Chapter 13

Geodesics on Riemannian Manifolds

13.1 Geodesics, Local Existence and Uniqueness

If (M, g) is a Riemannian manifold, then the concept of *length* makes sense for any piecewise smooth (in fact, C^1) curve on M.

Then, it possible to define the structure of a metric space on M, where $d(p, q)$ is the greatest lower bound of the length of all curves joining p and q.

Curves on M which locally yield the shortest distance between two points are of great interest. These curves called *geodesics* play an important role and the goal of this chapter is to study some of their properties.

Given any $p \in M$, for every $v \in T_pM$, the *(Riemannian) norm* of v, denoted $\|v\|$, is defined by

$$\|v\| = \sqrt{g_p(v, v)}.$$

The Riemannian inner product, $g_p(u, v)$, of two tangent vectors, $u, v \in T_pM$, will also be denoted by $\langle u, v \rangle_p$, or simply $\langle u, v \rangle$.

Definition 13.1. Given any Riemannian manifold, M, a *smooth parametric curve* (for short, *curve*) on M is a map, $\gamma\colon I \to M$, where I is some open interval of $\mathbb{R}$. For a closed interval, $[a, b] \subseteq \mathbb{R}$, a map $\gamma\colon [a, b] \to M$ is a *smooth curve from $p = \gamma(a)$ to $q = \gamma(b)$* iff γ can be extended to a smooth curve $\widetilde{\gamma}\colon (a - \epsilon, b + \epsilon) \to M$, for some $\epsilon > 0$. Given any two points, $p, q \in M$, a continuous map, $\gamma\colon [a, b] \to M$, is a *piecewise smooth curve from p to q* iff

(1) There is a sequence $a = t_0 < t_1 < \cdots < t_{k-1} < t_k = b$ of numbers, $t_i \in \mathbb{R}$, so that each map, $\gamma_i = \gamma \restriction [t_i, t_{i+1}]$, called a *curve segment* is a smooth curve, for $i = 0, \ldots, k-1$.

(2) $\gamma(a) = p$ and $\gamma(b) = q$.

The set of all piecewise smooth curves from p to q is denoted by $\Omega(M; p, q)$ or briefly by $\Omega(p, q)$ (or even by Ω, when p and q are understood).

The set $\Omega(M; p, q)$ is an important object sometimes called the *path space* of M (from p to q).

Unfortunately it is an infinite-dimensional manifold, which makes it hard to investigate its properties.

Observe that at any junction point, $\gamma_{i-1}(t_i) = \gamma_i(t_i)$, there may be a jump in the velocity vector of γ.

We let $\gamma'((t_i)_+) = \gamma_i'(t_i)$ and $\gamma'((t_i)_-) = \gamma_{i-1}'(t_i)$.

Given any curve, $\gamma \in \Omega(M; p, q)$, the *length*, $L(\gamma)$, of γ is defined by

$$\begin{aligned} L(\gamma) &= \sum_{i=0}^{k-1} \int_{t_i}^{t_{i+1}} \|\gamma'(t)\| \, dt \\ &= \sum_{i=0}^{k-1} \int_{t_i}^{t_{i+1}} \sqrt{g(\gamma'(t), \gamma'(t))} \, dt. \end{aligned}$$

It is easy to see that $L(\gamma)$ is unchanged by a monotone reparametrization (that is, a map $h\colon [a, b] \to [c, d]$, whose derivative, h', has a constant sign).

Let us now assume that our Riemannian manifold, (M, g), is equipped with the Levi-Civita connection and thus, for every curve, γ, on M, let $\frac{D}{dt}$ be the associated covariant derivative along γ, also denoted $\nabla_{\gamma'}$

Definition 13.2. Let (M, g) be a Riemannian manifold. A curve, $\gamma\colon I \to M$, (where $I \subseteq \mathbb{R}$ is any interval) is a *geodesic* iff $\gamma'(t)$ is parallel along γ, that is, iff

$$\frac{D\gamma'}{dt} = \nabla_{\gamma'}\gamma' = 0.$$

If M was embedded in $\mathbb{R}^d$, a geodesic would be a curve, γ, such that the acceleration vector, $\gamma'' = \frac{D\gamma'}{dt}$, is normal to $T_{\gamma(t)}M$.

By Proposition 12.11, $\|\gamma'(t)\| = \sqrt{g(\gamma'(t), \gamma'(t))}$ is constant, say $\|\gamma'(t)\| = c$.

If we define the *arc-length* function, $s(t)$, relative to a, where a is any chosen point in I, by

$$s(t) = \int_a^t \sqrt{g(\gamma'(t), \gamma'(t))}\, dt = c(t-a), \qquad t \in I,$$

we conclude that for a geodesic, $\gamma(t)$, the parameter, t, is an affine function of the arc-length.

The geodesics in $\mathbb{R}^n$ are the straight lines parametrized by constant velocity.

The geodesics of the 2-sphere are the great circles, parametrized by arc-length.

The geodesics of the Poincaré half-plane are the lines $x = a$ and the half-circles centered on the x-axis.

The geodesics of an ellipsoid are quite fascinating. They can be completely characterized and they are parametrized by elliptic functions (see Hilbert and Cohn-Vossen [26], Chapter 4, Section and Berger and Gostiaux [6], Section 10.4.9.5).

If M is a submanifold of $\mathbb{R}^n$, geodesics are curves whose acceleration vector, $\gamma'' = (D\gamma')/dt$ is normal to M (that is, for every $p \in M$, γ'' is normal to T_pM).

In a local chart, (U, φ), since a geodesic is characterized by the fact that its velocity vector field, $\gamma'(t)$, along γ is parallel, by Proposition 12.5, it is the solution of the following system of second-order ODE's in the unknowns, u_k:

$$\frac{d^2u_k}{dt^2} + \sum_{ij} \Gamma^k_{ij} \frac{du_i}{dt} \frac{du_j}{dt} = 0, \qquad k = 1, \ldots, n,$$

with $u_i = pr_i \circ \varphi \circ \gamma$ ($n = \dim(M)$).

The standard existence and uniqueness results for ODE's can be used to prove the following proposition:

Proposition 13.1. *Let (M, g) be a Riemannian manifold. For every point, $p \in M$, and every tangent vector, $v \in T_pM$, there is some interval, $(-\eta, \eta)$, and a unique geodesic,*

$$\gamma_v \colon (-\eta, \eta) \to M,$$

satisfying the conditions

$$\gamma_v(0) = p, \qquad \gamma_v'(0) = v.$$

The following proposition is used to prove that every geodesic is contained in a unique maximal geodesic (*i.e*, with largest possible domain):

Proposition 13.2. *For any two geodesics,* $\gamma_1 \colon I_1 \to M$ *and* $\gamma_2 \colon I_2 \to M$*, if* $\gamma_1(a) = \gamma_2(a)$ *and* $\gamma_1'(a) = \gamma_2'(a)$*, for some* $a \in I_1 \cap I_2$*, then* $\gamma_1 = \gamma_2$ *on* $I_1 \cap I_2$*.*

Propositions 13.1 and 13.2 imply that for every $p \in M$ and every $v \in T_pM$, there is a unique geodesic, denoted γ_v, such that $\gamma(0) = p$, $\gamma'(0) = v$, and the domain of γ is the largest possible, that is, cannot be extended.

We call γ_v a *maximal geodesic* (with initial conditions $\gamma_v(0) = p$ and $\gamma_v'(0) = v$).

Observe that the system of differential equations satisfied by geodesics has the following homogeneity property: If $t \mapsto \gamma(t)$ is a solution of the above system, then for every constant, c, the curve $t \mapsto \gamma(ct)$ is also a solution of the system.

We can use this fact together with standard existence and uniqueness results for ODE's to prove the proposition below.

Proposition 13.3. *Let (M, g) be a Riemannian manifold. For every point, $p_0 \in M$, there is an open subset, $U \subseteq M$, with $p_0 \in U$, and some $\epsilon > 0$, so that: For every $p \in U$ and every tangent vector, $v \in T_pM$, with $\|v\| < \epsilon$, there is a unique geodesic,*

$$\gamma_v \colon (-2, 2) \to M,$$

satisfying the conditions

$$\gamma_v(0) = p, \qquad \gamma_v'(0) = v.$$

If $\gamma_v \colon (-\eta, \eta) \to M$ is a geodesic with initial conditions $\gamma_v(0) = p$ and $\gamma_v'(0) = v \neq 0$, for any constant, $c \neq 0$, the curve, $t \mapsto \gamma_v(ct)$, is a geodesic defined on $(-\eta/c, \eta/c)$ (or $(\eta/c, -\eta/c)$ if $c < 0$) such that $\gamma'(0) = cv$. Thus,

$$\gamma_v(ct) = \gamma_{cv}(t), \qquad ct \in (-\eta, \eta).$$

Besides the notion of the gradient of a function, there is also the notion of Hessian.

Now that we have geodesics at our disposal, we also have a method to compute the Hessian, a task which is generally quite complex.

Given a smooth function $f\colon M \to \mathbb{R}$ on a Riemannian manifold M, recall that the *gradient* $\mathrm{grad}\, f$ of f is the vector field uniquely defined by the condition

$$\langle (\mathrm{grad}\, f)_p, u\rangle_p = df_p(u) = u(f), \qquad \text{for all } u \in T_pM \text{ and all } p \in M.$$

Definition 13.3. The *Hessian* $\mathrm{Hess}(f)$ (or $\nabla^2(f)$) of a function $f \in C^\infty(M)$ is defined by

$$\begin{aligned}\mathrm{Hess}(f)(X,Y) &= X(Y(f)) - (\nabla_X Y)(f)\\ &= X(df(Y)) - df(\nabla_X Y),\end{aligned}$$

for all vector fields $X, Y \in \mathfrak{X}(M)$.

Since ∇ is torsion-free, we get

$$\begin{aligned}\mathrm{Hess}(f)(X,Y) &= X(Y(f)) - (\nabla_X Y)(f)\\ &= Y(X(f)) - (\nabla_Y X)(f)\\ &= \mathrm{Hess}(f)(Y,X),\end{aligned}$$

which means that the Hessian is *symmetric*.

Proposition 13.4. *The Hessian is given by the equation*

$$\mathrm{Hess}(f)(X, Y) = \langle \nabla_X (\mathrm{grad}\, f), Y \rangle, \quad X, Y \in \mathfrak{X}(M).$$

Given any function $f \in C^\infty(M)$, for any $p \in M$ and for any $u \in T_pM$, the value of the Hessian $\mathrm{Hess}_p(f)(u, u)$ can be computed using geodesics.

Indeed, for any geodesic $\gamma \colon [0, \epsilon] \to M$ such that $\gamma(0) = p$ and $\gamma'(0) = u$, we have

$$\mathrm{Hess}_p(u, u) = \frac{d^2}{dt^2} f(\gamma(t))\Big|_{t=0}.$$

Since the Hessian is a symmetric bilinear form, we obtain $\mathrm{Hess}_p(u, v)$ by polarization; that is,

$$\begin{aligned}\mathrm{Hess}_p(u, v) = \frac{1}{2}(&\mathrm{Hess}_p(u + v, u + v) \\ &- \mathrm{Hess}_p(u, u) - \mathrm{Hess}_p(v, v)).\end{aligned}$$

Let us find the Hessian of the function $f\colon \mathbf{SO}(3) \to \mathbb{R}$ defined in the second example of Section 5.5, with

$$f(R) = (u^\top R v)^2.$$

We found that

$$df_R(X) = 2u^\top X v u^\top R v, \quad X \in R\mathfrak{so}(3)$$

and that the gradient is given by

$$(\mathrm{grad}(f))_R = u^\top R v R(R^\top u v^\top - v u^\top R).$$

To compute the Hessian, we use the curve $\gamma(t) = Re^{tB}$, where $B \in \mathfrak{so}(3)$.

Indeed, it can be shown (see Section 17.4, Proposition 17.19) that the metric induced by the inner product

$$\langle B_1, B_2 \rangle = \mathrm{tr}(B_1^\top B_2) = -\mathrm{tr}(B_1 B_2)$$

on $\mathfrak{so}(n)$ is bi-invariant, and so the curve γ is a geodesic.

First, we compute

$$\begin{aligned}(f(\gamma(t)))'(t) &= ((u^\top Re^{tB}v)^2)'(t) \\ &= 2u^\top Re^{tB}vu^\top RBe^{tB}v,\end{aligned}$$

and then

$$\begin{aligned}\mathrm{Hess}_R(RB, RB) &= (f(\gamma(t)))''(0) \\ &= (2u^\top Re^{tB}vu^\top RBe^{tB}v)'(0) \\ &= 2u^\top RBvu^\top RBv \\ &\quad + 2u^\top Rvu^\top RBR^\top RBv.\end{aligned}$$

By polarization, we obtain

$$\begin{aligned}\mathrm{Hess}_R(X, Y) &= 2u^\top Xvu^\top Yv \\ &\quad + u^\top Rvu^\top XR^\top Yv \\ &\quad + u^\top Rvu^\top YR^\top Xv,\end{aligned}$$

with $X, Y \in R\mathfrak{so}(3)$.

13.2 The Exponential Map

The idea behind the exponential map is to parametrize a Riemannian manifold, M, locally near any $p \in M$ in terms of a map from the tangent space T_pM to the manifold, this map being defined in terms of geodesics.

Definition 13.4. Let (M, g) be a Riemannian manifold. For every $p \in M$, let $\mathcal{D}(p)$ (or simply, $\mathcal{D}$) be the open subset of T_pM given by

$$\mathcal{D}(p) = \{v \in T_pM \mid \gamma_v(1) \quad \text{is defined}\},$$

where γ_v is the unique maximal geodesic with initial conditions $\gamma_v(0) = p$ and $\gamma_v'(0) = v$. The *exponential map* is the map, $\exp_p \colon \mathcal{D}(p) \to M$, given by

$$\exp_p(v) = \gamma_v(1).$$

It is easy to see that $\mathcal{D}(p)$ is *star-shaped*, which means that if $w \in \mathcal{D}(p)$, then the line segment $\{tw \mid 0 \leq t \leq 1\}$ is contained in $\mathcal{D}(p)$.

In view of the remark made at the end of the previous section, the curve

$$t \mapsto \exp_p(tv), \qquad tv \in \mathcal{D}(p)$$

is the geodesic, γ_v, through p such that $\gamma_v'(0) = v$. Such geodesics are called *radial geodesics*. The point, $\exp_p(tv)$, is obtained by running along the geodesic, γ_v, an arc length equal to $t \left\| v \right\|$, starting from p.

In general, $\mathcal{D}(p)$ is a proper subset of T_pM.

Definition 13.5. A Riemannian manifold, (M, g), is *geodesically complete* iff $\mathcal{D}(p) = T_pM$, for all $p \in M$, that is, iff the exponential, $\exp_p(v)$, is defined for all $p \in M$ and for all $v \in T_pM$.

Equivalently, (M, g) is geodesically complete iff every geodesic can be extended indefinitely.

Geodesically complete manifolds have nice properties, some of which will be investigated later.

Observe that $d(\exp_p)_0 = \mathrm{id}_{T_pM}$.

It follows from the inverse function theorem that $\exp_p$ is a diffeomorphism from some open ball in T_pM centered at 0 to M.

The following stronger proposition plays a crucial role:

Proposition 13.5. *Let (M, g) be a Riemannian manifold. For every point, $p \in M$, there is an open subset, $W \subseteq M$, with $p \in W$ and a number $\epsilon > 0$, so that*

(1) Any two points q_1, q_2 of W are joined by a unique geodesic of length $< \epsilon$.

(2) This geodesic depends smoothly upon q_1 and q_2, that is, if $t \mapsto \exp_{q_1}(tv)$ is the geodesic joining q_1 and q_2 $(0 \leq t \leq 1)$, then $v \in T_{q_1}M$ depends smoothly on (q_1, q_2).

(3) For every $q \in W$, the map $\exp_q$ is a diffeomorphism from the open ball, $B(0, \epsilon) \subseteq T_qM$, to its image, $U_q = \exp_q(B(0, \epsilon)) \subseteq M$, with $W \subseteq U_q$ and U_q open.

For any $q \in M$, an open neighborhood of q of the form, $U_q = \exp_q(B(0, \epsilon))$, where $\exp_q$ is a diffeomorphism from the open ball $B(0, \epsilon)$ onto U_q, is called a *normal neighborhood*.

Definition 13.6. Let (M, g) be a Riemannian manifold. For every point, $p \in M$, the *injectivity radius of M at p*, denoted $i(p)$, is the least upper bound of the numbers, $r > 0$, such that $\exp_p$ is a diffeomorphism on the open ball $B(0, r) \subseteq T_pM$. The *injectivity radius, $i(M)$, of M* is the greatest lower bound of the numbers, $i(p)$, where $p \in M$.

For every $p \in M$, we get a chart, (U_p, φ), where $U_p = \exp_p(B(0, i(p)))$ and $\varphi = \exp^{-1}$, called a *normal chart*.

If we pick any orthonormal basis, $(e_1, \ldots, e_n)$, of T_pM, then the x_i's, with $x_i = pr_i \circ \exp^{-1}$ and pr_i the projection onto $\mathbb{R}e_i$, are called *normal coordinates* at p (here, $n = \dim(M)$).

These are defined up to an isometry of T_pM.

The following proposition shows that Riemannian metrics do not admit any local invariants of order one:

Proposition 13.6. *Let (M, g) be a Riemannian manifold. For every point, $p \in M$, in normal coordinates at p,*

$$g\left(\frac{\partial}{\partial x_i}, \frac{\partial}{\partial x_j}\right)_p = \delta_{ij} \qquad \textit{and} \qquad \Gamma^k_{ij}(p) = 0.$$

For the next proposition, known as *Gauss Lemma*, we need to define *polar coordinates* on T_pM.

If $n = \dim(M)$, observe that the map,
$(0, \infty) \times S^{n-1} \longrightarrow T_pM - \{0\}$, given by

$$(r, v) \mapsto rv, \qquad r > 0,\ v \in S^{n-1}$$

is a diffeomorphism, where S^{n-1} is the sphere of radius $r = 1$ in T_pM.

Then, the map, $(0, i(p)) \times S^{n-1} \longrightarrow U_p - \{p\}$ given by

$$(r, v) \mapsto \exp_p(rv), \qquad 0 < r < i(p),\ v \in S^{n-1}$$

is also a diffeomorphism.

Proposition 13.7. *(Gauss Lemma) Let (M, g) be a Riemannian manifold. For every point, $p \in M$, the images, $\exp_p(S(0, r))$, of the spheres, $S(0, r) \subseteq T_pM$, centered at 0 by the exponential map, $\exp_p$, are orthogonal to the radial geodesics $r \mapsto \exp_p(rv)$ through p for all $r < i(p)$, with $v \in S^{n-1}$. This means that for any differentiable curve $t \mapsto v(t)$ on the unit sphere S^{n-1}, the corresponding curve on M*

$$t \mapsto \exp_p(rv(t)) \quad \textit{with } r \textit{ fixed},$$

is orthogonal to the radial geodesic

$$r \mapsto \exp_p(rv(t)) \quad \textit{with } t \textit{ fixed } (0 < r < i(p)).$$

Furthermore, in polar coordinates, the pull-back metric, $\exp^* g$, induced on T_pM is of the form

$$exp^* g = dr^2 + g_r,$$

where g_r is a metric on the unit sphere, S^{n-1}, with the property that g_r/r^2 converges to the standard metric on S^{n-1} (induced by $\mathbb{R}^n$) when r goes to zero (here, $n = \dim(M)$).

Consider any piecewise smooth curve

$$\omega\colon [a,b] \to U_p - \{p\}.$$

We can write each point $\omega(t)$ uniquely as

$$\omega(t) = \exp_p(r(t)v(t)),$$

with $0 < r(t) < i(p)$, $v(t) \in T_pM$ and $\|v(t)\| = 1$.

Proposition 13.8. *Let (M,g) be a Riemannian manifold. We have*

$$\int_a^b \|\omega'(t)\| \, dt \geq |r(b) - r(a)|,$$

where equality holds only if the function r is monotone and the function v is constant. Thus, the shortest path joining two concentric spherical shells, $\exp_p(S(0,r_1))$ and $\exp_p(S(0,r_2))$, is a radial geodesic.

We now get the following important result from Proposition 13.7 and Proposition 13.8:

Theorem 13.9. *Let (M, g) be a Riemannian manifold. Let W and ϵ be as in Proposition 13.5 and let $\gamma\colon [0,1] \to M$ be the geodesic of length $< \epsilon$ joining two points q_1, q_2 of W. For any other piecewise smooth path, ω, joining q_1 and q_2, we have*

$$\int_0^1 \|\gamma'(t)\| \, dt \leq \int_0^1 \|\omega'(t)\| \, dt$$

where equality holds only if the images $\omega([0,1])$ and $\gamma([0,1])$ coincide. Thus, γ is the shortest path from q_1 to q_2.

Here is an important consequence of Theorem 13.9.

Corollary 13.10. *Let* (M, g) *be a Riemannian manifold. If* $\omega\colon [0, b] \to M$ *is any curve parametrized by arc-length and* ω *has length less than or equal to the length of any other curve from* $\omega(0)$ *to* $\omega(b)$*, then* ω *is a geodesic.*

Definition 13.7. *Let* (M, g) *be a Riemannian manifold. A geodesic,* $\gamma\colon [a, b] \to M$*, is minimal iff its length is less than or equal to the length of any other piecewise smooth curve joining its endpoints.*

Theorem 13.9 asserts that any sufficiently small segment of a geodesic is minimal.

On the other hand, a long geodesic may not be minimal. For example, a great circle arc on the unit sphere is a geodesic. If such an arc has length greater than π, then it is not minimal.

Minimal geodesics are generally not unique. For example, any two antipodal points on a sphere are joined by an infinite number of minimal geodesics.

A *broken geodesic* is a piecewise smooth curve as in Definition 13.1, where each curve segment is a geodesic.

Proposition 13.11. *A Riemannian manifold, (M, g), is connected iff any two points of M can be joined by a broken geodesic.*

In general, if M is connected, then it is not true that any two points are joined by a geodesic. However, this will be the case if M is geodesically complete, as we will see in the next section.

Next, we will see that a Riemannian metric induces a distance on the manifold whose induced topology agrees with the original metric.

13.3 Complete Riemannian Manifolds, the Hopf-Rinow Theorem and the Cut Locus

Every connected Riemannian manifold, (M, g), is a metric space in a natural way.

Furthermore, M is a complete metric space iff M is geodesically complete.

In this section, we explore briefly some properties of complete Riemannian manifolds.

Proposition 13.12. *Let (M, g) be a connected Riemannian manifold. For any two points, $p, q \in M$, let $d(p, q)$ be the greatest lower bound of the lengths of all piecewise smooth curves joining p to q. Then, d is a metric on M and the topology of the metric space, (M, d), coincides with the original topology of M.*

The distance, d, is often called the *Riemannian distance* on M. For any $p \in M$ and any $\epsilon > 0$, the *metric ball of center p and radius ϵ* is the subset, $B_\epsilon(p) \subseteq M$, given by

$$B_\epsilon(p) = \{q \in M \mid d(p, q) < \epsilon\}.$$

The next proposition follows easily from Proposition 13.5:

Proposition 13.13. *Let (M, g) be a connected Riemannian manifold. For any compact subset, $K \subseteq M$, there is a number $\delta > 0$ so that any two points, $p, q \in K$, with distance $d(p, q) < \delta$ are joined by a unique geodesic of length less than δ. Furthermore, this geodesic is minimal and depends smoothly on its endpoints.*

Recall from Definition 13.5 that (M, g) is geodesically complete iff the exponential map, $v \mapsto \exp_p(v)$, is defined for all $p \in M$ and for all $v \in T_pM$.

We now prove the following important theorem due to *Hopf and Rinow* (1931):

Theorem 13.14. *(Hopf-Rinow) Let* (M, g) *be a connected Riemannian manifold. If there is a point,* $p \in M$*, such that* $\exp_p$ *is defined on the entire tangent space,* T_pM*, then any point,* $q \in M$*, can be joined to* p *by a minimal geodesic. As a consequence, if* M *is geodesically complete, then any two points of* M *can be joined by a minimal geodesic.*

Proof. The most beautiful proof is Milnor's proof in [37], Chapter 10, Theorem 10.9.

Theorem 13.14 implies the following result (often known as the *Hopf-Rinow Theorem*):

Theorem 13.15. *Let (M, g) be a connected, Riemannian manifold. The following statements are equivalent:*

(1) The manifold (M, g) is geodesically complete, that is, for every $p \in M$, every geodesic through p can be extended to a geodesic defined on all of $\mathbb{R}$.

(2) For every point, $p \in M$, the map $\exp_p$ is defined on the entire tangent space, T_pM.

(3) There is a point, $p \in M$, such that $\exp_p$ is defined on the entire tangent space, T_pM.

(4) Any closed and bounded subset of the metric space, (M, d), is compact.

(5) The metric space, (M, d), is complete (that is, every Cauchy sequence converges).

In view of Theorem 13.15, a connected Riemannian manifold, (M, g), is geodesically complete iff the metric space, (M, d), is complete.

We will refer simply to M as a *complete Riemannian manifold* (it is understood that M is connected).

Also, by (4), every compact, Riemannian manifold is complete.

If we remove any point, p, from a Riemannian manifold, M, then $M - \{p\}$ is not complete since every geodesic that formerly went through p yields a geodesic that can't be extended.

Assume (M, g) is a complete Riemannian manifold. Given any point, $p \in M$, it is interesting to consider the subset, $\mathcal{U}_p \subseteq T_pM$, consisting of all $v \in T_pM$ such that the geodesic

$$t \mapsto \exp_p(tv)$$

is a minimal geodesic up to $t = 1 + \epsilon$, for some $\epsilon > 0$.

The subset $\mathcal{U}_p$ is open and star-shaped and it turns out that $\exp_p$ is a diffeomorphism from $\mathcal{U}_p$ onto its image, $\exp_p(\mathcal{U}_p)$, in M.

The left-over part, $M - \exp_p(\mathcal{U}_p)$ (if nonempty), is actually equal to $\exp_p(\partial\mathcal{U}_p)$ and it is an important subset of M called the *cut locus of p*.

Proposition 13.16. *Let (M, g) be a complete Riemannian manifold. For any geodesic, $\gamma\colon [0, a] \to M$, from $p = \gamma(0)$ to $q = \gamma(a)$, the following properties hold:*

(i) If there is no geodesic shorter than γ between p and q, then γ is minimal on $[0, a]$.

(ii) If there is another geodesic of the same length as γ between p and q, then γ is no longer minimal on any larger interval, $[0, a + \epsilon]$.

(iii) If γ is minimal on any interval, I, then γ is also minimal on any subinterval of I.

Again, assume (M, g) is a complete Riemannian manifold and let $p \in M$ be any point. For every $v \in T_pM$, let

$$\begin{aligned} I_v \;=\; \{s \in \mathbb{R} \cup \{\infty\} \mid & \text{ the geodesic } \quad t \mapsto \exp_p(tv) \\ & \text{is minimal on } [0, s]\}. \end{aligned}$$

It is easy to see that I_v is a closed interval, so $I_v = [0, \rho(v)]$ (with $\rho(v)$ possibly infinite).

It can be shown that if $w = \lambda v$, then $\rho(v) = \lambda\rho(w)$, so we can restrict our attention to unit vectors, v.

It can also be shown that the map, $\rho\colon S^{n-1} \to \mathbb{R}$, is continuous, where S^{n-1} is the unit sphere of center 0 in T_pM, and that $\rho(v)$ is bounded below by a strictly positive number.

Definition 13.8. Let (M, g) be a complete Riemannian manifold and let $p \in M$ be any point. Define $\mathcal{U}_p$ by

$$\begin{aligned} \mathcal{U}_p &= \left\{ v \in T_pM \;\middle|\; \rho\left(\frac{v}{\|v\|}\right) > \|v\| \right\} \\ &= \{v \in T_pM \mid \rho(v) > 1\} \end{aligned}$$

and the *cut locus* of p by

$$\mathrm{Cut}(p) = \exp_p(\partial\mathcal{U}_p) = \{\exp_p(\rho(v)v) \mid v \in S^{n-1}\}.$$

The set $\mathcal{U}_p$ is open and star-shaped.

The boundary, $\partial\mathcal{U}_p$, of $\mathcal{U}_p$ in T_pM is sometimes called the *tangential cut locus* of p and is denoted $\widetilde{\mathrm{Cut}}(p)$.

Remark: The cut locus was first introduced for convex surfaces by Poincaré (1905) under the name *ligne de partage*.

According to Do Carmo [17] (Chapter 13, Section 2), for Riemannian manifolds, the cut locus was introduced by J.H.C. Whitehead (1935).

But it was Klingenberg (1959) who revived the interest in the cut locus and showed its usefuleness.

Proposition 13.17. *Let* (M, g) *be a complete Riemannian manifold. For any point,* $p \in M$*, the sets* $\exp_p(\mathcal{U}_p)$ *and* $\mathrm{Cut}(p)$ *are disjoint and*

$$M = \exp_p(\mathcal{U}_p) \cup \mathrm{Cut}(p).$$

Observe that the injectivity radius, $i(p)$, of M at p is equal to the distance from p to the cut locus of p:

$$i(p) = d(p, \mathrm{Cut}(p)) = \inf_{q \in \mathrm{Cut}(p)} d(p, q).$$

Consequently, the injectivity radius, $i(M)$, of M is given by

$$i(M) = \inf_{p \in M} d(p, \mathrm{Cut}(p)).$$

If M is compact, it can be shown that $i(M) > 0$. It can also be shown using Jacobi fields that $\exp_p$ is a diffeomorphism from $\mathcal{U}_p$ onto its image, $\exp_p(\mathcal{U}_p)$.

Thus, $\exp_p(\mathcal{U}_p)$ is diffeomorphic to an open ball in $\mathbb{R}^n$ (where $n = \dim(M)$) and the cut locus is closed.

Hence, the manifold, M, is obtained by gluing together an open n-ball onto the cut locus of a point. In some sense the topology of M is "contained" in its cut locus.

Given any sphere, S^{n-1}, the cut locus of any point, p, is its antipodal point, $\{-p\}$.

In general, the cut locus is very hard to compute. In fact, even for an ellipsoid, the determination of the cut locus of an arbitrary point was a matter of conjecture for a long time. This conjecture was settled around 2011.

13.4 The Calculus of Variations Applied to Geodesics; The First Variation Formula

Given a Riemannian manifold, (M, g), the path space, $\Omega(p, q)$, was introduced in Definition 13.1.

It is an "infinite dimensional" manifold. By analogy with finite dimensional manifolds we define a kind of tangent space to $\Omega(p, q)$ at a point ω.

In this section, it is convenient to assume that paths in $\Omega(p, q)$ are parametrized over the interval $[0, 1]$.

Definition 13.9. For every "point" $\omega \in \Omega(p, q)$, we define the "*tangent space*", $T_\omega\Omega(p, q)$, of $\Omega(p, q)$ at ω, to be the space of all piecewise smooth vector fields, W, along ω, for which $W(0) = W(1) = 0$ (we may assume that our paths, ω, are parametrized over $[0, 1]$).

Now, if $F\colon \Omega(p,q) \to \mathbb{R}$ is a real-valued function on $\Omega(p,q)$, it is natural to ask what the induced "tangent map",

$$dF_\omega\colon T_\omega\Omega(p,q) \to \mathbb{R},$$

should mean (here, we are identifying $T_{F(\omega)}\mathbb{R}$ with $\mathbb{R}$).

Observe that $\Omega(p,q)$ is not even a topological space so the answer is far from obvious!

In the case where $f\colon M \to \mathbb{R}$ is a function on a manifold, there are various equivalent ways to define df, one of which involves curves.

For every $v \in T_pM$, if $\alpha\colon (-\epsilon,\epsilon) \to M$ is a curve such that $\alpha(0) = p$ and $\alpha'(0) = v$, then we know that

$$df_p(v) = \left.\frac{d(f(\alpha(t)))}{dt}\right|_{t=0}.$$

We may think of α as a small *variation* of p. Recall that p is a *critical point* of f iff $df_p(v) = 0$, for all $v \in T_pM$.

Rather than attempting to define dF_ω (which requires some conditions on F), we will mimic what we did with functions on manifolds and define what is a *critical path* of a function, $F\colon \Omega(p,q) \to \mathbb{R}$, using the notion of *variation*.

Now, geodesics from p to q are special paths in $\Omega(p,q)$ and they turn out to be the critical paths of the *energy function*,

$$E_a^b(\omega) = \int_a^b \|\omega'(t)\|^2 \, dt,$$

where $\omega \in \Omega(p,q)$, and $0 \leq a < b \leq 1$.

Definition 13.10. Given any path, $\omega \in \Omega(p, q)$, a *variation of ω (keeping endpoints fixed)* is a function, $\widetilde{\alpha}\colon (-\epsilon, \epsilon) \to \Omega(p, q)$, for some $\epsilon > 0$, such that

(1) $\widetilde{\alpha}(0) = \omega$

(2) There is a subdivision, $0 = t_0 < t_1 < \cdots < t_{k-1} < t_k = 1$ of $[0, 1]$ so that the map

$$\alpha\colon (-\epsilon, \epsilon) \times [0, 1] \to M$$

defined by $\alpha(u, t) = \widetilde{\alpha}(u)(t)$ is smooth on each strip $(-\epsilon, \epsilon) \times [t_i, t_{i+1}]$, for $i = 0, \ldots, k - 1$.

If U is an open subset of $\mathbb{R}^n$ containing the origin and if we replace $(-\epsilon, \epsilon)$ by U in the above, then $\widetilde{\alpha}\colon U \to \Omega(p, q)$ is called an *n-parameter variation* of ω.

The function α is also called a *variation* of ω.

Since each $\widetilde{\alpha}(u)$ belongs to $\Omega(p, q)$, note that

$$\alpha(u, 0) = p, \quad \alpha(u, 1) = q, \quad \text{for all } u \in (-\epsilon, \epsilon).$$

The function, $\widetilde{\alpha}$, may be considered as a "smooth path" in $\Omega(p, q)$, since for every $u \in (-\epsilon, \epsilon)$, the map $\widetilde{\alpha}(u)$ is a curve in $\Omega(p, q)$ called a *curve in the variation (or longitudinal curve of the variation)*.

The "tangent vector" $\frac{d\widetilde{\alpha}}{du}(0) \in T_\omega\Omega(p, q)$ is defined to be the vector field W along ω given by

$$W_t = \frac{\partial \alpha}{\partial u}(u, t)\bigg|_{u=0}.$$

By definition,

$$\frac{d\widetilde{\alpha}}{du}(0)_t = W_t, \quad t \in [0, 1].$$

Clearly, $W \in T_\omega\Omega(p, q)$. In particular, $W(0) = W(1) = 0$.

The vector field, W, is also called the *variation vector field* associated with the variation α.

Besides the curves in the variation, $\widetilde{\alpha}(u)$ (with $u \in (-\epsilon, \epsilon)$), for every $t \in [0, 1]$, we have a curve, $\alpha_t \colon (-\epsilon, \epsilon) \to M$, called a *transversal curve of the variation*, defined by

$$\alpha_t(u) = \widetilde{\alpha}(u)(t),$$

and W_t is equal to the velocity vector, $\alpha_t'(0)$, at the point $\omega(t) = \alpha_t(0)$.

For ϵ sufficiently small, the vector field, W_t, is an infinitesimal model of the variation $\widetilde{\alpha}$.

We can show that for any $W \in T_\omega\Omega(p, q)$ there is a variation, $\widetilde{\alpha} \colon (-\epsilon, \epsilon) \to \Omega(p, q)$, which satisfies the conditions

$$\widetilde{\alpha}(0) = \omega, \qquad \frac{d\widetilde{\alpha}}{du}(0) = W.$$

As we said earlier, given a function, $F \colon \Omega(p, q) \to \mathbb{R}$, we do not attempt to define the differential, dF_ω, but instead, the notion of critical path.

Definition 13.11. Given a function, $F\colon \Omega(p, q) \to \mathbb{R}$, we say that a path, $\omega \in \Omega(p, q)$, is a *critical path* for F iff

$$\left.\frac{dF(\widetilde{\alpha}(u))}{du}\right|_{u=0} = 0,$$

for every variation, $\widetilde{\alpha}$, of ω (which implies that the derivative $\left.\frac{dF(\widetilde{\alpha}(u))}{du}\right|_{u=0}$ is defined for every variation, $\widetilde{\alpha}$, of ω).

For example, if F takes on its minimum on a path ω_0 and if the derivatives $\frac{dF(\widetilde{\alpha}(u))}{du}$ are all defined, then ω_0 is a critical path of F.

We will apply the above to two functions defined on $\Omega(p, q)$:

(1) The *energy function* (also called *action integral*):

$$E_a^b(\omega) = \int_a^b \|\omega'(t)\|^2 \, dt.$$

(We write $E = E_0^1$.)

(2) The *arc-length function*,

$$L_a^b(\omega) = \int_a^b \|\omega'(t)\| \, dt.$$

The quantities $E_a^b(\omega)$ and $L_a^b(\omega)$ can be compared as follows: if we apply the Cauchy-Schwarz's inequality,

$$\left(\int_a^b f(t)g(t)dt\right)^2 \leq \left(\int_a^b f^2(t)dt\right)\left(\int_a^b g^2(t)dt\right)$$

with $f(t) \equiv 1$ and $g(t) = \|\omega'(t)\|$, we get

$$(L_a^b(\omega))^2 \leq (b-a)E_a^b,$$

where equality holds iff g is constant; that is, iff the parameter t is proportional to arc-length.

Now, suppose that there exists a minimal geodesic, γ, from p to q. Then,

$$E(\gamma) = L(\gamma)^2 \leq L(\omega)^2 \leq E(\omega),$$

where the equality $L(\gamma)^2 = L(\omega)^2$ holds only if ω is also a minimal geodesic, possibly reparametrized.

On the other hand, the equality $L(\omega) = E(\omega)^2$ can hold only if the parameter is proportional to arc-length along ω.

This proves that $E(\gamma) < E(\omega)$ unless ω is also a minimal geodesic. We just proved:

Proposition 13.18. *Let (M, g) be a complete Riemannian manifold. For any two points, $p, q \in M$, if $d(p, q) = \delta$, then the energy function, $E\colon \Omega(p, q) \to \mathbb{R}$, takes on its minimum, δ^2, precisely on the set of minimal geodesics from p to q.*

Next, we are going to show that the critical paths of the energy function are exactly the geodesics. For this, we need the *first variation formula*.

Let $\widetilde{\alpha}\colon (-\epsilon, \epsilon) \to \Omega(p, q)$ be a variation of ω and let

$$W_t = \frac{\partial \alpha}{\partial u}(u, t)\bigg|_{u=0}$$

be its associated variation vector field.

Furthermore, let

$$V_t = \frac{d\omega}{dt} = \omega'(t),$$

the velocity vector of ω and

$$\Delta_t V = V_{t_+} - V_{t_-},$$

the discontinuity in the velocity vector at t, which is nonzero only for $t = t_i$, with $0 < t_i < 1$ (see the definition of $\gamma'((t_i)_+)$ and $\gamma'((t_i)_-)$ just after Definition 13.1).

Theorem 13.19. *(First Variation Formula) For any path, $\omega \in \Omega(p, q)$, we have*

$$\frac{1}{2}\frac{dE(\widetilde{\alpha}(u))}{du}\bigg|_{u=0} = -\sum_i \langle W_t, \Delta_t V\rangle - \int_0^1 \left\langle W_t, \frac{D}{dt}V_t \right\rangle dt,$$

where $\widetilde{\alpha}\colon (-\epsilon, \epsilon) \to \Omega(p, q)$ is any variation of ω.

Intuitively, the first term on the right-hand side shows that varying the path ω in the direction of decreasing "kink" tends to decrease E.

The second term shows that varying the curve in the direction of its acceleration vector, $\frac{D}{dt}\omega'(t)$, also tends to reduce E.

A geodesic, γ, (parametrized over $[0, 1]$) is smooth on the entire interval $[0, 1]$ and its acceleration vector, $\frac{D}{dt}\gamma'(t)$, is identically zero along γ. This gives us half of

Theorem 13.20. *Let (M, g) be a Riemanian manifold. For any two points, $p, q \in M$, a path, $\omega \in \Omega(p, q)$ (parametrized over $[0, 1]$), is critical for the energy function, E, iff ω is a geodesic.*

Remark: If $\omega \in \Omega(p, q)$ is parametrized by arc-length, it is easy to prove that

$$\left.\frac{dL(\widetilde{\alpha}(u))}{du}\right|_{u=0} = \frac{1}{2} \left.\frac{dE(\widetilde{\alpha}(u))}{du}\right|_{u=0}.$$

As a consequence, a path, $\omega \in \Omega(p, q)$ is critical for the arc-length function, L, iff it can be reparametrized so that it is a geodesic

In order to go deeper into the study of geodesics we need Jacobi fields and the "second variation formula", both involving a curvature term.

Therefore, we now proceed with a more thorough study of curvature on Riemannian manifolds.