1. Covariant derivative and parallel transport

In this section all manifolds we consider are without boundary. All connections will be assumed to be Levi-Civita connections of a given metric. We will denote all time derivatives with a dot, $\frac{df}{dt} = \dot{f}$.

Proposition/Definition 1.1. Let M be manifold with a Riemannian metric. Let $c : (a, b) \to M$ be a smooth map from an interval. A *vector field along* c is a smooth map $V : (a, b) \to TM$ so that $V(t) \in TM_{c(t)}$. There is a unique operation $\frac{D}{dt}$ satisfying:

- $\frac{D}{dt}(V(t) + W(t)) = \frac{D}{dt}V(t) + \frac{D}{dt}W(t)$.
- $\frac{D}{dt}(f(t)V(t)) = \frac{D}{dt}V(t) + \dot{f}(t)V(t)$ for any smooth function $f : (a, b) \to \mathbb{R}$.
- Whenever $V(t) = X(c(t))$ for a vector field $X : M \to TM$ on M, $\frac{D}{dt}V(t) = \nabla_{\dot{c}(t)}X$ (notice that this is a local statement, and holds whenever $\dot{c}(t) \neq 0$).

We call this operation $\frac{D}{dt}$ the *covariant derivative.*

Proof: Choose a coordinate chart, then we can write $c(t) = (c_1(t), \ldots, c_n(t))$ and $V(t) = \sum_{i=1}^{n} v_i(t)\partial_i$. Suppose that such an operation exists. Then using the first two properties, we see $\frac{D}{dt}V(t) = \sum_i \dot{v}_i(t)\partial_i + v_i(t)\frac{D}{dt}\partial_i$. Using the final property, we get that $\frac{D}{dt}\partial_i = \sum_j \dot{c}_j(t)\nabla_{\partial_j}\partial_i = \sum_{j,k} \dot{c}_j(t)\Gamma^k_{ij}\partial_k$. By reindexing and putting these together, we get

$$\frac{D}{dt}V(t) = \sum_k \left(\dot{v}_k(t) + \sum_{i,j} v_i(t)\dot{c}_j(t)\Gamma^k_{ij} \right) \partial_k. \tag{1.1}$$

Since we get a formula this shows that $\frac{D}{dt}$ is unique, it is also easy to check that the above formula satisfies the three properties. □

Notice that $\frac{D}{dt}V(t)$ may be non-zero even at points where $\dot{c} = 0$. For example, if c is a constant map, then $\frac{D}{dt}$ is just the ordinary derivative of $V(t)$, thought of as a curve in the vector space $TM_{c(0)}$. Note also that $\frac{D}{dt}$ is Leibnitzian over $\langle \cdot, \cdot \rangle$, i.e. $\frac{d}{dt}\langle V(t), W(t)\rangle = \langle \frac{D}{dt}V(t), W(t)\rangle + \langle V(t), \frac{D}{dt}W(t)\rangle$.

Notice that, if the curve c is given, then equation 1.1 is a first order ordinary differential equation in v_k. Therefore, given any $V(0) \in TM_{c(0)}$, there is a unique vector field $V(t)$ along c satisfying $\frac{D}{dt}V(t) = 0$. We call this vector field the *parallel transport of* $V(0)$ *along* c. For any t, parallel transport is an isomorphism $TM_{c(0)} \to TM_{c(t)}$. Indeed, since $\frac{d}{dt}\langle V(t), V(t)\rangle = 2\langle \frac{D}{dt}V(t), V(t)\rangle = 0$, we see that the length of $V(t)$ does not change with parallel transport.

We emphasize that the parallel transport of a vector depends on the curve c, not just its homotopy class! This dependence is the essence of curvature.

Definition 1.2. A curve $c : (a, b) \to M$ is called a *geodesic* if $\frac{D}{dt}(\dot{c}(t)) = 0$.

Said differently, a curve is geodesic if and only if its velocity vector is parallel. Looking at equation 1.1, we see that in local coordinates c is a geodesic if and only if

$$\ddot{c}_k(t) + \sum_{i,j} \dot{c}_i(t)\dot{c}_j(t)\Gamma^k_{ij} = 0 \quad \text{for all } k = 1, \ldots, n.$$

This is a second order ordinary differential equation where the leading term is homogeneous, therefore we also have existence and uniqueness of solutions matching given initial conditions.

Proposition 1.3. *Given $q \in M$ and $v \in TM_q$, there is a unique geodesic $c : (-\varepsilon, \varepsilon) \to M$ satisfying $c(0) = q$, $\dot{c}(0) = v$.*

Let $q, p \in M$, and $v \in TM_q$. If there is a geodesic $c : (-\varepsilon, 1+\varepsilon)$ satisfying $c(0) = q$, $\dot{c}(0) = v$, and $c(1) = p$, we write $p = \exp_q(v)$. By uniqueness of geodesics the functional notation is justified. $\exp_q$ is not necessarily defined on all of TM_q, since the geodesic with initial data q, v may not be defined at $t = 1$. However, it is a smooth map wherever it is defined, since the existence/uniqueness theorem for ODE's also gives smooth dependence on initial conditions.

We claim that $\exp_q$ is defined on an open ball in TM_q around the origin. Indeed, choose a unit vector $v \in TM_q$. Then we get a geodesic $c : (-\varepsilon, \varepsilon)$ satisfying $c(0) = q$ and $\dot{c}(0) = v$. Then the curve $\widetilde{c} : (-2, 2) \to M$ given by $\widetilde{c}(t) = c(\frac{t}{2\varepsilon})$ is also a geodesic, satisfying $\dot{\widetilde{c}}(0) = \frac{1}{2\varepsilon} v$. Since the unit sphere is compact we can choose a uniform ε which works for all v. Summing up:

Definition 1.4. For every point $q \in M$ there is a ε so that $\exp_q : B(\varepsilon) \to M$ is a well-defined smooth map (here $B(\varepsilon) \subseteq TM_q$ is the unit ball centered at the origin). We call $\exp_q$ the *exponential map* at the point q.

2. Arclength and geodesics

Definition 2.1. Let M be a manifold with a Riemannian metric, and let $c : [a, b] \to M$ be a curve. We define the *arclength* of c to be $\ell(c) = \int_a^b ||\dot{c}(t)|| dt$. Given two points $p, q \in M$, we define the *distance* between p and q to be the infimum of arclengths among all paths connecting p to q. We denote this distance by $\rho(p, q)$.

Let $v \in TM_q$, and define $c(t) = \exp_q(tv)$ as a curve in M. By the time-scaling property of geodesics used above, c is the geodesic through q with initial velocity v. Differentiating with respect to t tells us that $d(\exp_q)_0 = \mathrm{id}$. Thus, if we take a small ball $B(\varepsilon) \subseteq TM_q$, we know that $\exp_q : B(\varepsilon) \to M$ is a diffeomorphism onto its image.

Proposition 2.2. *Let $q \in M$ and choose $\varepsilon > 0$ so that $\exp_q : B(\varepsilon) \to M$ is a diffeomorphism onto its image. Then for any $p = \exp_q(v)$, $\rho(p, q) = ||v||$. Furthermore, the geodesic $c(t) = \exp_q(tv)$ is the unique curve satisfying $\rho(p, q) = \ell(c)$, up to reparametrization.*

This proposition has a number of consequences. First of all it implies that ρ is a metric in the point-set topology sense: it is obviously symmetric and it satisfies the triangle inequality because a path from p to q to r is in particular a path from p to r. To show it is non-degenerate, choose a neighborhood $\exp_q(B(\varepsilon))$ around q. No point $p \in \exp_q(B(\varepsilon))$ can satisfy $\rho(p, q) = 0$ except for $p = q$, since $|| \cdot ||$ is nondegenerate. But any curve to a point outside $\exp_q(B(\varepsilon))$ must have arclength at least ε, since it passes through $\partial \exp_q(B(\varepsilon))$.

The same argument shows that the metric topology induced by ρ agrees with the manifold topology on M: small metric balls around q are the same as the images under $\exp_q$ of vector space balls in TM_q, but since $\exp_q$ is a diffeomorphism on small scales these are also a basis for the manifold topology.

Also, notice that any curve which minimizes arclength must minimize arclength locally. Since being a geodesic is a local property of curves, the proposition implies that arclength minimizers are geodesic:

Corollary 2.3. *Let $c : [a, b] \to M$ be a path satisfying $\ell(c) = \rho(c(a), c(b))$. Then, after reparametrizing, c is a geodesic.*

Proof: Suppose c minimizes arclength, and let $q = c(t)$ be a point in the image of c. Choose a chart $\exp_q(B(\varepsilon)$ around q. Then, for any δ, $c|_[t, t+\delta]$ is minimizes the arclength between q and $c(t+\delta)$ (if there were a shorter path between q and $c(t+\delta)$, we could define a shorter path between $c(a)$ and $c(b)$ just by changing c on this subinterval). By Proposition 2.2 $c|_[t, t+\delta]$ is a reparametrization of a geodesic. But this holds for every point q in the image of c, so c is a reparametrized geodesic everywhere. □

To prove Proposition 2.2, we first prove a Lemma.

Lemma 2.4. *Let $\exp_q : B(\varepsilon) \to M$ be a diffeomorphism onto its image, and let $S_\delta = \exp_q(S^{n-1}(\delta))$ be the image of the sphere of radius δ under the exponential map ($\delta < \varepsilon$). Then the geodesics through q are orthogonal to S_δ.*

Proof: Any path in S_δ can be written as $\exp_q(\delta v(t))$, where $v : (-a, a) \to TM_q$ is a curve satisfying $||v(t)|| = 1$. Let $f : (\varepsilon, \varepsilon) \times (-a, a) \to M$ be defined by $f(r, t) = \exp_q(rv(t))$. We need to show $\langle \frac{\partial f}{\partial r}, \frac{\partial f}{\partial t} \rangle = 0$. Notice that $\frac{\partial f}{\partial t} = 0$ when $r = 0$, therefore to show that $\langle \frac{\partial f}{\partial r}, \frac{\partial f}{\partial t} \rangle = 0$ it suffices to show that $\frac{\partial}{\partial r} \langle \frac{\partial f}{\partial r}, \frac{\partial f}{\partial t} \rangle = 0$.

Using compatibility with the metric, we see that

$$\frac{\partial}{\partial r} \left\langle \frac{\partial f}{\partial r}, \frac{\partial f}{\partial t} \right\rangle = \left\langle \frac{D}{dr} \frac{\partial f}{\partial r}, \frac{\partial f}{\partial t} \right\rangle + \left\langle \frac{\partial f}{\partial r}, \frac{D}{dr} \frac{\partial f}{\partial t} \right\rangle.$$

$\frac{D}{dr} \frac{\partial f}{\partial r} = 0$ everywhere, because for fixed t, $r \mapsto f(r, t)$ is a geodesic. $\frac{D}{dr} \frac{\partial f}{\partial t} = \frac{D}{dt} \frac{\partial f}{\partial r}$, since the connection is symmetric. Therefore

$$\frac{\partial}{\partial r} \left\langle \frac{\partial f}{\partial r}, \frac{\partial f}{\partial t} \right\rangle = \left\langle \frac{\partial f}{\partial r}, \frac{D}{dt} \frac{\partial f}{\partial r} \right\rangle = \frac{1}{2} \frac{\partial}{\partial t} \left\langle \frac{\partial f}{\partial r}, \frac{\partial f}{\partial r} \right\rangle = \frac{1}{2} \frac{\partial}{\partial t} ||v(t)||^2 = 0.$$

This completes the proof. □

Proof of Proposition 2.2: Let $c : [a, b] \to \exp_q(B(\varepsilon))$ be a curve from q to p. We can find functions $r : [a, b] \to [0, \varepsilon]$ and $v : [a, b] \to S^{n-1}(1) \subseteq TM_q$ so that $c(t) = \exp_q(r(t)v(t))$ which means $c(t) = f(r(t), t)$ in the notation of the above proof. Then $\dot{c}(t) = \frac{\partial f}{\partial t} + \dot{r}(t) \frac{\partial f}{\partial r}$, so

$$||\dot{c}(t)||^2 = \left|\left| \frac{\partial f}{\partial t} \right|\right|^2 + |\dot{r}|^2 \left|\left| \frac{\partial f}{\partial r} \right|\right|^2 = \left|\left| \frac{\partial f}{\partial t} \right|\right|^2 + |\dot{r}|^2$$

since $\left|\left| \frac{\partial f}{\partial r} \right|\right| = ||v|| = 1$. In particular, we see that $||\dot{c}|| \geqslant |\dot{r}| \left|\left| \frac{\partial f}{\partial r} \right|\right|$, and equality if and only if $\dot{v} = 0$, which means that the image of c is contained in the geodesic $t \mapsto \exp_q(tv)$. Then

$$\ell(c) = \int_a^b ||\dot{c}|| dt \geqslant \int_a^b |\dot{r}| \, dt \geqslant \int_a^b \dot{r} dt = r(b).$$

On the other hand, c is a reparametrization of the geodesic $t \mapsto \exp_q(tv)$ if and only if it is of the form $c(t) = \exp_q(r(t)v)$ for some increasing function r. This is precisely the conditions that make the above inequalities into equalities. □

3. Geodesics and the second fundamental form

Suppose that $c : (a, b) \to M$ is a geodesic. Since $||\dot{c}||$ is constant, we see that, unless c is a constant map, it is locally embedded.

If $c : (a, b) \to Q$ is any locally embedded curve and $Q \subseteq M$, then we can choose $V : M \to TM$ to be a vector field satisfying $V(c(t)) = \dot{c}(t)$, and therefore

$$\frac{D^M}{dt}\dot{c}(t) = \nabla^M_{\dot{c}(t)} V = \nabla^Q_{\dot{c}(t)} V + \mathrm{II}(V, \dot{c}(t)) = \frac{D^Q}{dt}\dot{c}(t) + \mathrm{II}(\dot{c}(t), \dot{c}(t)).$$

This formula has a number of nice interpretations relating geodesics to the second fundamental form.

- Let Q be a submanifold of M. Then $\mathrm{II} = 0$ everywhere if and only if every geodesic in Q is also a geodesic in M.
- In particular, let c be a locally embedded curve, and let Q be the image of c. Then c is geodesic if and only if it has constant speed and $\mathrm{II} = 0$ everywhere.
- On the other hand, let $q \in Q$ and let $v \in TQ_q$. Let $c : (-\varepsilon, \varepsilon) \to Q$ be the geodesic satisfying the conditions $c(0) = q$ and $\dot{c}(0) = v$. Then $\mathrm{II}(v, v) = \frac{D^M}{dt}\dot{c}(t)$. Since $\mathrm{II}(v, w) = \frac{1}{2}\left(\mathrm{II}(v + w, v + w) - \mathrm{II}(v, v) - \mathrm{II}(w, w)\right)$, this shows we can completely calculate II just by calculating the covariant derivative in M of geodesics in Q.

4. Geodesics and completeness

Theorem 4.1. *Let M be a connected manifold with a Riemannian metric. The following are equivalent:*

(i) The metric ρ is complete.
(ii) At every point $q \in M$, $\exp_q$ can be defined on all of TM_q.
(iii) There is a point $q \in M$ so that $\exp_q$ can be defined on all of TM_q.
(iv) For every point $q \in M$, $\exp_q : TM_q \to M$ is surjective.
(v) There is a point $q \in M$ so that $\exp_q : TM_q \to M$ is surjective.

Proof: Some of the implications are trivial: clearly *(ii)* implies *(iii)*, and *(iv)* implies *(v)*.

We show that *(iii)* implies *(v)*. Let $R = \rho(q, p)$, and choose an $\varepsilon > 0$ so that $\exp_q|_{B(\varepsilon)}$ is a diffeomorphism onto its image. Let $\exp_q(\varepsilon v)$ be the point in $S_\varepsilon = \exp_q(\partial B(\varepsilon))$ which is closest to p, $\rho(x, p) = \rho(S_\varepsilon, p)$, note that x exists since S_ε is compact. We claim that $\exp_q(Rv) = p$.

Consider the set of real numbers $A = \{t \in [0, R]; \rho(\exp_q(tv), p) = R - t\}$, we need to show that $R \in A$. Notice that A is a closed set, and let $r = \sup A \in A$. Suppose for the sake of contradiction that $r < R$, so $x = \exp_q(rv) \neq p$. Choose $\delta > 0$ so that $\exp_x : B(\delta) \to M$ is a diffeomorphism onto its image, and so that $\delta < R - r$. Again there is a point $y = \exp_x(\delta w)$ with $||w|| = 1$ which minimizes the distance between p and $\exp_x(\partial B(\delta))$.

Since every curve between x and p must pass through $\exp_x(\partial B(\delta))$, it follows that $R - r = \rho(x, p) = \delta + \rho(y, p)$. By the triangle inequality, $\rho(q, y) \leqslant \rho(q, x) + \rho(x, y) = r + \delta$. But also by the triangle inequality, $R = \rho(q, p) \leqslant \rho(q, y) + \rho(y, p) = \rho(q, y) + R - r - \delta$. Therefore $\rho(q, y) = r + \delta$.

The piecewise geodesic which travels from q to x for time r and then from x to y for time δ has exactly this length. Thus Corollary 2.3 says that this piecewise geodesic is actually a geodesic, so $y = \exp_q((r+\delta)v)$. But then by definition, $r+\delta \in A$. This contradicts the definition of r, and completes the proof that *(iii)* implies *(v)*. The proof that *(ii)* implies *(iv)* is simply by applying this proof to every point $q \in M$.

Next we show that *(v)* implies *(i)*. Let $K \subseteq M$ be any closed set with diameter $r < \infty$, to show that ρ is complete it suffices to show that K is compact. Let $R = \rho(q, K) + r$, then by the triangle inequality every point $k \in K$ satisfies $\rho(q,k) \leqslant R$. Therefore, since $\exp_q$ is surjective, we must have that $K \subseteq \exp_q(B(R))$, where $B(R) \subseteq TM_q$ is the closed ball of radius R. But then K is a closed subset of a compact space, and therefore it is compact.

Finally, it remains to show that *(i)* implies *(ii)*, we prove the contrapositive. Let $c : (a,b) \to M$ be a geodesic with $b < \infty$ which cannot be extended further. Then the sequence $x_n = c(b - \frac{1}{n})$ cannot converge, because if it did we could define $c(b) = \lim x_n$ and then extend $c : (a, b+\varepsilon)$ by applying the existence theorem for geodesics at the point $c(b)$. But for any geodesic, $\rho(c(t), c(s)) = |t-s|$. Therefore x_n is a Cauchy sequence. □

Often times a Riemannian metric which satisfies *(ii)* (and therefore all five properties) is called *geodesically complete*, or simply *complete*. Notice that every metric on a compact manifold is complete. More generally, if Q is a submanifold of M and M has a complete metric, then the restriction of the metric to Q is complete if and only if Q is properly embedded. In particular, every manifold has a complete metric, since every manifold has a proper embedding into $\mathbb{R}^N$.

Typically when studying global differential geometry on open manifolds we restrict ourselves to complete metrics. The reason being that incomplete metrics are "too easy" to construct by cutting and pasting. For example, the only open surfaces with complete flat metrics are the standard $\mathbb{R}^2$, the cylinder $S^1 \times \mathbb{R}$, and the Mobius strip $S^1 \widetilde{\times} \mathbb{R}$. However, every open surface has an incomplete flat metric.