ is the diameter of (M,g), defined as

$$\mathrm{diam}(M,g) = \sup_{p,q\in M} d(p,q).$$

But for noncompact manifolds M, we may have $\mathrm{inj}(M,g) = 0$ or $+\infty$.

For any $\rho < \mathrm{inj}_p(M,g)$, we have $B_\rho(0) \subset T_pM \cap \mathcal{E}$, where $B_\rho(0)$ is the ball of radius ρ in T_pM centered at 0.

Definition 1.5. We will call $B_\rho(p) = \exp_p(B_\rho(0))$ the *geodesic ball* of radius ρ centered at p in M, and its boundary $S_\rho(p) = \partial B_\rho(p)$ the *geodesic sphere* of radius ρ centered at p in M.

Now let γ be any geodesic starting at p. Then $\exp_p^{-1}(\gamma\cap B_\rho(p)$ is the line segment in $B_\rho(0) \subset T_pM$ starting at 0 in the direction $\dot\gamma$ whose length is $|\gamma \cap B_\rho(p)|$. As a consequence, the geodesics starting at p of lengths less than ρ are exactly the images under $\exp_p$ of line segments starting at 0 in the ball $B_\rho(0)$. So

Corollary 1.6. *Suppose $p \in M$ and $\rho < \mathrm{inj}_p(M,g)$. Then for any $q \in B_\rho(p)$, there is a unique geodesic connecting p to q whose length is less than ρ.*

Remark. No matter how p and q are closed to each other, one might be able to find other geodesics connecting p to q whose length is longer. To see this, one can look at cylinders or torus, in which case one can always find infinitely many geodesics that connecting arbitrary two points p and q.

Another consequence of lemma 1.2 is to construct, for each Jacobi field X along a geodesic γ, a geodesic variation whose variation field is X.

Theorem 1.7. *A vector field X along γ is a Jacobi field if and only if X is the variation field of some geodesic variation of γ.*

Proof. Last time we have seen that the variation field of any geodesic variation is a Jacobi field.

Now suppose X is a Jacobi field along γ. We will denote

$$Y_{\gamma(a)} = \nabla_{\dot\gamma(a)}X.$$

<u>Case 1: $X_{\gamma(a)} \neq 0$.</u> Let $\xi : (-\varepsilon,\varepsilon) \to M$ be a geodesic with initial conditions

$$\xi(0) = \gamma(a), \quad \dot\xi(0) = X_{\gamma(a)}.$$

Let $T(s), W(s)$ be parallel vector fields along ξ with

$$T(0) = \dot\gamma(a) \quad \text{and} \quad W(0) = Y_{\gamma(a)}.$$

Define

$$f : [a,b]\times(-\varepsilon,\varepsilon) \to M, \quad (t,s) \mapsto f(t,s) = \exp_{\xi(s)}((t-a)(T(s)+sW(s))).$$

Then f is a geodesic variation of γ. Let V be the variation field of f. Since both V and X are Jacobi fields along γ, to show $V = X$, it is enough to show $V(a) = X_{\gamma(a)}$ and $\nabla_{\dot\gamma(a)} V = Y_{\gamma(a)}$. The first one follows from

$$V(a) = f_s(a,0) = \frac{d}{ds}\Big|_{s=0} \xi(s) = X_{\gamma(a)},$$

and the second one follows from

$$\nabla_{\dot\gamma(a)} V = \nabla_{f_t} f_s|_{t=a,s=0} = \nabla_{f_s} f_t|_{t=a,s=0} = \nabla_{X_{\gamma(a)}}(T(s) + sW(s)) = W(0) = Y_{\gamma(a)}.$$

Case 2: $X_{\gamma(a)} = 0$. The geodesic variation above becomes

$$f(t,s) = \exp_{\gamma(a)}((t-a)(\dot\gamma(a) + sY_{\gamma(a)})).$$

So

$$V(t) = f_s(t,0) = (d\exp_{\gamma(a)})_{(t-a)\dot\gamma(a)}((t-a)Y_{\gamma(a)}) = (t-a)(d\exp_{\gamma(a)})_{(t-a)\dot\gamma(a)}(Y_{\gamma(a)}).$$

It follows that $V(a) = 0$ and

$$\begin{aligned}\nabla_{\dot\gamma(a)} V &= [\nabla_{\dot\gamma(a)}(t-a)](d\exp_{\gamma(a)})_{(t-a)\dot\gamma(a)}(Y_{\gamma(a)}) + (t-a)\nabla_{\dot\gamma(a)}(d\exp_{\gamma(a)})_{(t-a)\dot\gamma(a)}(Y_{\gamma(a)})\big|_{t=a} \\ &= (d\exp_{\gamma(a)})_0 Y_{\gamma(a)} \\ &= Y_{\gamma(a)}.\end{aligned}$$

□

Remark. The proof for the case $X_{\gamma(a)} \neq 0$ does not apply to the case $X_{\gamma(a)} = 0$. In fact, if $X_{\gamma(a)} \neq 0$, the variation is not a proper variation at $\gamma(a)$. Note that each geodesic is defined on an open interval. So the variation can be think of as a variation

$$f : (a - \varepsilon_0, b + \varepsilon_0) \times (-\varepsilon, \varepsilon) \to M$$

and $\gamma(a)$ is an *inner point* of the image of f. As a result, f_s and f_t are vector fields defined on the two dimensional graph of f near $\gamma(a)$, for which we can apply the commutativity. If $X_{\gamma(a)} = 0$, then the variation is a proper variation at $\gamma(a)$. In this case, although one can extend each geodesic a bit, $\gamma(a)$ is no longer an inner point of the image. As a result, the vectors $\frac{\partial f}{\partial s}$ and $\frac{\partial f}{\partial t}$ are no longer vector fields near $\gamma(a)$ (although they are still vector fields near other points).

2. The Gauss Lemma

Now let $(p, X_p) \in \mathcal{E}$. By definition, $\exp_p$ maps the point $X_p \in T_pM$ to the point $\exp_p(X_p) \in M$. In general, the differential of the $\exp_p$ at X_p is no longer the identity. However, we have

Lemma 2.1 (The Gauss lemma)**.** *Let $(p, X_p) \in \mathcal{E}$. Then for any $Y_p \in T_pM = T_{X_p}(T_pM)$, we have*

$$\langle (d\exp_p)_{X_p} X_p, (d\exp_p)_{X_p} Y_p \rangle_{\exp_p(X_p)} = \langle X_p, Y_p \rangle_p.$$

Proof. By linearity, it's enough to check the lemma for $Y_p = X_p$ and $Y_p \perp X_p$.
Case 1: $Y_p = X_p$. If we denote $\gamma(t) = \exp(tX_p)$, then $X_p = \dot\gamma(0)$ and

$$(d\exp_p)_{X_p} X_p = \frac{d}{dt}\Big|_{t=1} \exp_p(tX_p) = \dot\gamma(1).$$

Since geodesics are always of constant speed, we conclude

$$\langle (d\exp_p)_{X_p} X_p, (d\exp_p)_{X_p} X_p\rangle = \langle \dot\gamma(1), \dot\gamma(1)\rangle = \langle \dot\gamma(0), \dot\gamma(0)\rangle = \langle X_p, X_p\rangle.$$

Case 2: $Y_p \perp X_p$. Let $\gamma_1(s)$ be a curve in the sphere of radius $|X_p|$ in T_pM so that $\gamma_1(0) = X_p$ and $\dot\gamma_1(0) = Y_p$ (Here we used the condition that $Y_p \perp X_p$). Since $(p, X_p) \in \mathcal{E}$, we see that there exists $\varepsilon > 0$ so that

$$(p, t\gamma_1(s)) \in \mathcal{E}$$

for all $0 < t < 1$ and $-\varepsilon < s < \varepsilon$.

Now let $A = \{(t, s) \mid 0 < t < 1, -\varepsilon < s < \varepsilon\}$ and consider the geodesic variation

$$f : A \to M, \quad (t, s) \mapsto \exp_p(t\gamma_1(s)) = \gamma(t; p, \gamma_1(s))$$

of $\gamma(t; p, X_p)$. Then

$$f_t(1, 0) = \frac{d}{dt}\Big|_{t=1} \exp(tX_p) = (d\exp_p)_{X_p} X_p,$$
$$f_s(1, 0) = \frac{d}{ds}\Big|_{s=0} \exp(\gamma_1(s)) = (d\exp_p)_{X_p} Y_p.$$

So

$$\langle (d\exp_p)_{X_p} X_p, (d\exp_p)_{X_p} Y_p\rangle = \langle f_t(1, 0), f_s(1, 0)\rangle.$$

On the other hand side, as in last time we have [c.f. line 2-line 6 in the proof of the first variation formula]

$$\frac{\partial}{\partial t}\langle f_t, f_s\rangle = \langle \nabla_{f_t} f_t, f_s\rangle + \frac{1}{2}\frac{\partial}{\partial s}\langle f_t, f_t\rangle.$$

Use the facts $f_t = \dot\gamma_s$, γ_s are geodesics, and [It is here that we need the fact that each γ_s is a *geodesic*!]

$$|f_t| = |\dot\gamma_s(t)| = |\dot\gamma_s(0)| = |\gamma_1(s)| = 1,$$

we see that $\langle f_t, f_s\rangle$ is independent of t. Since

$$\lim_{t\to 0} f_s(t, 0) = \lim_{t\to 0} \frac{d}{ds}\Big|_{s=0} \exp_p(t\gamma_1(s)) = \lim_{t\to 0} d(\exp_p)_{tX_p}(tY_p) = 0,$$

we conclude $\langle f_t(1, 0), f_s(1, 0)\rangle = 0$, which proves the lemma. □

Geometrically, the Gauss lemma implies

Corollary 2.2 (The Geometric Gauss Lemma)**.** *For any $\rho < \mathrm{inj}_p(M, g)$ and any $q \in S_\rho(p)$, the shortest geodesic connecting p to q is orthogonal to $S_\rho(p)$.*