LECTURE 16: VARIATIONS OF LENGTH AND ENERGY

Although we defined geodesics as "self-parallel" curves, in the last several lectures we have seen that on Riemannian manifolds, geodesics are closely related to "length minimizing" curves:

- (Lecture 13) on any Riemannian manifold, in a small neighborhood of any point, geodesics are precisely the shortest curves connecting endpoints.
- (Lecture 14 and 15) on any complete Riemannian manifold, in each path-homotopy class, there exists a length minimizing curve and it is a geodesic.

On the other hand, we also know the existence of geodesics which are not length minimizing in the given path homotopy class [e.g. closed geodesics on S^m]. In what follows we take a closer look at the relation between geodesics and the length functional.

1. Geodesics as critical points of energy functional

¶ The Euler-Lagrange equation.

For any $p, q \in M$, consider

$$\mathcal{C}_{pq} = \{\gamma : [a, b] \to M \mid \gamma \text{ is piecewise smooth and } \gamma(a) = p, \gamma(b) = q\}.$$

One may ask: what property distinguish geodesics in $\mathcal{C}_{pq}$ from other curves? One of the answers should be "length-minimizing", at least locally. Now let's attack this problem by studying the length functional directly.

Recall that the length of a piecewise smooth curve $\gamma : [a, b] \to (M, g)$ is

$$L(\gamma) = \text{Length}(\gamma) = \int_a^b |\dot{\gamma}(t)| dt.$$

To find the minimum of such a functional, for simplicity let's first assume that γ is inside a coordinate patch, and thus is given by a vector-valued function $x(t) = (x^1(t), \cdots, x^m(t))$. Consider a very general question in variational analysis:

> Given a smooth function $f = f(t, x, \dot{x})$, find all the minimizer of the functional
>
> $$I(x) = \int_a^b f(t, x(t), \dot{x}(t)) dt$$
>
> in the set of all smooth curves $x(t) = (x^1(t), \cdots, x^m(t))$ with fixed endpoints $x(a) = p, x(b) = q$.

Since this "space of smooth curves" is huge (namely, of "infinitely dimensional"), one cannot apply usual methods in calculus to find the minimizer. Fortunately, there is a new branch of mathematics called variational calculus that is invented to handle such problems. The idea is: convert the one "infinitely dimensional" problem [in which we have infinitely many directions to move] to infinitely many "one-dimensional problems" [in which we fix one direction to move]. Here is how it works in this example: Since we are studying the functional on curves with fixed endpoints, we may fix any smooth map $y(t) = (y^1(t), \cdots, y^m(t))$ with $y(a) = y(b) = 0$ and consider the corresponding one-parameter family of curves of the form $x(t)+\varepsilon y(t)$. So if $x = x(t)$ is a minimizer of I, then we must have

$$\begin{aligned} 0 = \frac{d}{d\varepsilon}\Big|_{\varepsilon=0} I(x+\varepsilon y) &= \frac{d}{d\varepsilon}\Big|_{\varepsilon=0} \int_a^b f(t, x+\varepsilon y, \dot{x}+\varepsilon\dot{y})dt \\ &= \int_a^b \left(\frac{\partial f}{\partial x^k}(t,x,\dot{x})y^k + \frac{\partial f}{\partial \dot{x}^k}(t,x,\dot{x})\dot{y}^k\right)dt \\ &= \int_a^b \left(\frac{\partial f}{\partial x^k}(t,x,\dot{x}) - \frac{d}{dt}\frac{\partial f}{\partial \dot{x}^k}(t,x,\dot{x})\right)y^k dt. \end{aligned}$$

As a result, we see that if x is a minimizer (or a critical point) of I, then

$$\frac{\partial f}{\partial x^k}(t,x,\dot{x}) = \frac{d}{dt}\frac{\partial f}{\partial \dot{x}^k}(t,x,\dot{x}), \quad 1 \le k \le m,$$

which is known as the *Euler-Lagrange equation* for the functional I.

¶ Arc length v.s. energy.

We may apply Euler-Lagrange equation above to the function

$$f(t,x(t),\dot{x}(t)) = (\langle \dot{x}(t), \dot{x}(t)\rangle_{x(t)})^{1/2} = (g_{ij}(x(t))\dot{x}^i(t)\dot{x}^j(t))^{1/2}.$$

However, since there is a square root, the computation could be a bit messy. It turn out that there is a small trick that can simplify the computation a lot: instead of the length functional, we can work on the *energy functional*:

$$E(\gamma) = \frac{1}{2}\int_a^b |\dot{\gamma}(t)|^2 dt.$$

By the Cauchy-Schwartz inequality, for each piecewise smooth curve γ,

$$L(\gamma)^2 = \left(\int_a^b |\dot{\gamma}(t)|dt\right)^2 \le \left(\int_a^b 1^2 dt\right)\left(\int_a^b |\dot{\gamma}(t)|^2 dt\right) = 2(b-a)E(\gamma),$$

with equality holds if and only if $|\dot{\gamma}(t)| \equiv$ constant. In particular we see that although the length functional $L(\gamma)$ is independent of the choice of parametrizations, the energy functional does depend on the parametrizations (and on the length of the interval $[a,b]$): among different parametrizations of γ on fixed $[a,b]$, $E(\gamma)$ is minimized on the "constant speed parametrization". As a consequence we can prove

Proposition 1.1. *A curve $\gamma : [a,b] \to M$ in $\mathcal{C}_{pq}$ minimize the energy functional $E(\gamma)$ if and only if it has constant speed and minimize the length functional $L(\gamma)$.*

Proof. Suppose $\gamma : [a,b] \to M$ minimize $E(\gamma)$ but there exists $\gamma' \in \mathcal{C}_{pq}$ such that $L(\gamma') < L(\gamma)$, then for the "constant speed re-parametrization" $\widetilde{\gamma} : [a,b] \to M$ of γ',

$$E(\widetilde{\gamma}) = \frac{1}{2(b-a)} L(\widetilde{\gamma})^2 = \frac{1}{2(b-a)} L(\gamma')^2 < \frac{1}{2(b-a)} L(\gamma)^2 \le E(\gamma'),$$

which is a contradiction. So any minimizer of $E(\gamma)$ must also minimize $L(\gamma)$.

Conversely, if $\gamma : [a,b] \to M$ has constant speed and minimize $L(\gamma)$, but there is another $\gamma' : [a,b] \to M$ in $\mathcal{C}_{pq}$ with $E(\gamma') < E(\gamma)$, then

$$L(\gamma') \le \sqrt{2(b-a)E(\gamma')} < \sqrt{2(b-a)E(\gamma)} = L(\gamma),$$

a contradiction. □

Since any piecewise smooth curve can be reparametrized to have constant speed, to minimize $L(\gamma)$, it is enough to minimize $E(\gamma)$ whose integrand is much simpler. Applying Euler-Lagrange equation to

$$f(t, x(t), \dot{x}(t)) = g_{ij}(x(t))\dot{x}^i(t)\dot{x}^j(t)$$

we get, for $1 \le k \le m$,

$$\frac{\partial g_{ij}}{\partial x^k}\dot{x}^i\dot{x}^j = \frac{d}{dt}\left(g_{kj}\dot{x}^j\right) + \frac{d}{dt}\left(g_{ik}\dot{x}^i\right) = 2\frac{\partial g_{kj}}{\partial x^i}\dot{x}^i\dot{x}^j + 2g_{kj}\ddot{x}^j,$$

which, as we have seen in Lecture 13, is equivalent to the geodesic equation

$$\ddot{x}^k + \Gamma^k_{ij}\dot{x}^i\dot{x}^j = 0, \quad 1 \le k \le m.$$

Amazingly enough, by this way we get not only all the geodesics that are length minimizing curves in C_{pq} and all the geodesics that are the length minimizing curves in each path homotopy class of $\mathcal{C}_{pq}$ [1], but in fact we get ALL the geodesics connecting p and q:

Theorem 1.2. *For a Riemannian manifold (M, g), a curve $\gamma : [a,b] \to M$ in $\mathcal{C}_{pq}$ is a geodesic if and only if it satisfies the Euler-Lagrange equation of the energy functional $E(\gamma)$.*

This gives another proof of the fact that any length minimizing curve is a geodesic, and also explains why there exist geodesics that are not length minimizing even in the given path-homotopy class: those curves are only "critical points" of E which need not be minimizing among all near-by curves. If one need to find the minimizing geodesics, then as usual one can further calculate the second order derivative $\frac{d^2}{d\varepsilon^2}|_{\varepsilon=0}E(x+\varepsilon y)$ in a coordinate system, using which one can show that geodesics are always length-minimizing locally in a neighborhood.

[1]Although these curves are not length minimizing in $\mathcal{C}_{pq}$, they are in fact length minimizing among "nearby curves", namely among curves of the form $x + \varepsilon y$ in the computation above for ε small enough, since these curves are in the same path-homotopy class.

2. Formulas for the first and second variations

The calculations above are thought-provoking but has the shortcoming that they are performed in a chart. In what follows we take a global way to calculate the first and second derivatives, and also study variations which could be more general (without fixing endpoints) or more restrictive (with geodesic variation).

¶ Variations.

For simplicity we start with smooth variations of a smooth curve:

Definition 2.1. Let $\gamma : [a, b] \to M$ be a smooth curve, and $\varepsilon > 0$.

(1) A smooth *variation* of γ is a smooth map $f : [a, b] \times (-\varepsilon, \varepsilon) \to M$ so that

$$f(t, 0) = \gamma(t)$$

for all $t \in [a, b]$. In what follows, we will also denote $\gamma_s(t) = f(t, s)$.

(2) A variation f is called *proper* if for every $s \in (-\varepsilon, \varepsilon)$,

$$\gamma_s(a) = \gamma(a) \qquad \text{and} \qquad \gamma_s(b) = \gamma(b).$$

(3) A variation is called a *geodesic variation* if each γ_s is a geodesic.

For simplicity we denote $R = [a, b] \times (-\varepsilon, \varepsilon)$. Let $f : R \to M$ be a smooth variation of γ. Then $E = f^*TM$ is a vector bundle over R, on which we have an induced linear connection $\widetilde{\nabla} = f^*\nabla$ (where ∇ is the Levi-Civita connection on (M, g)). To be rigorous, in what follows we will calculate via $\widetilde{\nabla}$, and refer to the appendix of this section for the definition and properties of $\widetilde{\nabla}$.

The variation f gives rise to two natural sections of E, namely,

$$f_s(t, s) := (df)_{t,s}(\frac{\partial}{\partial s}) \in T_{f(t,s)}M = E_{t,s}$$

and

$$f_t(t, s) := (df)_{t,s}(\frac{\partial}{\partial t}) \in T_{f(t,s)}M = E_{t,s},$$

where $\frac{\partial}{\partial s}$ and $\frac{\partial}{\partial t}$ are the coordinate vector fields on R. Note that by definition,

$$f_t(t_0, s_0) = \dot{\gamma}_{s_0}(t_0).$$

We are mainly interested in the restriction of the sections f_s and f_t to $s = 0$, which are in fact "vector fields along γ". Obviously we have $\dot{\gamma}(t) = f_t(t, 0) = (df)_{t,0}(\frac{\partial}{\partial t})$.

Definition 2.2. We will call

$$V(t) := f_s(t, 0) = (df)_{t,0}(\frac{\partial}{\partial s})$$

the *variation field* of the variation f.

Note that if γ is an embedded curve and f is an embedding, then both $\dot{\gamma}(t)$ and $V(t)$ can be regarded as vector fields on M along γ in a natural way, and the computations below can be carried out via ∇ instead of $\widetilde{\nabla}$.

¶ The first variation formula of energy for smooth variations.

Now we compute the variation of E along given variation (without fixing endpoints): Let $f(t,s)$ be a smooth variation of a smooth curve $\gamma:[a,b]\to M$. By Proposition 3.11 and Proposition 3.13, the derivative of $E(\gamma_s)$ is

$$\frac{d}{ds}E(\gamma_s)=\frac{1}{2}\int_a^b\frac{\partial}{\partial s}\langle\dot\gamma_s(t),\dot\gamma_s(t)\rangle dt=\int_a^b\langle\widetilde{\nabla}_{\partial/\partial s}f_t,f_t\rangle dt=\int_a^b\langle\widetilde{\nabla}_{\partial/\partial t}f_s,f_t\rangle dt.$$

Applying metric compatibility (i.e. Proposition 3.11) again, we get

$$\int_a^b\langle\widetilde{\nabla}_{\partial/\partial t}f_s,f_t\rangle dt=\int_a^b\frac{\partial}{\partial t}\langle f_s,f_t\rangle dt-\int_a^b\langle f_s,\widetilde{\nabla}_{\partial/\partial t}f_t\rangle dt=\langle f_s,f_t\rangle|_{t=a}^{t=b}-\int_a^b\langle f_s,\widetilde{\nabla}_{\partial/\partial t}f_t\rangle dt.$$

So we get

Theorem 2.3 (The First Variation of Energy). *Given any smooth variation $f(t,s)$ of a smooth curve $\gamma:[a,b]\to M$,*

$$\frac{d}{ds}E(\gamma_s)=\int_a^b\langle\widetilde{\nabla}_{\partial/\partial t}f_s,f_t\rangle dt=\langle f_s(t,s),f_t(t,s)\rangle|_{t=a}^{t=b}-\int_a^b\langle f_s,\widetilde{\nabla}_{\partial/\partial t}f_t\rangle dt.$$

In particular,

$$\left.\frac{d}{ds}\right|_{s=0}E(\gamma_s)=-\int_a^b\left\langle V(t),\nabla_{\dot\gamma(t)}\dot\gamma\right\rangle dt-\langle V(a),\dot\gamma(a)\rangle+\langle V(b),\dot\gamma(b)\rangle.$$

In particular, if f is a proper smooth variation, then

$$\left.\frac{d}{ds}\right|_{s=0}E(\gamma_s)=-\int_a^b\left\langle V(t),\nabla_{\dot\gamma(t)}\dot\gamma\right\rangle dt.$$

Again we see that γ is a geodesic (i.e. $\nabla_{\dot\gamma}\dot\gamma=0$) if and only if γ is a critical point of the energy functional E among all proper variations.

¶ The first variation formula of length for smooth variations.

Use the same way, one can calculate the first variation of the length. A trick to simplify the computation is the following observation:

$$\frac{\partial}{\partial s}|\dot\gamma_s(t)|=\frac{\partial}{\partial s}\langle f_t,f_t\rangle^{\frac{1}{2}}=\frac{1}{2}\frac{1}{|f_t|}\frac{\partial}{\partial s}\langle f_t,f_t\rangle=\frac{1}{|f_t|}\langle\widetilde{\nabla}_{\partial/\partial t}f_s,f_t\rangle=\langle\widetilde{\nabla}_{\partial/\partial t}f_s,\frac{f_t}{|f_t|}\rangle.$$

Then following the same computation, one gets

Theorem 2.4 (The First Variation of Length). *Let $f(t,s)$ be a smooth variation of a smooth curve γ. Then*

$$\left.\frac{d}{ds}\right|_{s=0}L(\gamma_s)=-\int_a^b\left\langle V(t),\nabla_{\dot\gamma(t)}\frac{\dot\gamma}{|\dot\gamma|}\right\rangle dt-\left\langle V(a),\frac{\dot\gamma(a)}{|\dot\gamma(a)|}\right\rangle+\left\langle V(b),\frac{\dot\gamma(b)}{|\dot\gamma(b)|}\right\rangle.$$

As an application we prove

Proposition 2.5. *Let S be a closed submanifold of (M, g). Suppose γ is a geodesic from $p \notin S$ to $q \in S$ with $L(\gamma) = d(p, S)$. Then γ is perpendicular to S.*

Proof. For any $v \in T_qS$, take a curve $\sigma : (-\varepsilon, \varepsilon) \to S$ with $\sigma(0) = q$ and $\dot\sigma(0) = v$. Let γ_s be a variation of γ with $\gamma_s(0) = p$ and $\gamma_s(l) = \sigma(s)$, where $l = L(\gamma)$. Then $V(a) = 0$ and $V(b) = v$, and bu the first variation formula,

$$0 = \frac{d}{ds}|_{s=0} L(\gamma_s) = \langle v, \dot\gamma(l)\rangle,$$

thus the conclusion follows. □

¶ Piecewise smooth curve.

More generally, one can consider piecewise smooth curves $\gamma : [a, b] \to M$, i.e. there exists a subdivision

$$a = t_0 < t_1 < t_2 < \cdots < t_k < t_{k+1} = b$$

such that γ is smooth on each interval $[t_i, t_{i+1}]$. We shall consider *piecewise smooth variations* of γ, which is a continuous map $f : [a, b] \times (-\varepsilon, \varepsilon) \to M$ so that f is smooth on each $[t_i, t_{i+1}] \times (-\varepsilon, \varepsilon)$ for each i. Note that this implies

- for each $s \in (-\varepsilon, \varepsilon)$, the curve $t \mapsto \gamma_s(t) = f(t, s)$ is piecewise smooth,
- for each $t \in [a, b]$, the curve $s \mapsto f(t, s)$ is smooth (so f_s is well-defined at t_i's).

Applying the previous theorems to each $[t_i, t_{i+1}] \times (-\varepsilon, \varepsilon)$, we get

Corollary 2.6. *Let f be a piecewise smooth variation of curve γ. Then*

$$\frac{d}{ds}\bigg|_{s=0} E(\gamma_s) = -\int_a^b \langle V(t), \nabla_{\dot\gamma}\dot\gamma\rangle\, dt - \langle V(a), \dot\gamma(a)\rangle + \langle V(b), \dot\gamma(b)\rangle - \sum_{i=1}^k \langle V(t_i), \dot\gamma(t_i^+) - \dot\gamma(t_i^-)\rangle$$

and

$$\frac{d}{ds}\bigg|_{s=0} L(\gamma_s) = -\int_a^b \left\langle V(t), \nabla_{\dot\gamma}\frac{\dot\gamma(t)}{|\dot\gamma|}\right\rangle dt - \left\langle V(a), \frac{\dot\gamma(a)}{|\dot\gamma(a)|}\right\rangle + \left\langle V(b), \frac{\dot\gamma(b)}{|\dot\gamma(b)|}\right\rangle - \sum_{i=1}^k \left\langle V(t_i), \frac{\dot\gamma(t_i^+)}{|\dot\gamma(t_i^+)|} - \frac{\dot\gamma(t_i^-)}{|\dot\gamma(t_i^-)|}\right\rangle.$$

The local computations above imply that among smooth curves, geodesics are critical points of the energy functional. A natural question is: If a curve is piecewise smooth, can it be a critical point of the energy functional? Of course for γ be a critical point of the energy functional, it must be a geodesic when restricted to any subinterval where it is smooth, or in other words, it must be "piecewise geodesic".

Corollary 2.7. *If a piecewise smooth curve γ is a critical point of the energy functional among proper variations, then it is C^1 and thus a geodesic.*

Proof. We can first choose proper variations with variation fields satisfying $V(t_i) = 0$ and deduce that $\nabla_{\dot\gamma}\dot\gamma = 0$ at any smooth point of γ. In particular, the first term in the right hand of the first variation formula vanishes. As a consequence, we have

$$\sum_{i=1}^{k} \langle V(t_i), \dot\gamma(t_i^+) - \dot\gamma(t_i^-)\rangle = 0$$

for any variation field V. Then for each i we can consider all variation fields so that $V(t_j) = 0$ for all $j \neq i$, and conclude that

$$\langle V(t_i), \dot\gamma(t_i^+) - \dot\gamma(t_i^-)\rangle = 0$$

for any $V(t_i) \in T_{\gamma(t_i)}$. It follows that $\dot\gamma(t_i^+) = \dot\gamma(t_i^-)$ and thus γ is C^1. □

¶ Piecewise smooth curve.

Finally we compute the second variation of energy. As in calculus, the second variation is mainly used near critical points, i.e. near geodesics. So we let $\gamma : [a, b] \to M$ be a geodesic, and $f(t, s)$ be a smooth variation of γ. According to Theorem 2.3, Proposition 3.11 and Proposition 3.13,

$$\begin{aligned}\frac{d^2}{ds^2}E(\gamma_s) &= \int_a^b \frac{\partial}{\partial s}\langle \widetilde\nabla_{\partial/\partial t} f_s, f_t\rangle dt \\ &= \int_a^b \langle \widetilde\nabla_{\partial/\partial s}\widetilde\nabla_{\partial/\partial t} f_s, f_t\rangle dt + \int_a^b \langle \widetilde\nabla_{\partial/\partial t} f_s, \widetilde\nabla_{\partial/\partial s} f_t\rangle dt \\ &= \int_a^b \langle \widetilde R(\frac{\partial}{\partial s}, \frac{\partial}{\partial t}) f_s, f_t\rangle dt + \int_a^b \langle \widetilde\nabla_{\partial/\partial t}\widetilde\nabla_{\partial/\partial s} f_s, f_t\rangle dt + \int_a^b \langle \widetilde\nabla_{\partial/\partial t} f_s, \widetilde\nabla_{\partial/\partial t} f_s\rangle dt.\end{aligned}$$

There are two $\widetilde\nabla_{\partial/\partial t}$ in the above formula. We may either apply Proposition 3.11 to the first one to get

$$\begin{aligned}\frac{d^2}{ds^2}E(\gamma_s) &= \int_a^b \frac{\partial}{\partial t}\langle \widetilde\nabla_{\partial/\partial s} f_s, f_t\rangle dt + \int_a^b \langle \widetilde R(\frac{\partial}{\partial s}, \frac{\partial}{\partial t}) f_s, f_t\rangle - \langle \widetilde\nabla_{\partial/\partial s} f_s, \widetilde\nabla_{\partial/\partial t} f_t\rangle + \langle \widetilde\nabla_{\partial/\partial t} f_s, \widetilde\nabla_{\partial/\partial t} f_s\rangle dt \\ &= \langle \widetilde\nabla_{\partial/\partial s} f_s, f_t\rangle\Big|_{(a,s)}^{(b,s)} + \int_a^b \langle \widetilde R(\frac{\partial}{\partial s}, \frac{\partial}{\partial t}) f_s, f_t\rangle - \langle \widetilde\nabla_{\partial/\partial s} f_s, \widetilde\nabla_{\partial/\partial t} f_t\rangle + \langle \widetilde\nabla_{\partial/\partial t} f_s, \widetilde\nabla_{\partial/\partial t} f_s\rangle dt,\end{aligned}$$

or apply Proposition 3.11 to both to get

$$\begin{aligned}\frac{d^2}{ds^2}E(\gamma_s) &= \int_a^b \frac{\partial}{\partial t}\big(\langle \widetilde\nabla_{\partial/\partial s} f_s, f_t\rangle + \langle f_s, \widetilde\nabla_{\partial/\partial t} f_s\rangle\big) dt \\ &\quad + \int_a^b \big(\langle \widetilde R(\frac{\partial}{\partial s}, \frac{\partial}{\partial t}) f_s, f_t\rangle - \langle \widetilde\nabla_{\partial/\partial s} f_s, \widetilde\nabla_{\partial/\partial t} f_t\rangle - \langle f_s, \widetilde\nabla_{\partial/\partial t}\widetilde\nabla_{\partial/\partial t} f_s\rangle\big) dt \\ &= \big(\langle \widetilde\nabla_{\partial/\partial s} f_s, f_t\rangle + \langle f_s, \widetilde\nabla_{\partial/\partial t} f_s\rangle\big)\Big|_{(a,s)}^{(b,s)} \\ &\quad + \int_a^b \big(\langle \widetilde R(\frac{\partial}{\partial s}, \frac{\partial}{\partial t}) f_s, f_t\rangle - \langle \widetilde\nabla_{\partial/\partial s} f_s, \widetilde\nabla_{\partial/\partial t} f_t\rangle - \langle f_s, \widetilde\nabla_{\partial/\partial t}\widetilde\nabla_{\partial/\partial t} f_s\rangle\big) dt.\end{aligned}$$

Note that by Proposition 3.13(a),

$$\langle \widetilde{R}(\frac{\partial}{\partial s}, \frac{\partial}{\partial t}) f_s, f_t\rangle|_{(t,0)} = \langle R(V(t), \dot\gamma(t))V(t), \dot\gamma(t)\rangle = \langle V, R(\dot\gamma, V)\dot\gamma\rangle(t).$$

So by letting $s=0$ in both formula, we get

Theorem 2.8 (The Second Variation of Energy)**.** *Let $f(t,s)$ be a smooth variation of a geodesic $\gamma : [a,b] \to M$, then*

$$\begin{aligned}\frac{d^2}{ds^2}\Big|_{s=0} E(\gamma_s) &= \langle \widetilde{\nabla}_{\partial/\partial s} f_s, \dot\gamma\rangle|_a^b + \int_a^b \big(\langle R(\dot\gamma, V)\dot\gamma, V\rangle + \langle \nabla_{\dot\gamma} V, \nabla_{\dot\gamma} V\rangle\big) dt \\ &= \big(\langle \widetilde{\nabla}_{\partial/\partial s} f_s, \dot\gamma\rangle + \langle V, \widetilde{\nabla}_{\dot\gamma} V\rangle\big)\big|_a^b - \int_a^b \big\langle V, \nabla_{\dot\gamma}\nabla_{\dot\gamma} V - R(\dot\gamma, V)\dot\gamma(t)\big\rangle dt.\end{aligned}$$

In particular, if the variation is proper, then $V(a) = V(b) = 0$ and

$$(\widetilde{\nabla}_{\partial/\partial s} f_s)|_{(a,s)} = \nabla_{V(a)} f_s(a,s) = 0, \qquad (\widetilde{\nabla}_{\partial/\partial s} f_s)|_{(b,s)} = \nabla_{V(b)} f_s(b,s) = 0$$

so we get

$$\begin{aligned}\frac{d^2}{ds^2}\Big|_{s=0} E(\gamma_s) &= \int_a^b \big(\langle R(\dot\gamma, V)\dot\gamma, V\rangle + \langle \nabla_{\dot\gamma} V, \nabla_{\dot\gamma} V\rangle\big) dt \\ &= \int_a^b \langle V(t), -\nabla_{\dot\gamma}\nabla_{\dot\gamma} V(t) + R(\dot\gamma, V)\dot\gamma(t)\rangle \, dt.\end{aligned}$$

Note that

$$\langle R(\dot\gamma, V)\dot\gamma, V\rangle = -Rm(\dot\gamma, V, \dot\gamma, V) = -K(\dot\gamma, V)\ \mathrm{Area}(\dot\gamma, V),$$

so we see that if (M,g) has non-positive sectional curvature, then for any proper variation of any geodesics,

$$\frac{d^2}{ds^2}\Big|_{s=0} E(\gamma_s) \geq 0.$$

As a result, any geodesic in non-positive sectional curvature space is locally minimizing among nearby curves.

3. Appendix: The induced connection (by Yulong Li)

¶ The pullback bundle.

Let M, N be two smooth manifolds, ∇ a connection on M and $\varphi : N \to M$ a smooth map. Then we may pullback the tangent bundle $\pi : TM \to M$ over M to a vector bundle $\pi' : E = \varphi^*(TM) \to N$ over N[known as the *pullback bundle*], where

$$E = \{(x,\ v) | x \in N,\ v \in TM,\ \varphi(x) = \pi(v)\} \subset N \times TM.$$

In other words, we just set the fiber E_x of $\pi' : E \to N$ to be the vector space $T_{\varphi(x)}M$. Denote by $\widetilde{\varphi} : E \to TM$ the induced bundle map that maps $(x, v) \in E$ to $v \in TM$. Then we have the following commutative diagram

$$\begin{array}{ccccccc}
 & & (x,\ v) & \longmapsto & v & & \\
 & & \cap & & \cap & & \\
(x,\ v) & \in & E & \xrightarrow{\ \widetilde{\varphi}\ } & TM & \ni & v \\
\downarrow & & \downarrow{\scriptstyle \pi'} & & \downarrow{\scriptstyle \pi} & & \downarrow \\
x & \in & N & \xrightarrow[\ \varphi\ ]{} & M & \ni & \pi(v) \\
 & & \cup & & \cup & & \\
 & & x & \longmapsto & \varphi(x) & &
\end{array}$$

In this construction, there are two natural ways to obtain sections on E:

- For any section $V \in \Gamma^\infty(TM)$, one can define an assignment
$$\begin{array}{rccl} \widetilde{\varphi}^*V : & N & \to & E \\ & x & \mapsto & (x,\ V_{\varphi(x)}). \end{array}$$
By this way any smooth vector field $V \in \Gamma^\infty(TM)$ gives rise to a smooth section $\widetilde{\varphi}^*V \in \Gamma^\infty(E)$ of E.
[Note: if E_i is a local frame of TM near $\varphi(x)$, then $\widetilde{\varphi}^*E_i$ is a local frame of E near x.]
- For any section $X \in \Gamma^\infty(TN)$, one can define an assignment
$$\begin{array}{rccl} d\varphi(X) : & N & \to & E \\ & x & \mapsto & (x,\ (d\varphi)_x(X_x)). \end{array}$$
By this way any smooth vector field $X \in \Gamma^\infty(TN)$ gives rise to a smooth section $d\varphi(X) \in \Gamma^\infty(E)$ of E.

The two constructions are related as follows: For $X \in \Gamma^\infty(TN), V \in \Gamma^\infty(TM)$,

$$d\varphi(X) = \widetilde{\varphi}^*V \Longleftrightarrow d\varphi_x(X_x) = V_{\varphi(x)} \Longleftrightarrow X, V \text{ are } \varphi\text{-related}.$$

In manifold theory we have seen that if X, V and Y, W are φ-related, then $[X, Y]$ and $[V, W]$ is φ-related. So we get

Proposition 3.1. *If $d\varphi(X) = \widetilde{\varphi}^*V, d\varphi(Y) = \widetilde{\varphi}^*W$, then $d\varphi([X, Y]) = \widetilde{\varphi}^*([V, W])$.*

To extend this proposition to more general vector fields on N, we let E_i be a local frame of TM near $\varphi(x)$, then $\widetilde{\varphi}^* E_i$ is a local frame of E near x.

Proposition 3.2. *If $X, Y \in \Gamma^\infty(TN)$ and $d\varphi(X) = X^i \widetilde{\varphi}^* E_i, d\varphi(Y) = Y^j \widetilde{\varphi}^* E_j$, then*

$$d\varphi([X,\ Y]) = X(Y^j)\widetilde{\varphi}^* E_j - Y(X^i)\widetilde{\varphi}^* E_i + X^i Y^j \widetilde{\varphi}^*([E_i,\ E_j]).$$

Proof. For any $x \in N$ and any $f \in C^\infty(M)$,

$$Y(\varphi^* f)(x) = (d\varphi_x)(Y_x) f = Y^j(x)(\widetilde{\varphi}^* E_j)_x(f) = Y^j(x)(E_j)_{\varphi(x)}(f) = (Y^j \varphi^*(E_j f))(x),$$

so $Y(\varphi^* f) = Y^j \varphi^*(E_j f)$ and thus

$$Y^j X_x\big(\varphi^*(E_j f)\big) - X^i Y_x\big(\varphi^*(E_i f)\big) = X^i Y^j \varphi^*(E_i E_j f - E_j E_i f) = X^i Y^j \varphi^*([E_i, E_j] f).$$

It follows that as vectors in $T_{\varphi(x)}M$ acting on $f \in C^\infty(M)$,

$$\begin{aligned}(d\varphi[X,Y])_{\varphi(x)} f &= [X,Y]_x(\varphi^* f)\\ &= X_x(Y\varphi^* f) - Y_x(X\varphi^* f)\\ &= X_x\big(Y^j \varphi^*(E_j f)\big) - Y_x\big(X^i \varphi^*(E_i f)\big)\\ &= X_x(Y^j)\varphi^*(E_j f) + Y^j X_x\big(\varphi^*(E_j f)\big) - Y_x(X^i)\varphi^*(E_i f) - X^i Y_x\big(\varphi^*(E_i f)\big)\\ &= X_x(Y^j)\varphi^*(E_j f) - Y_x(X^i)\varphi^*(E_i f) + X^i(x) Y^j(x)\varphi^*([E_i, E_j] f).\end{aligned}$$

So we get, as sections in $\Gamma^\infty(E)$,

$$d\varphi([X,\ Y]) = X(Y^j)\widetilde{\varphi}^* E_j - Y(X^i)\widetilde{\varphi}^* E_i + X^i Y^j \widetilde{\varphi}^*([E_i,\ E_j]).$$

□

¶ The induced connection on the pullback bundle.

Since each fiber $E_x = T_{\varphi(x)}M$, one may transplant structures on TM to E. For example, if (M, g) is a Riemannian manifold, then the metric structure on TM gives rise to a metric structure on the pullback bundle E in the natural way, namely one just endow each fiber $E_x = T_{\varphi(x)}M$ with the inner product $g_{\varphi(x)}$.

Although it is not that obvious, we may also transplant linear connections on TM to E:

Proposition 3.3. *Given any linear connection ∇ on TM, there exists a unique linear connection $\widetilde{\nabla}$ on E satisfying*

$$\widetilde{\nabla}_u(\widetilde{\varphi}^* V) = \widetilde{\varphi}^*(\nabla_{(d\varphi)_x u} V) \tag{1}$$

for any $x \in N$, $u \in T_x N$ and $V \in \Gamma^\infty(TM)$.

Proof. We first prove the uniqueness. Assume $\widetilde{\nabla}$ exists. For any $x_0 \in N$ and any local frame $\{E_i\}_{i=1}^m$ around $\varphi(x_0)$, $\{\widetilde{\varphi}^* E_i\}_{i=1}^m$ is a local frame around x_0. Thus for any $s \in \Gamma^\infty(E)$, we may write

$$s(x) = s^i(x)\widetilde{\varphi}^* E_i(x)$$

for x near x_0. It follows from Leibniz rule and (1) that for any $u \in T_xN$,

$$\widetilde{\nabla}_u s = u(s^i)\widetilde{\varphi}^* E_i(x) + s^i(x)\widetilde{\varphi}^*(\nabla_{(d\varphi)_x u} E_i). \tag{2}$$

So $\widetilde{\nabla}$ is determined by ∇ uniquely.

As for the existence, it is sufficient to show (2) is independent of the choice of the local frame $\{E_i\}$ [and thus gives rise to a map $\widetilde{\nabla} : \Gamma^\infty(TN) \times \Gamma^\infty(E) \to \Gamma^\infty(E)$ via linearity in $\Gamma^\infty(TN)$] and defines a linear connection on E. Let $E_i' = f_i^j E_j$ be another frame field around $\varphi(x)$. Write $s = \bar{s}^i \widetilde{\varphi}^* E_i'$, then $s^i = \bar{s}^j(\varphi^* f_j^i)$ and thus

$$\begin{aligned} u(s^i)\widetilde{\varphi}^* E_i + s^i \widetilde{\varphi}^* \nabla_{(d\varphi)_x u} E_i &= u(\bar{s}^j)(\varphi^* f_j^i)\widetilde{\varphi}^*(E_i) + \bar{s}^j u(\varphi^* f_j^i)\widetilde{\varphi}^*(E_i) + \bar{s}^j(\varphi^* f_j^i)\varphi^*\nabla_{(d\varphi)_x u} E_i \\ &= u(\bar{s}^j)\widetilde{\varphi}^* E_j' + \bar{s}^j \widetilde{\varphi}^*(\nabla_{(d\varphi)_x u} E_j'). \end{aligned}$$

Thus (2) is independent of the choices of the frame field. The verification of linearity and Leibniz's rule is straightforward. □

Definition 3.4. For any smooth map $\varphi : N \to M$ and any linear connection ∇ on M, the unique linear connection $\widetilde{\nabla}$ on $E = \varphi^*(TM)$ defined above is called the *induced connection* of ∇ by φ on E.

To compute $\widetilde{\nabla}_X s$, one usually fix a local frame E_i of TM near $\varphi(x)$, and expend both $d\varphi(X)$ and s in the induced local frame $\widetilde{\varphi}^* E_i$. Then we may rewrite (2) as

Lemma 3.5. *Fix a local frame E_i of TM near $\varphi(x)$. Suppose $X \in \Gamma^\infty(TN)$ satisfies $d\varphi(X) = X^i \widetilde{\varphi}^* E_i$, then for any section $s = s^k \widetilde{\varphi}^* E_k$,*

$$\widetilde{\nabla}_X s = X(s^k)\widetilde{\varphi}^* E_k + X^i s^k \widetilde{\varphi}^*(\nabla_{E_i} E_k).$$

To get a better understanding, we point out two extreme cases:

- If φ is a diffeomorphism, then E is isomorphic to TN, and $\widetilde{\nabla}$ is obtained by "transplanting everything from M to N via φ in the obvious way".
- If φ is a constant map: Suppose $\varphi(x) \equiv y$ for all $x \in N$, then E is the trivial bundle $N \times T_yM$. In this case any $s \in \Gamma^\infty(E)$ can be written as $s = s^i(x)e_i$, where e_i is a basis of T_yM, and
$$\widetilde{\nabla}_X s = (Xs^i)e_i$$
(which is independent of ∇).

¶ Basic properties of the induced connection.

Let $\varphi : N \to M$ be smooth, $E = \varphi^*(TM)$ the pullback bundle, ∇ a linear connection on M, and $\widetilde{\nabla}$ the induced linear connection on E. It is not surprising that the induced connection $\widetilde{\nabla}$ inherits many nice properties from ∇.

First we study the relation between the curvature tensor

$$\widetilde{R}(X,Y)s := \widetilde{\nabla}_X\widetilde{\nabla}_Y s - \widetilde{\nabla}_Y\widetilde{\nabla}_X s - \widetilde{\nabla}_{[X,Y]} s$$

of the induced connection and the curvature tensor of the original linear connection:

Proposition 3.6. *Let ∇ be a linear connection on M, $\widetilde{\nabla}$ its induced connection on E. then for any $X,\ Y \in \Gamma^\infty(TN)$ and any $s \in \Gamma^\infty(E)$, we have*

$$\big(\widetilde{R}(X,\ Y)s\big)|_x = R\big((d\varphi)_x X_x,\ (d\varphi)_x Y_x\big)\widetilde{\varphi}\big(s|_x\big), \quad \forall x \in N. \tag{3}$$

Proof. For simplicity we take a local frame E_i of TM near $\varphi(x)$ such that $[E_i, E_j] = 0$ [e.g. take the coordinate vector fields]. We can simplify further by assuming $[X, Y]_x = 0$ since both sides of (3) depends only on X_x and Y_x, Again write $d\varphi(X) = X^i\widetilde{\varphi}^*E_i$, $d\varphi(Y) = Y^j\widetilde{\varphi}^*E_j$ and $s = s^k\widetilde{\varphi}^*E_k$. Then by Lemma 3.5,

$$\begin{aligned}\widetilde{\nabla}_X\widetilde{\nabla}_Y s =& \widetilde{\nabla}_X\left(Y(s^k)\widetilde{\varphi}^*E_k + Y^j s^k\widetilde{\varphi}^*(\nabla_{E_j}E_k)\right)\\ =& X(Y(s^k))\widetilde{\varphi}^*E_k + X^iY(s^k)\widetilde{\varphi}^*(\nabla_{E_i}E_k)\\ &+ X(Y^js^k)\widetilde{\varphi}^*(\nabla_{E_j}E_k) + X^iY^js^k\widetilde{\varphi}^*(\nabla_{E_i}\nabla_{E_j}E_k).\end{aligned}$$

Evaluating at the point x, and using the facts $[X, Y]_x = 0$ (which, together with the fact $[E_i, E_j] = 0$ and Proposition 3.2 implies $X_x(Y^j) = Y_x(X^j)$) we get

$$\begin{aligned}\big(\widetilde{R}(X,\ Y)s\big)|_x = \big(\widetilde{\nabla}_X\widetilde{\nabla}_Y s - \widetilde{\nabla}_Y\widetilde{\nabla}_X s\big)|_x &= X^iY^js^k\widetilde{\varphi}^*(\nabla_{E_i}\nabla_{E_j}E_k - \nabla_{E_j}\nabla_{E_i}E_k)|_x\\ &= R\big((d\varphi)_xX_x,\ (d\varphi)_xY_x\big)\widetilde{\varphi}\big(s|_x\big).\end{aligned}$$

□

Next let's turn to torsion-freeness. Although it makes no sense to talk about torsion-freeness for a connection on a general vector bundle, since we can't exchange X and s in the expression $\nabla_X s$. However, as we have seen, every smooth vector field $Y \in \Gamma^\infty(TN)$ gives rise to a smooth section $d\varphi(Y) \in \Gamma^\infty(E)$, so we could restrict ourselves to such sections and thus talk about "partial-torsion-freeness" for the induced linear connection:

Proposition 3.7. *If ∇ is torsion free, then for any $X,\ Y \in \Gamma^\infty(TN)$, we have*

$$\widetilde{\nabla}_X d\varphi(Y) - \widetilde{\nabla}_Y d\varphi(X) - d\varphi([X,\ Y]) = 0.$$

Proof. Fix a local frame E_i of TM near $\varphi(x)$, then $\widetilde{\varphi}^*E_i$ is a local frame of E. Write $d\varphi(X) = X^i\widetilde{\varphi}^*E_i$ and $d\varphi(Y) = Y^i\widetilde{\varphi}^*E_i$ near x. Then by Lemma 3.5,

$$\widetilde{\nabla}_X d\varphi(Y) = X(Y^j)\widetilde{\varphi}^*E_j + X^iY^j\widetilde{\varphi}^*(\nabla_{E_i}E_j),$$

which implies

$$\begin{aligned}\widetilde{\nabla}_X d\varphi(Y) - \widetilde{\nabla}_Y d\varphi(X) =& X(Y^j)\widetilde{\varphi}^*E_j + X^iY^j\widetilde{\varphi}^*(\nabla_{E_i}E_j) - Y(X^i)\widetilde{\varphi}^*E_i - X^iY^j\widetilde{\varphi}^*(\nabla_{E_j}E_i)\\ =& X(Y^j)\widetilde{\varphi}^*E_j - Y(X^i)\widetilde{\varphi}^*E_i + X^iY^j\widetilde{\varphi}^*([E_i,\ E_j]).\end{aligned}$$

Now the conclusion follows from Proposition 3.2, □

Finally the metric compatibility. As we have mentioned, any Riemannian metric on M induces a metric structure on the pullback bundle E.

Proposition 3.8. *If g is a Riemannian metric on M, and ∇ is compatible with g, then the induced connection $\widetilde{\nabla}$ is compatible with the induced metric on E, i.e.*

$$X\langle s_1,\ s_2\rangle = \langle \widetilde{\nabla}_X s_1,\ s_2\rangle + \langle s_1,\ \widetilde{\nabla}_X s_2\rangle$$

for any $X \in \Gamma^\infty(TN)$ and any $s_1, s_2 \in \Gamma^\infty(E)$.

Proof. It is enough to prove this property at a point x. Take a local orthonormal frame E_i of TM around $\varphi(x)$. Write $s_j = s_j^i E_i$ for $j = 1, 2$ and denote $u = X_x$, then

$$\begin{aligned}\langle \widetilde{\nabla}_u s_1,\ s_2\rangle + \langle s_1,\ \widetilde{\nabla}_u s_2\rangle =& u(s_1^i)s_2^j(x)\delta_{ij} + s_1^i(x)s_2^j(x)\langle \nabla_{(d\varphi)_x u}E_i,\ E_j\rangle \\ &+ s_1^i(x)u(s_2^j)\delta_{ij} + s_1^i(x)s_2^j(x)\langle E_i,\ \nabla_{(d\varphi)_x u}E_j\rangle \\ =& \sum_{i=1}^m u(s_1^i s_2^i) = u\langle s_1,\ s_2\rangle.\end{aligned}$$

□

¶ The use of the induced connection.

Why should we study the induced connection? Because it provides us the correct language to perform and explain computations computations when we are handling vector fields associated to maps. Here are two applications. Let M be a smooth manifold with a linear connection ∇.

(1) Let $\gamma : [a, b] \to M$ be a smooth curve. We defined the concept of geodesic via $\nabla_{\dot\gamma}\dot\gamma = 0$. However, in this definition we vaguely assumed that γ is an embedded curve, otherwise $\dot\gamma$ need not be " a vector field along γ". On the other hand, from the differential equations of a geodesic, obviously we allow geodesics to be non-embedded curves (like the constant geodesics or geodesics with self-intersections). The correct way is to explain the expression $\nabla_{\dot\gamma}\dot\gamma$ via the induced connection in the case $\gamma : [a, b] \to M$ is not an embedding. We will use $\frac{d}{dt}$ to represent the canonical coordinate vector field on $[a, b] \subset \mathbb{R}$. We first extend the concept of parallel vector fields to sections of $\gamma^*(TM)$ which are not necessary vector fields on M:

Definition 3.9. Let $\gamma : [a, b] \to M$ be a smooth curve.
(a) We say $X : [a, b] \to TM$ is a *smooth vector field along* γ if X is smooth and $X(t) \in T_{\gamma(t)}M$ for all t. In other words, if X is a smooth section of $E = \gamma^*(TM)$.
(b) Let X be a smooth vector field along γ.We say X is *parallel* along γ if

$$\widetilde{\nabla}_{\frac{d}{dt}}\widetilde{\gamma}^* X = 0, \quad \forall t.$$

Note that $\dot\gamma$ is always a smooth vector field along γ. So we may define

Definition 3.10. We say a smooth map $\gamma : [a, b] \to M$ is a *geodesic* if $\dot\gamma$ is parallel along γ in the sense of Definition 3.9, i.e.,

$$\widetilde{\nabla}_{\frac{d}{dt}} \tilde{\gamma}^* \dot\gamma = 0, \quad \forall t.$$

Of course if γ is an embedded curve, then these definitions reduce to the old definitions that we are familiar with.

Applying Proposition 3.8 to this setting, we get

Proposition 3.11. *If ∇ is a metric-compatible linear connection, then for any smooth vector fields V, W along γ,*

$$\frac{d}{dt}\langle V, W\rangle = \langle \widetilde{\nabla}_{d/dt} V, W\rangle + \langle V, \widetilde{\nabla}_{d/dt} W\rangle.$$

(2) Now we turn to variations of a curve. Let

$$f(t, s) : R = [a, b] \times (-\varepsilon,\ \varepsilon) \to M$$

be a smooth map which is a variation of a smooth curve γ. Then

$$f_s(t, s) = df_{t,s}(\frac{\partial}{\partial s}) \in T_{f(t,s)}M = E_{t,s}$$

and

$$f_t(t, s) = df_{t,s}(\frac{\partial}{\partial t}) \in T_{f(t,s)}M = E_{t,s}$$

are sections of the induced bundle $E = f^*TM$ over R. By definition we have $f(t, 0) = \gamma(t)$ and

$$f_t(t, 0) = \dot\gamma.$$

Definition 3.12. We will call

$$V(t) = f_s(t, 0)$$

the *variation field* of f along γ. [It is a section on $\widetilde{\gamma}^*(TM)$.]

Applying Proposition 3.6, Proposition 3.7 to this setting, we get

Proposition 3.13. *Let $f(t, s)$ be any smooth variation. Then*

(a) for any (t_0, s_0),

$$\big(\widetilde{R}(\frac{\partial}{\partial t},\ \frac{\partial}{\partial s}) f_t\big)|_{(t_0,s_0)} = R(f_t(t_0, s_0),\ f_s(t_0, s_0)) f_t(t_0, s_0).$$

In particular,

$$\big(\widetilde{R}(\frac{\partial}{\partial t},\ \frac{\partial}{\partial s}) f_t\big)|_{(t,0)} = R(\dot\gamma(t), V(t))\dot\gamma(t).$$

(b) If ∇ is torsion free, then

$$\widetilde{\nabla}_{\frac{\partial}{\partial s}} f_t = \widetilde{\nabla}_{\frac{\partial}{\partial t}} f_s.$$

In what follows we will mainly use our old notations ∇, in the understanding that if one expression makes no sense, one should explain it in the language of induced connection $\widetilde{\nabla}$.