Geodesics on a Riemannian Manifold via the Calculus of Variations

Sterling Scarlett, Grant Talbert, Arystan Shokan

December 2025

Contents

1 Motivation

One of the first geometric facts we learn is that, in the Euclidean plane, the shortest path between two points is a straight line. This statement is so familiar that it is easy to forget how much structure is hidden in it: we are using both the linear structure of $\mathbb{R}^n$ and its standard inner product to talk about lengths, angles, and straightness. As soon as we leave the flat world of Euclidean space, the situation becomes less obvious.

For example, on the surface of the Earth, airplanes do not follow straight line segments in $\mathbb{R}^3$, but rather arcs of great circles on the sphere. These arcs are locally distance minimizing: near any point on such a curve, if you look only at sufficiently short subsegments, they realize the shortest path between their endpoints along the surface of the Earth. From the intrinsic point of view of the sphere, they play exactly the role that straight lines play in $\mathbb{R}^n$. This leads to the central question of this paper:

How can we make sense of "straightest" or "shortest" curves at constant veloctiy on an arbitrary smooth manifold?

The modern answer begins with the notion of a *Riemannian metric*. Informally, a Riemannian metric g on a smooth manifold M gives each tangent space T_pM the structure of an inner product space, in a way that varies smoothly with the point p. Once such a metric is chosen, we can measure the length of tangent vectors, and by integration we obtain lengths of curves. From the lengths of curves we define a distance function $d_g(p, q)$ by taking an infimum over all curves from p to q.

Curves that "realize" this distance (or at least make it stationary) are called *geodesics*. Geodesics play the role of straight lines in this more general setting. They are important both for purely geometric reasons and for applications: in physics, for instance, geodesics in a Lorentzian manifold represent the trajectories of freely-falling particles in general relativity.

There are several different but equivalent ways to characterize geodesics:

- As locally length-minimizing curves.
- As curves with zero covariant acceleration, $\nabla_{\dot\gamma}\dot\gamma = 0$.
- As critical points of the energy functional

  $$E(\gamma) = \frac{1}{2}\int_a^b \|\dot\gamma(t)\|_g^2 \, \mathrm{d}t$$

  under variations that fix the endpoints.

The last point is the variational point of view, and it is the main focus of this paper. Instead of guessing the geodesic equation and then justifying it, we will start from the length and energy functionals on the space of curves and derive the geodesic equation as an Euler–Lagrange equation. This approach fits naturally with classical problems in the calculus of variations and gives a conceptually clean derivation of the geodesic equation.

1.1 A motivating example: the punctured plane

Before getting into definitions, it is useful to see that the distance function behaves in a slightly subtle way even in a very simple example.

Example 1.1 (The punctured plane)**.** Let $M = \mathbb{R}^2 \setminus \{0\}$ with the Riemannian metric g induced by the standard Euclidean inner product on $\mathbb{R}^2$. Fix a point $p \in M$ and let $q = -p$. In the full plane $\mathbb{R}^2$, the unique straight line segment from p to q has length $2\|p\|$ and realizes the Euclidean distance between p and q.

However, in M this straight segment is not allowed, because it passes through the origin, which has been removed. Any piecewise smooth curve $\gamma\colon [a, b] \to M$ with $\gamma(a) = p$ and $\gamma(b) = q$ must "go around" the origin. Intuitively, we still expect the distance between p and q to be $2\|p\|$, but there will be no curve in M that actually achieves this length.

Example 1.2. Take $p = (1, 0)$ and $q = (-1, 0)$, and let $M = \mathbb{R}^2 \setminus \{0\}$ as above. For each $\varepsilon > 0$ consider the curve γ_ε that goes from $(1, 0)$ to $(\varepsilon, 0)$ along the x-axis, then follows a semicircle of radius ε around the origin to $(-\varepsilon, 0)$, and then goes from $(-\varepsilon, 0)$ to $(-1, 0)$ along the x-axis.

The first and last segments have total length $2(1-\varepsilon)$, and the semicircular arc has length $\pi\varepsilon$. Writing L_g for the length of a curve γ_ε,

$$L_g(\gamma_\varepsilon) = 2(1-\varepsilon) + \pi\varepsilon = 2 + (\pi - 2)\varepsilon.$$

As $\varepsilon \to 0^+$ we get $L_g(\gamma_\varepsilon) \to 2$. If we view distance d_g as the infimum over lengths of piecewise smooth curves from p to q, then $d_g(p, q) = 2$.

On the other hand, every curve in M from p to q must go around the origin, and hence has length strictly greater than 2: if a curve could achieve length exactly 2, it would have to coincide with the straight line segment from $(1, 0)$ to $(-1, 0)$, and that passes through the origin. Thus $d_g(p, q) = 2$, but there is no curve γ in M with $L_g(\gamma) = d_g(p, q)$.

This example illustrates both the usefulness and the subtlety of the distance function. It also hints at the importance of global assumptions like completeness in the study of geodesics. In a *complete* or *geodesically complete* and connected Riemannian manifold, any two points can be joined by a geodesic of minimal length ([Spi79, Ch9 Thm. 18]).

1.2 Goals and outline

The main goal of this paper is to give a self-contained derivation of the geodesic equation on a Riemannian manifold using the calculus of variations, and to connect this equation with the intuitive idea of geodesics as shortest curves. Roughly speaking, we will:

- Define Riemannian metrics, curve length, and the induced distance function d_g.
- Introduce variations of curves and define the energy functional.
- Derive the first variation formula for the energy.
- Show that critical points of the energy are precisely smooth curves satisfying the geodesic equation $\nabla_{\dot\gamma}\dot\gamma = 0$.
- Relate geodesics to length-minimizing curves and discuss examples.

The structure of the paper is as follows. In Section 2 we review Riemannian metrics, the length of curves, and the Riemannian distance. In Section 3 we define variations of curves and the energy functional. Section 4 is devoted to the first variation formula. In Section 5 we characterize geodesics as curves satisfying a second-order ODE, the geodesic equation, and show the equivalence with being critical points of the energy. In Section 6 we relate energy and length and explain how geodesics arise as locally minimizing curves. Finally, in Section 7 we compute geodesics explicitly in some important examples.

2 Riemannian metrics, length, and distance

Throughout, M will denote a smooth (C^∞) manifold of dimension n.

2.1 Riemannian metrics

Definition 2.1. Let M be a smooth manifold. A *Riemannian metric* g on M is a smooth section of $T^*M \otimes T^*M$. More intuitively, g is a smooth assignment to each point $p \in M$ of an inner product

$$g_p \colon T_pM \times T_pM \to \mathbb{R}$$

such that the map

$$M \times TM \times TM \to \mathbb{R}, \qquad (p, v, w) \mapsto g_p(v, w)$$

is smooth when restricted to $TM \times TM$ over M. We usually write $\langle v, w \rangle_g = g_p(v, w)$ when $v, w \in T_pM$, and $\|v\|_g = \sqrt{\langle v, v\rangle_g}$. A *Riemannian manifold* is a manifold M together with a Riemannian metric g.

All manifolds admit at least one Riemannian metric by [Lee12] Proposition 13.3, but oftentimes this metric may not be desirable.

Example 2.2. Let $(U, (x^1, \dots, x^n))$ be a coordinate chart on M. The coordinate vector fields $\partial_{x^1}, \dots, \partial_{x^n}$ form a basis for T_pM at each $p \in U$. The metric g is then determined by its components

$$g_{ij}(p) = g_p\left(\frac{\partial}{\partial x^i}\Big|_p, \frac{\partial}{\partial x^j}\Big|_p \right),$$

which form a smooth positive-definite symmetric matrix (g_{ij}). We often write

$$g = \sum_{i,j=1}^{n} g_{ij}(x)\, \mathrm{d}x^i \otimes \mathrm{d}x^j.$$

Unless otherwise specified, all manifolds in this paper are assumed to be Riemannian manifolds with Riemannian metric g.

2.2 Length of curves

Definition 2.3. Let $\gamma\colon [a, b] \to M$ a piecewise C^∞ curve on a manifold M. The *velocity* of γ at t is $\dot\gamma(t) = \mathrm{d}\gamma/\mathrm{d}t \in T_{\gamma(t)}M$. The *length* of γ with respect to g is

$$L_g(\gamma) = \int_a^b \|\dot\gamma(t)\|_g \,\mathrm{d}t = \int_a^b \sqrt{g_{\gamma(t)}(\dot\gamma(t), \dot\gamma(t))}\,\mathrm{d}t.$$

It is straightforward to check that the length is invariant under orientation-preserving reparametrizations: if $\phi\colon [c, d] \to [a, b]$ is a smooth, strictly increasing bijection and $\tilde\gamma = \gamma \circ \phi$, then $L_g(\tilde\gamma) = L_g(\gamma)$.

Example 2.4. In $\mathbb{R}^n$ with the standard metric $g = \sum_{i=1}^n \mathrm{d}x^i \otimes \mathrm{d}x^i$, the length reduces to the usual length formula

$$L_g(\gamma) = \int_a^b \sqrt{\sum_{i=1}^{n} \left(\frac{\mathrm{d}\gamma^i}{\mathrm{d}t}(t) \right)^2}\,\mathrm{d}t,$$

where $\gamma(t) = (\gamma^1(t), \dots, \gamma^n(t))$.

2.3 The Riemannian distance function

Definition 2.5. Let (M, g) be a connected Riemannian manifold. For $p, q \in M$ the *Riemannian distance* between p and q is

$$d_g(p,q) = \inf\{L_g(\gamma) \mid \gamma\colon [a,b] \to M \text{ piecewise } C^\infty,\ \gamma(a) = p,\ \gamma(b) = q\}.$$

We now prove that this defines a metric on M.

Proposition 2.6. *The function $d_g\colon M \times M \to \mathbb{R}$ is a metric, i.e.*

(i) $d_g(p,q) \geq 0$ *and* $d_g(p,q) = 0$ *if and only if* $p = q$,

(ii) $d_g(p,q) = d_g(q,p)$, *and*

(iii) $d_g(p,r) \leq d_g(p,q) + d_g(q,r)$ *for all* $p, q, r \in M$.

Proof. (i) By definition, $L_g(\gamma) \geq 0$ for every curve γ, so $d_g(p,q) \geq 0$. If $p = q$, the constant curve $\gamma(t) \equiv p$ has length zero, so $d_g(p,p) = 0$. Conversely, suppose $d_g(p,q) = 0$. Let $(U, (x^1, \ldots, x^n))$ be a coordinate chart containing p, and fix the Euclidean norm $\|\cdot\|$ on $\mathbb{R}^n$. Since g is positive definite and smooth, there exist constants $0 < c < C < \infty$ such that

$$c\,\|v\| \leq \|v\|_g \leq C\,\|v\|$$

for all v in tangent spaces over a small neighborhood of p. If $q \neq p$ is sufficiently close to p, then there is a smooth curve γ in U from p to q, and its Euclidean length is bounded below by some positive number depending on $\|x(q) - x(p)\|$. The inequalities above imply $L_g(\gamma) \geq c\,L_{\mathrm{Eucl}}(\gamma) > 0$, so $d_g(p,q) > 0$. Thus if $d_g(p,q) = 0$, we must have $p = q$.

(ii) Symmetry is clear: if γ is a curve from p to q, then the reversed curve $\tilde{\gamma}(t) = \gamma(a + b - t)$ has the same length and goes from q to p. Taking infima gives $d_g(p,q) = d_g(q,p)$.

(iii) For the triangle inequality, fix $p, q, r \in M$ and $\varepsilon > 0$. Choose piecewise smooth curves γ_1 from p to q and γ_2 from q to r such that

$$L_g(\gamma_1) \leq d_g(p,q) + \varepsilon, \qquad L_g(\gamma_2) \leq d_g(q,r) + \varepsilon.$$

Define the concatenated curve

$$\gamma(t) = \begin{cases} \gamma_1(2t), & t \in [0, \frac{1}{2}], \\ \gamma_2(2t-1), & t \in [\frac{1}{2}, 1]. \end{cases}$$

Then

$$L_g(\gamma) = L_g(\gamma_1) + L_g(\gamma_2) \leq d_g(p,q) + d_g(q,r) + 2\varepsilon.$$

Taking the infimum over all such γ gives $d_g(p,r) \leq d_g(p,q) + d_g(q,r) + 2\varepsilon$, and since $\varepsilon > 0$ was arbitrary, we obtain the triangle inequality. $\square$

It is also true that the metric topology induced by d_g agrees with the original manifold topology. A full proof requires some more work, but the idea is that in local coordinates, the inequality $c\,\|v\| \leq \|v\|_g \leq C\,\|v\|$ implies that the d_g-balls and the Euclidean balls define the same notion of "small neighborhood." We will not need the precise details later, so see [Lee12] Theorem 13.29 for a full proof.

3 Variations and the energy functional

As noted earlier, a geodesic is intuitively understood to be a curve of minimal length, a locally length-minimizing curve. We also noted that this notion does not always agree with the Riemannian distance function, as seen in Example 1.2. To make precise when these two notions agree, we first provide a classification of geodesics. This classification will be the focus of much of this paper.

In order to find critical points of the length or energy functional, we need a way to perturb a given curve. For this, we look to the calculus of variations.

3.1 Variations of curves

Definition 3.1. Let $\gamma\colon [a,b] \to M$ be a piecewise C^∞ curve. A *variation* of γ is a smooth map

$$\alpha\colon (-\varepsilon, \varepsilon) \times [a,b] \to M$$

for some $\varepsilon > 0$ such that:

(i) $\alpha(0,t) = \gamma(t)$ for all $t \in [a,b]$;

(ii) There is a partition
$$a = t_0 < t_1 < \cdots < t_N = b$$
such that for each fixed u, the curve
$$\bar{\alpha}(u)\colon [a,b] \to M, \qquad \bar{\alpha}(u)(t) = \alpha(u,t),$$
is C^∞ on each closed subinterval $[t_i, t_{i+1}]$.

We say that the variation *fixes endpoints* if $\alpha(u,a)$ and $\alpha(u,b)$ are independent of u.

Definition 3.2. Let α be a variation of γ. The *variation vector field* along γ is the vector field V along γ defined by
$$V(t) = \left.\frac{\partial \alpha}{\partial u}\right|_{u=0}(t) \in T_{\gamma(t)}M.$$
If α fixes endpoints, then $V(a) = V(b) = 0$.

Remark 3.3. Conversely, given a sufficiently nice vector field V along γ, one can construct a variation α with variation field V. We will use this in the proof of the geodesic equation.

3.2 The energy functional

Directly working with the length functional

$$L_g(\gamma) = \int_a^b \|\dot{\gamma}(t)\|_g \, \mathrm{d}t$$

is possible but somewhat technical because of the square root in the definition of the norm. It is more convenient to work with the *energy functional*, whose integrand is quadratic in the velocity.

Definition 3.4. Let $\gamma\colon [a,b] \to M$ be piecewise C^∞. The *energy* of γ is

$$E_a^b(\gamma) = \frac{1}{2}\int_a^b \|\dot{\gamma}(t)\|_g^2 \, \mathrm{d}t = \frac{1}{2}\int_a^b g_{\gamma(t)}\big(\dot{\gamma}(t), \dot{\gamma}(t)\big) \, \mathrm{d}t.$$

If we fix endpoints $p, q \in M$ and an interval $[a, b]$, we can think of E_a^b as a functional on the space of piecewise C^∞ curves $\gamma\colon [a, b] \to M$ with $\gamma(a) = p$ and $\gamma(b) = q$. We will say that γ is a *critical point* of E_a^b if the derivative

$$\frac{\mathrm{d}}{\mathrm{d}u} E_a^b(\bar{\alpha}(u))\bigg|_{u=0}$$

vanishes for every variation α that fixes endpoints.

Proposition 3.5. *Let $\gamma\colon [a, b] \to M$ be piecewise C^∞. Then*

$$L_g(\gamma)^2 \leq 2(b-a)\, E_a^b(\gamma),$$

with equality if and only if $\|\dot{\gamma}(t)\|_g$ is constant on $[a, b]$.

Proof. By the Cauchy–Schwarz inequality on L^2 applied to $\|\dot{\gamma}(t)\|_g$ and the constant function at 1,

$$\left(\int_a^b \|\dot{\gamma}(t)\|_g \ \mathrm{d}t\right)^2 \leq (b-a) \int_a^b \|\dot{\gamma}(t)\|^2 \ \mathrm{d}t,$$

with equality if and only if $|\gamma(t)|$ is constant almost everywhere. The left-hand side is $L_g(\gamma)^2$, while the right-hand side is

$$(b-a) \int_a^b \|\dot{\gamma}(t)\|_g^2 \ \mathrm{d}t = 2(b-a)\, E_a^b(\gamma).$$

$\square$

This shows that among curves with fixed endpoints and fixed parameter interval $[a, b]$, those with constant speed are the ones for which energy and length are most tightly related. In particular, any curve that minimizes energy among all curves with the same endpoints and parameter interval must have constant speed. We will later see that this relation admits a converse in the sense that critical points for L_g are reparameterizations of those for E_a^b.

4 The first variation of energy

We now compute the derivative of the energy functional along a variation. This is the key step in deriving the geodesic equation.

4.1 Left and right derivatives

Because we allow piecewise smooth curves, it is useful to introduce the notation for left and right derivatives at the break points.

Let $\gamma\colon [a, b] \to M$ be piecewise C^∞ with partition $a = t_0 < \cdots < t_N = b$ such that γ is C^∞ on each open subinterval (t_i, t_{i+1}).

Definition 4.1. For $1 \leq i \leq N-1$, the *left* and *right* derivatives of γ at t_i are defined by

$$\dot{\gamma}(t_i^-) = \lim_{t \to t_i^-} \dot{\gamma}(t), \qquad \dot{\gamma}(t_i^+) = \lim_{t \to t_i^+} \dot{\gamma}(t),$$

computed in local coordinates. We define the *jump* at t_i by

$$\Delta_{t_i}\dot{\gamma} = \dot{\gamma}(t_i^+) - \dot{\gamma}(t_i^-).$$

We also set

$$\Delta_{t_0}\dot{\gamma} = \dot{\gamma}(t_0^+), \qquad \Delta_{t_N}\dot{\gamma} = -\dot{\gamma}(t_N^-).$$

Remark 4.2. If γ is C^∞ on the entire interval $[a, b]$, then all jumps vanish except for the jumps at the endpoints $\Delta_{t_0}\dot{\gamma}$ and $\Delta_{t_N}\dot{\Gamma}$.

4.2 The first variation formula

We will repeatedly use the following linear-algebraic lemma.

Lemma 4.3. *Let $W\colon [a,b] \to \mathbb{R}^n$ be continuous. Suppose*

$$\int_a^b \langle V(t), W(t)\rangle \,\mathrm{d}t = 0$$

for every smooth $V\colon [a,b] \to \mathbb{R}^n$ with $V(a) = V(b) = 0$. Then $W(t) \equiv 0$ on $[a,b]$.

Proof. Fix $t_0 \in (a,b)$ and suppose $W(t_0) \neq 0$. Let $e_1, \ldots, e_n$ be the standard basis of $\mathbb{R}^n$, and write $W(t_0) = \sum_{k=1}^n w_k e_k$. Then at least one w_k is nonzero; without loss of generality $w_1 \neq 0$.

By continuity, there exists $\delta > 0$ such that $\langle e_1, W(t)\rangle$ has the same sign as w_1 for all $t \in (t_0 - \delta, t_0 + \delta)$. Choose a smooth bump function $\varphi\colon [a,b] \to \mathbb{R}$ such that φ is supported in $(t_0 - \delta, t_0 + \delta)$ and $\varphi(t) \geq 0$ with $\varphi(t_0) > 0$.

Define $V(t) = \varphi(t)e_1$. Then V is smooth, $V(a) = V(b) = 0$, and

$$\int_a^b \langle V(t), W(t)\rangle \,\mathrm{d}t = \int_a^b \varphi(t)\,\langle e_1, W(t)\rangle \,\mathrm{d}t.$$

The integrand is nonnegative and positive near t_0, so the integral is strictly positive. This contradicts the assumption that the integral is always zero. Hence $W(t) = 0$ for all t. □

Corollary 4.4. *Let $W : [a,b] \to \mathbb{R}$ be continuous. Suppose that*

$$\int_a^b V(t)W(t)\,\mathrm{d}t = 0$$

for every smooth $V : [a,b] \to \mathbb{R}$ with $V(a) = V(b) = 0$. Then $W(t) \equiv 0$ on $[a,b]$.

Proof. Taking $n = 1$ in Lemma 4.3, note that $\langle V(t), W(t)\rangle = V(t)W(t)$. □

Since manifolds have local coordinates, and thus can be locally identified with $\mathbb{R}^n$, we can invoke Lemma 4.3 on manifolds provided we restrict ourselves within a single chart.

Before we state and prove the first variation formula, we introduce some notation.

Definition 4.5. Let (M,g) be a Riemannian n-manifold. The *Christoffel symbol* Γ_{ij}^k is defined as

$$\Gamma_{ij}^k = \frac{1}{2}\sum_{m=1}^n g^{km}\left(\frac{\partial g_{jm}}{\partial x^i} + \frac{\partial g_{im}}{\partial x^j} - \frac{\partial g_{ij}}{\partial x^m}\right).$$

Define the *Levi-Civita connection* $\nabla : \mathfrak{X}(M) \times \mathfrak{X}(M) \to \mathfrak{X}(M)$ on (M,g) as the map

$$\nabla(\partial_{x^i}, \partial_{x^j}) = \sum_{k=1}^n \Gamma_{ij}^k \partial_{x^k}$$

which is $C^\infty(M)$-linear (and thus $\mathbb{R}$-linear via constant functions) in the first argument and satisfies the following product rule in the second: for any smooth $f : M \to \mathbb{R}$,

$$\nabla_X(fY) = f\nabla_X Y + (Xf)Y.$$

We generally write $\nabla(X,Y)$ as $\nabla_X Y$. We will often call $\nabla_X Y$ the *covariant derivative* of Y in the direction of X.

Remark 4.6. Given $p \in M$, we can evaluate the Christoffel symbols $\Gamma^k_{ij}(p)$ at the point p by taking a coordinate representation of g in a chart around p, and then evaluating the partial dertivatives of g_{ij} at p. Moreover, this representation is manifestly independent of the choice of coordinates by invariance of g under choice of coordinates.

For more properties of the Levi-Civita connection, see [Lee18] Theorem 5.10. For more generalities on connections and covariant derivatives, see [Lee18] chapters 4 and 5. We will only need the coordinate representation of the Levi-Civita connection using the Christoffel symbols for this paper.

Theorem 4.7 (First variation of energy)**.** *Let (M, g) be a Riemannian manifold and $\gamma\colon [a, b] \to M$ a piecewise C^∞ curve with partition $a = t_0 < \cdots < t_N = b$. Let α be a variation of γ with variation vector field V. Then*

$$\frac{\mathrm{d}}{\mathrm{d}u} E^b_a(\bar{\alpha}(u))\Big|_{u=0} = -\int_a^b \langle V(t), \nabla_{\dot{\gamma}}\dot{\gamma}(t)\rangle_g \,\mathrm{d}t - \sum_{i=0}^{N} \langle V(t_i), \Delta_{t_i}\dot{\gamma}\rangle_g .$$

Here ∇ is the Levi-Civita connection of (M, g).

Proof. We first express the energy in local coordinates and then differentiate. Choose a coordinate chart $(U, (x^1, \ldots, x^n))$ such that $\gamma([t_i, t_{i+1}]) \subseteq U$. We may refine the partition if necessary to ensure full containment within U. Write

$$\gamma(t) = (\gamma^1(t), \ldots, \gamma^n(t))$$

and

$$\dot{\gamma}(t) = \sum_{k=1}^{n} \frac{\mathrm{d}\gamma^k}{\mathrm{d}t}(t) \frac{\partial}{\partial x^k}\Big|_{\gamma(t)}.$$

Let g_{ij} be the components of g in these coordinates. Then the integrand of the energy on $[t_i, t_{i+1}]$ is

$$\frac{1}{2}\left\|\dot{\gamma}(t)\right\|_g^2 = \frac{1}{2}\sum_{j,\ell=1}^{n} g_{j\ell}(\gamma(t)) \frac{\mathrm{d}\gamma^j}{\mathrm{d}t}(t)\frac{\mathrm{d}\gamma^\ell}{\mathrm{d}t}(t).$$

Define

$$F(x, y) = \frac{1}{2}\sum_{j,\ell=1}^{n} g_{j\ell}(x) y^j y^\ell, \qquad x \in U,\ y \in \mathbb{R}^n.$$

Since we are working in coordinates, we can identify $T_{\gamma(t)}U \cong \mathbb{R}^n$, yielding

$$E^b_a(\gamma) = \sum_{i=0}^{N-1}\int_{t_i}^{t_{i+1}} F(\gamma(t), \dot{\gamma}(t))\,\mathrm{d}t.$$

The energy functional now resembles a standard form for which the calculus of variations is applicable.

Now consider a variation $\alpha(u, t)$ with $\alpha(0, t) = \gamma(t)$. In coordinates, write

$$\alpha(u, t) = (\alpha^1(u, t), \ldots, \alpha^n(u, t)).$$

For each fixed u, we have a curve $t \mapsto \alpha(u, t)$, and we denote its components by $\alpha^k(u, t)$. The velocity is

$$\frac{\partial\alpha}{\partial t}(u, t) = \sum_{k=1}^{n} \frac{\partial\alpha^k}{\partial t}(u, t)\frac{\partial}{\partial x^k}\Big|_{\alpha(u,t)}.$$

Then

$$E_a^b(\bar{\alpha}(u)) = \sum_{i=0}^{N-1} \int_{t_i}^{t_{i+1}} F\left(\alpha(u,t), \frac{\partial \alpha}{\partial t}(u,t)\right) \, \mathrm{d}t.$$

Differentiating with respect to u under the integral sign, we get

$$\frac{\mathrm{d}}{\mathrm{d}u} E_a^b(\bar{\alpha}(u)) = \sum_{i=0}^{N-1} \int_{t_i}^{t_{i+1}} \left[\sum_{k=1}^{n} \frac{\partial F}{\partial x^k} \frac{\partial \alpha^k}{\partial u} + \sum_{k=1}^{n} \frac{\partial F}{\partial y^k} \frac{\partial}{\partial t}\left(\frac{\partial \alpha^k}{\partial u}\right)\right] \mathrm{d}t.$$

Evaluating at $u = 0$ and using that $\alpha(0,t) = \gamma(t)$, we find

$$\begin{aligned} \frac{\mathrm{d}}{\mathrm{d}u} E_a^b(\bar{\alpha}(u))\Big|_{u=0} = \sum_{i=0}^{N-1} \int_{t_i}^{t_{i+1}} & \left[\sum_{k=1}^{n} \frac{\partial F}{\partial x^k}\Big|_{(\gamma,\dot{\gamma})} \frac{\partial \alpha^k}{\partial u}(0,t) \right. \\ & \left. + \sum_{k=1}^{n} \frac{\partial F}{\partial y^k}\Big|_{(\gamma,\dot{\gamma})} \frac{\partial}{\partial t}\left(\frac{\partial \alpha^k}{\partial u}(0,t)\right)\right] \mathrm{d}t. \end{aligned}$$

We integrate the second term by parts on each interval $[t_i, t_{i+1}]$:

$$\begin{aligned} \int_{t_i}^{t_{i+1}} \sum_{k=1}^{n} \frac{\partial F}{\partial y^k} \frac{\partial}{\partial t}\left(\frac{\partial \alpha^k}{\partial u}(0,t)\right) \mathrm{d}t = & \left[\sum_{k=1}^{n} \frac{\partial F}{\partial y^k} \frac{\partial \alpha^k}{\partial u}(0,t)\right]_{t_i}^{t_{i+1}} \\ & - \int_{t_i}^{t_{i+1}} \sum_{k=1}^{n} \frac{\partial}{\partial t}\left(\frac{\partial F}{\partial y^k}\right) \frac{\partial \alpha^k}{\partial u}(0,t) \, \mathrm{d}t. \end{aligned}$$

Summing over i and combining terms, we obtain

$$\begin{aligned} \frac{\mathrm{d}}{\mathrm{d}u} E_a^b(\bar{\alpha}(u))\Big|_{u=0} = & \sum_{i=0}^{N-1} \int_{t_i}^{t_{i+1}} \sum_{k=1}^{n} \left(\frac{\partial F}{\partial x^k} - \frac{\partial}{\partial t}\left(\frac{\partial F}{\partial y^k}\right)\right) \frac{\partial \alpha^k}{\partial u}(0,t) \, \mathrm{d}t \\ & + \sum_{i=0}^{N-1} \left[\sum_{k=1}^{n} \frac{\partial F}{\partial y^k} \frac{\partial \alpha^k}{\partial u}(0,t)\right]_{t_i}^{t_{i+1}}. \end{aligned}$$

The last sum telescopes. Evaluating the boundary terms at the internal points $t_1, \ldots, t_{N-1}$ yields expressions involving the left and right derivatives of γ at those points. One checks that this boundary contribution can be written as

$$-\sum_{i=0}^{N} \langle V(t_i), \Delta_{t_i} \dot{\gamma} \rangle_g \,,$$

where V is the variation vector field and $\Delta_{t_i}\dot{\gamma}$ is the jump in the velocity at t_i. We omit some routine algebra; the main point is that the terms at t_i involve differences of the form

$$\left\langle V(t_i), \dot{\gamma}(t_i^+) \right\rangle_g - \left\langle V(t_i), \dot{\gamma}(t_i^-) \right\rangle_g = \langle V(t_i), \Delta_{t_i} \dot{\gamma} \rangle_g \,.$$

The remaining integral gives the Euler–Lagrange part. A direct computation shows that

$$\frac{\partial F}{\partial x^k} - \frac{\partial}{\partial t}\left(\frac{\partial F}{\partial y^k}\right) = -\left\langle \frac{\partial}{\partial x^k}, \nabla_{\dot{\gamma}} \dot{\gamma} \right\rangle_g ,$$

where ∇ is the Levi-Civita connection. Therefore

$$\sum_{k=1}^{n} \left(\frac{\partial F}{\partial x^k} - \frac{\partial}{\partial t}\left(\frac{\partial F}{\partial y^k}\right)\right) \frac{\partial \alpha^k}{\partial u}(0,t) = -\langle V(t), \nabla_{\dot{\gamma}} \dot{\gamma}(t) \rangle_g \,.$$

Substituting this back into the expression for the derivative completes the proof. □

5 Geodesics and the geodesic equation

We can now define geodesics and characterize them as solutions of a second order ODE.

Definition 5.1. A piecewise C^∞ curve $\gamma\colon [a,b] \to M$ is called a *geodesic* if it is a critical point of the energy functional E_a^b among all variations that fix endpoints.

Using the first variation formula, we obtain a more geometric description.

Theorem 5.2. *A piecewise C^∞ curve $\gamma\colon [a,b] \to M$ is a critical point of E_a^b for fixed endpoints if and only if:*

(i) γ is in fact C^∞ on $[a,b]$ (so all jumps $\Delta_{t_i}\dot\gamma$ vanish), and

(ii) γ satisfies the geodesic equation

$$\nabla_{\dot\gamma}\dot\gamma(t) = 0 \quad \textit{for all } t \in [a,b].$$

Proof. ($\implies$) Assume γ is a critical point of E_a^b. Let α be any variation of γ fixing endpoints, with variation vector field V. Then by Theorem 4.7,

$$0 = \frac{\mathrm{d}}{\mathrm{d}u} E_a^b(\bar\alpha(u))\Big|_{u=0} = -\int_a^b \langle V, \nabla_{\dot\gamma}\dot\gamma\rangle_g \,\mathrm{d}t - \sum_{i=0}^N \langle V(t_i), \Delta_{t_i}\dot\gamma\rangle_g \,.$$

We first show that the jump terms vanish. Fix an index i and choose a variation α whose variation field V is supported very close to t_i, with $V(a) = V(b) = 0$ and $V(t_j) = 0$ for all $j \neq i$ (this can be done by building V from a bump function and then integrating to get α; see remark 5.3). For such a variation, the integral term can be made arbitrarily small, but the sum over i reduces to the single term $-\langle V(t_i), \Delta_{t_i}\dot\gamma\rangle_g$. Since the whole expression must be zero for all such V, we conclude that

$$\langle V(t_i), \Delta_{t_i}\dot\gamma\rangle_g = 0$$

for all $V(t_i) \in T_{\gamma(t_i)}M$. This forces $\Delta_{t_i}\dot\gamma = 0$.

Thus $\dot\gamma(t)$ has no jumps and is continuous on $[a,b]$. Since γ is C^∞ on each subinterval and $\dot\gamma$ is continuous at the break points, standard results on ODEs imply that γ is in fact C^∞ on all of $[a,b]$.

With γ now smooth, all $\Delta_{t_i}\dot\gamma = 0$, so the first variation formula simplifies to

$$\int_a^b \langle V, \nabla_{\dot\gamma}\dot\gamma\rangle_g \,\mathrm{d}t = 0$$

for every smooth vector field V along γ with $V(a) = V(b) = 0$. By working in local coordinates and applying Lemma 4.3, we see that this forces $\nabla_{\dot\gamma}\dot\gamma(t) = 0$ for all t.

($\impliedby$) Conversely, suppose γ is C^∞ and satisfies $\nabla_{\dot\gamma}\dot\gamma = 0$. Let α be a variation that fixes endpoints. Then all jumps vanish other than the boundary jumps, but V vanishes at the boundary since α fixes endpoints, so the sum component vanishes. The integral term in the first variation formula also vanishes since the covariant derivative vanishes. Hence $\mathrm{d}E_a^b(\bar\alpha(u))/\,\mathrm{d}u|_{u=0} = 0$ for all α, so γ is a critical point. ∎

Remark 5.3. To justify the existence of variations with a prescribed variation field V, one can proceed as follows. In local coordinates, define

$$\alpha(u,t) = \exp_{\gamma(t)}(uV(t)),$$

where exp is the exponential map associated to g. For small u, this is well-defined and yields a smooth map α with $\alpha(0,t) = \gamma(t)$ and variation field V. If we want the endpoints fixed, we arrange $V(a) = V(b) = 0$. We will not develop the full theory of the exponential map here; instead, we can work locally and patch together variations using bump functions.

In local coordinates, the geodesic equation takes a more explicit form.

Proposition 5.4 (Geodesic equation in coordinates)**.** *A smooth curve $\gamma\colon [a,b] \to M$ with coordinate representation $\gamma(t) = (\gamma^1(t), \ldots, \gamma^n(t))$ is a geodesic if and only if its components satisfy*

$$\frac{\mathrm{d}^2\gamma^k}{\mathrm{d}t^2}(t) + \sum_{i,j=1}^{n} \Gamma^k_{ij}(\gamma(t)) \frac{\mathrm{d}\gamma^i}{\mathrm{d}t}(t) \frac{\mathrm{d}\gamma^j}{\mathrm{d}t}(t) = 0, \qquad k = 1, \ldots, n.$$

Proof. Recall definition 4.5 which defines ∇ on basis vectors. It suffices to show that by imposing bilinearity and the product rule, we recover the claimed ODE. We have

$$\begin{aligned}
\nabla_{\dot\gamma}\dot\gamma &= \nabla_{\dot\gamma}\left(\sum_{i=1}^{n} \frac{\mathrm{d}\gamma^i}{\mathrm{d}t}\frac{\partial}{\partial x^i}\right) \\
&= \sum_{i=1}^{n} \frac{\mathrm{d}\gamma}{\mathrm{d}t}\left(\frac{\mathrm{d}\gamma^i}{\mathrm{d}t}\right)\frac{\partial}{\partial x^i} + \frac{\mathrm{d}\gamma^i}{\mathrm{d}t}\nabla_{\dot\gamma}\frac{\partial}{\partial x^i} \\
&= \sum_{i=1}^{n} \frac{\mathrm{d}^2\gamma^i}{\mathrm{d}t^2}\frac{\partial}{\partial x^i} + \frac{\mathrm{d}\gamma^i}{\mathrm{d}t}\nabla_{\dot\gamma}\frac{\partial}{\partial x^i} \\
&= \sum_{i=1}^{n} \frac{\mathrm{d}^2\gamma^i}{\mathrm{d}t^2}\frac{\partial}{\partial x^i} + \frac{\mathrm{d}\gamma^i}{\mathrm{d}t}\sum_{j=1}^{n}\frac{\mathrm{d}\gamma^j}{\mathrm{d}t}\nabla_{\partial_i}\frac{\partial}{\partial x^j} \\
&= \sum_{i=1}^{n} \frac{\mathrm{d}^2\gamma^i}{\mathrm{d}t^2}\frac{\partial}{\partial x^i} + \sum_{j=1}^{n}\sum_{k=1}^{n}\frac{\mathrm{d}\gamma^i}{\mathrm{d}t}\frac{\mathrm{d}\gamma^j}{\mathrm{d}t}\Gamma^k_{ij}\frac{\partial}{\partial x^k} \\
&= \left(\sum_{k=1}^{n}\frac{\mathrm{d}^2\gamma^k}{\mathrm{d}t^2}\frac{\partial}{\partial x^k}\right) + \left(\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n}\frac{\mathrm{d}\gamma^i}{\mathrm{d}t}\frac{\mathrm{d}\gamma^j}{\mathrm{d}t}\Gamma^k_{ij}\frac{\partial}{\partial x^k}\right) \\
&= \sum_{k=1}^{n}\left(\frac{\mathrm{d}^2\gamma^k}{\mathrm{d}t^2} + \sum_{i,j=1}^{n}\Gamma^k_{ij}\frac{\mathrm{d}\gamma^i}{\mathrm{d}t}\frac{\mathrm{d}\gamma^j}{\mathrm{d}t}\right)\frac{\partial}{\partial x^k}.
\end{aligned}$$

Here we have noted that $\dot\gamma$ acts on $\dot\gamma^i$ by $\dot\gamma(\dot\gamma^i) = \ddot\gamma^i$. Thus $\nabla_{\dot\gamma}\dot\gamma = 0$ if and only if each coordinate component satisfies the claimed ODE. □

6 Length versus energy and minimizing properties

We now relate the variational characterization of geodesics (via energy) to the more geometric intuition of geodesics as locally length-minimizing curves.

6.1 Critical points of length vs. critical points of energy

Let $\gamma\colon [a,b] \to M$ be a smooth curve with $\dot\gamma(t) \neq 0$ for all t. Define its *arc-length parameter* by

$$s(t) = \int_a^t \|\dot\gamma(t)\|_g \, \mathrm{d}t.$$

Then $s\colon [a,b] \to [0, L_g(\gamma)]$ is smooth and strictly increasing, hence a diffeomorphism onto its image. We can invert it to obtain $t(s)$ and define the reparametrized curve

$$\tilde{\gamma}(s) = \gamma(t(s)).$$

By construction, $\tilde{\gamma}$ has unit speed:

$$\|\dot{\tilde{\gamma}}(s)\|_g = 1$$

for all s. In particular, $L_g(\tilde{\gamma}) = L_g(\gamma)$. We will refer to this as he *arc-length reparameterization* of γ.

The length functional is less convenient to differentiate because of the square root, but for unit-speed curves it is closely related to the energy.

Theorem 6.1. *Let $\gamma\colon [a,b] \to M$ be a smooth curve with $\dot{\gamma}(t) \neq 0$ for all t, and let $\tilde{\gamma}$ be its arc-length reparametrization. Then:*

(i) γ is a critical point of the energy functional E_a^b with fixed endpoints if and only if its arc-length reparametrization $\tilde{\gamma}$ is a critical point of the length functional.

(ii) Any smooth curve σ that is critical for the length functional (with fixed endpoints) and has nonvanishing velocity can be reparametrized to a geodesic.

Sketch of proof. A full proof requires computing the first variation of the length functional, which is similar to, but slightly more complicated than, the computation for the energy. The idea is as follows.

For a unit-speed curve $\tilde{\gamma}$, the length and energy functionals satisfy

$$L_g(\tilde{\gamma}) = \int_0^{L_g(\gamma)} 1 \, \mathrm{d}s = L_g(\gamma),$$

and

$$E_0^{L_g(\gamma)}(\tilde{\gamma}) = \frac{1}{2} \int_0^{L_g(\gamma)} 1^2 \, \mathrm{d}s = \frac{1}{2} L_g(\gamma).$$

Thus up to a constant factor, L and E coincide on unit-speed curves. The first variation of L along variations that respect unit speed is proportional to the first variation of E.

More concretely, if $\tilde{\alpha}$ is a variation of $\tilde{\gamma}$ through unit-speed curves with fixed endpoints, then

$$\frac{\mathrm{d}}{\mathrm{d}u} L_g(\tilde{\alpha}(u))\Big|_{u=0} = \frac{1}{\|\dot{\tilde{\gamma}}\|_g^2} \frac{\mathrm{d}}{\mathrm{d}u} E(\tilde{\alpha}(u))\Big|_{u=0}.$$

Therefore, the vanishing of the first variation of E is equivalent to the vanishing of the first variation of L under such variations.

The second statement follows by applying arc-length reparametrization and then using the characterization of geodesics as critical points of E from Theorem 5.2. Details can be filled in by adapting the proof of the first variation formula to L instead of E. □

6.2 Local minimizing property of geodesics

In general, geodesics are critical points of the length functional, not necessarily global minimizers. Nevertheless, sufficiently short geodesic segments do minimize length between their endpoints.

Theorem 6.2 (Local minimizing property)**.** *Let (M, g) be a Riemannian manifold and $p \in M$. There exists a neighborhood U of p such that for any $q \in U$ there is a unique geodesic segment γ from p to q lying in U, and γ is the unique minimizing curve between p and q.*

Idea of proof. The proof uses the exponential map $\exp_p\colon T_pM \to M$, which is defined by $\exp_p(v) = \gamma_v(1)$, where γ_v is the unique geodesic with $\gamma_v(0) = p$ and $\dot\gamma_v(0) = v$. For sufficiently small v, the map $\exp_p$ is a diffeomorphism onto a neighborhood U of p. This induces coordinates in which geodesics through p correspond to straight lines in T_pM, and one can use this to show that the radial geodesic from p to q is the unique minimizing curve in U.

The full proof is beyond the scope of this paper, so we refer the reader to [Lee18] chapter 6, and in particular Proposition 6.11. $\square$

7 Examples

We now compute geodesics explicitly in some important examples, illustrating how the general theory plays out in practice.

7.1 Euclidean space $\mathbb{R}^n$

Consider $M = \mathbb{R}^n$ with the standard Euclidean metric

$$g = \sum_{i=1}^{n} \mathrm{d}x^i \otimes \mathrm{d}x^i.$$

In the standard coordinates $(x^1, \dots, x^n)$ we have $g_{ij} = \delta_{ij}$. All partial derivatives $\partial g_{ij}/\partial x^k$ are zero, so all Christoffel symbols vanish:

$$\Gamma^k_{ij} = 0 \quad \text{for all } i, j, k.$$

The geodesic equation therefore reduces to

$$\frac{\mathrm{d}^2\gamma^k}{\mathrm{d}t^2} = 0, \qquad k = 1, \dots, n.$$

The general solution is

$$\gamma^k(t) = A^k t + B^k,$$

so γ is an affine map:

$$\gamma(t) = At + B$$

for some constant vectors $A, B \in \mathbb{R}^n$. These are precisely straight lines. Thus our abstract definition of geodesics recovers the familiar fact that geodesics in Euclidean space are straight lines with constant velocity.

7.2 The round two-sphere S^2

Let $M = S^2 \subset \mathbb{R}^3$ be the unit sphere with the metric induced by the Euclidean inner product. To compute the geodesic equation, we introduce spherical coordinates.

Write a point on S^2 as

$$(\cos\theta\cos\varphi,\ \cos\theta\sin\varphi,\ \sin\theta),$$

where $\theta \in (-\frac{\pi}{2}, \frac{\pi}{2})$ is latitude and $\varphi \in (0, 2\pi)$ is longitude. In these coordinates, a straightforward computation shows that the induced metric is

$$g = \mathrm{d}\theta^2 + \cos^2\theta\ \mathrm{d}\varphi^2.$$

(If one uses colatitude instead of latitude, one gets $g = \mathrm{d}\theta^2 + \sin^2\theta\ \mathrm{d}\varphi^2$; both descriptions are equivalent up to a coordinate change.)

For definiteness, let us take

$$g_{\theta\theta} = 1, \qquad g_{\varphi\varphi} = \sin^2\theta, \qquad g_{\theta\varphi} = g_{\varphi\theta} = 0.$$

Then the inverse matrix has

$$g^{\theta\theta} = 1, \qquad g^{\varphi\varphi} = \frac{1}{\sin^2\theta}, \qquad g^{\theta\varphi} = g^{\varphi\theta} = 0.$$

Using the formula for Christoffel symbols, we compute the nonzero ones:

$$\Gamma^{\theta}_{\varphi\varphi} = \frac{1}{2} g^{\theta\theta}\left(\frac{\partial g_{\varphi\theta}}{\partial x^{\varphi}} + \frac{\partial g_{\varphi\theta}}{\partial x^{\varphi}} - \frac{\partial g_{\varphi\varphi}}{\partial x^{\theta}}\right) = -\sin\theta\cos\theta,$$

$$\Gamma^{\varphi}_{\theta\varphi} = \Gamma^{\varphi}_{\varphi\theta} = \frac{1}{2} g^{\varphi\varphi}\left(\frac{\partial g_{\varphi\varphi}}{\partial x^{\theta}} + \frac{\partial g_{\theta\varphi}}{\partial x^{\varphi}} - \frac{\partial g_{\theta\varphi}}{\partial x^{\varphi}}\right) = \cot\theta.$$

All other Γ^k_{ij} vanish.

Therefore, a curve $\gamma(t) = (\theta(t), \varphi(t))$ on S^2 is a geodesic if and only if it satisfies

$$\frac{\mathrm{d}^2\theta}{\mathrm{d}t^2} - \sin\theta\cos\theta\left(\frac{\mathrm{d}\varphi}{\mathrm{d}t}\right)^2 = 0, \tag{7.1}$$

$$\frac{\mathrm{d}^2\varphi}{\mathrm{d}t^2} + 2\cot\theta\,\frac{\mathrm{d}\theta}{\mathrm{d}t}\,\frac{\mathrm{d}\varphi}{\mathrm{d}t} = 0. \tag{7.2}$$

Claim: The solution to this system of differential equations are great circles:

Proof: Let $S^2 = \{\mathbf{x} \in \mathbb{R}^3 : \|\mathbf{x}\| = 1\}$ be the unit sphere. Consider a C^2-regular curve $\gamma : I \to S^2$, where $I \subset \mathbb{R}$ is an interval, parametrized in spherical coordinates:

$$\gamma(t) = \big(\sin\theta(t)\cos\varphi(t),\ \sin\theta(t)\sin\varphi(t),\ \cos\theta(t)\big),$$

with $\theta(t) \in (0, \pi)$, $\varphi(t) \in \mathbb{R}$ for all $t \in I$. We assume regularity: $(\dot\theta(t), \dot\varphi(t)) \neq (0, 0)$ for all t.

The spherical orthonormal frame is:

$$\begin{aligned}
\mathbf{e}_r &= (\sin\theta\cos\varphi,\ \sin\theta\sin\varphi,\ \cos\theta),\\
\mathbf{e}_\theta &= (\cos\theta\cos\varphi,\ \cos\theta\sin\varphi,\ -\sin\theta),\\
\mathbf{e}_\varphi &= (-\sin\varphi,\ \cos\varphi,\ 0).
\end{aligned}$$

These satisfy $\mathbf{e}_r \times \mathbf{e}_\theta = \mathbf{e}_\varphi$, $\mathbf{e}_\theta \times \mathbf{e}_\varphi = \mathbf{e}_r$, $\mathbf{e}_\varphi \times \mathbf{e}_r = \mathbf{e}_\theta$, and $\|\mathbf{e}_r\| = \|\mathbf{e}_\theta\| = \|\mathbf{e}_\varphi\| = 1$.

The velocity vector is:

$$\mathbf{v}(t) = \dot\gamma(t) = \dot\theta\,\mathbf{e}_\theta + \sin\theta\,\dot\varphi\,\mathbf{e}_\varphi.$$

Differentiating $\mathbf{v}$ with respect to t (noting $\mathbf{e}_\theta, \mathbf{e}_\varphi$ depend on t through θ, φ), we obtain the acceleration:

$$\begin{aligned}
\mathbf{a}(t) &= \frac{d}{dt}\Big(\dot\theta\,\mathbf{e}_\theta + \sin\theta\,\dot\varphi\,\mathbf{e}_\varphi\Big)\\
&= \ddot\theta\,\mathbf{e}_\theta + \dot\theta\,\dot{\mathbf{e}}_\theta + \frac{d}{dt}(\sin\theta\,\dot\varphi)\,\mathbf{e}_\varphi + \sin\theta\,\dot\varphi\,\dot{\mathbf{e}}_\varphi.
\end{aligned}$$

We compute the time derivatives of the frame vectors. First, the partial derivatives:

$$\begin{aligned}
\frac{\partial\mathbf{e}_r}{\partial\theta} &= \mathbf{e}_\theta, & \frac{\partial\mathbf{e}_r}{\partial\varphi} &= \sin\theta\,\mathbf{e}_\varphi,\\
\frac{\partial\mathbf{e}_\theta}{\partial\theta} &= -\mathbf{e}_r, & \frac{\partial\mathbf{e}_\theta}{\partial\varphi} &= \cos\theta\,\mathbf{e}_\varphi,\\
\frac{\partial\mathbf{e}_\varphi}{\partial\theta} &= 0, & \frac{\partial\mathbf{e}_\varphi}{\partial\varphi} &= -\sin\theta\,\mathbf{e}_r - \cos\theta\,\mathbf{e}_\theta.
\end{aligned}$$

Thus, by the chain rule:

$$\dot{\mathbf{e}}_\theta = \dot\theta\,\frac{\partial \mathbf{e}_\theta}{\partial\theta} + \dot\varphi\,\frac{\partial \mathbf{e}_\theta}{\partial\varphi} = -\dot\theta\,\mathbf{e}_r + \dot\varphi\cos\theta\,\mathbf{e}_\varphi,$$
$$\dot{\mathbf{e}}_\varphi = \dot\theta\,\frac{\partial \mathbf{e}_\varphi}{\partial\theta} + \dot\varphi\,\frac{\partial \mathbf{e}_\varphi}{\partial\varphi} = \dot\varphi(-\sin\theta\,\mathbf{e}_r - \cos\theta\,\mathbf{e}_\theta).$$

Substituting these into $\mathbf{a}(t)$ gives:

$$\begin{aligned}\mathbf{a}(t) = \ddot\theta\,\mathbf{e}_\theta &+ \dot\theta\big(-\dot\theta\,\mathbf{e}_r + \dot\varphi\cos\theta\,\mathbf{e}_\varphi\big)\\ &+ \big(\cos\theta\,\dot\theta\dot\varphi + \sin\theta\,\ddot\varphi\big)\mathbf{e}_\varphi\\ &+ \sin\theta\,\dot\varphi\big(-\dot\varphi\sin\theta\,\mathbf{e}_r - \dot\varphi\cos\theta\,\mathbf{e}_\theta\big).\end{aligned}$$

Collecting components:

$$\begin{aligned}\mathbf{a}(t) = (\ddot\theta &- \sin\theta\cos\theta\,\dot\varphi^2)\mathbf{e}_\theta\\ &+ \big(\sin\theta\,\ddot\varphi + 2\cos\theta\,\dot\theta\dot\varphi\big)\mathbf{e}_\varphi\\ &+ \big(-\dot\theta^2 - \sin^2\theta\,\dot\varphi^2\big)\mathbf{e}_r.\end{aligned}$$

Equations (7.1) and (7.2) are exactly the conditions that the coefficients of $\mathbf{e}_\theta$ and $\mathbf{e}_\varphi$ in (**??**) vanish. Indeed, multiplying (7.2) by $\sin\theta$ yields:

$$\sin\theta\,\ddot\varphi + 2\cos\theta\,\dot\theta\dot\varphi = 0.$$

Thus, under the given system, $\mathbf{a}(t)$ is purely radial:

$$\mathbf{a}(t) = -\big(\dot\theta^2 + \sin^2\theta\,\dot\varphi^2\big)\,\mathbf{e}_r = \lambda(t)\,\gamma(t),$$

where $\lambda(t) = -\big(\dot\theta^2 + \sin^2\theta\,\dot\varphi^2\big)$ and we used $\gamma(t) = \mathbf{e}_r$.

Define the angular momentum per unit mass vector:

$$\mathbf{L}(t) = \gamma(t)\times\mathbf{v}(t).$$

Differentiating:

$$\begin{aligned}\frac{d\mathbf{L}}{dt} &= \dot\gamma\times\mathbf{v} + \gamma\times\mathbf{a}\\ &= \mathbf{v}\times\mathbf{v} + \gamma\times(\lambda\gamma)\\ &= 0 + \lambda\,(\gamma\times\gamma) = 0.\end{aligned}$$

Hence $\mathbf{L}$ is constant: $\mathbf{L}(t)\equiv\mathbf{L}_0\in\mathbb{R}^3$.

Since $\mathbf{L}_0$ is constant, for all t:

$$\gamma(t)\cdot\mathbf{L}_0 = \gamma(t)\cdot\big(\gamma(t)\times\mathbf{v}(t)\big) = 0,$$

by the scalar triple product identity. Thus $\gamma(I)$ lies in the plane $\Pi = \{\mathbf{x}\in\mathbb{R}^3 : \mathbf{x}\cdot\mathbf{L}_0 = 0\}$ through the origin.

We now consider two cases:

Case 1: $\mathbf{L}_0 \neq 0$. Then Π is a well-defined plane through the origin. The intersection $\Pi\cap S^2$ is a great circle. Since $\gamma(t)\in S^2$ for all t and $\gamma(t)\in\Pi$, we have $\gamma(I)\subseteq\Pi\cap S^2$, i.e., the curve lies on a great circle.

Case 2: $\mathbf{L}_0 = 0$. Then $\gamma(t) \parallel \mathbf{v}(t)$ for all t (since $\gamma \times \mathbf{v} = 0$). But $\gamma(t)$ is a unit vector, so differentiating $\gamma \cdot \gamma = 1$ gives $\gamma \cdot \mathbf{v} = 0$. Combined with $\gamma \times \mathbf{v} = 0$, this implies $\mathbf{v}(t) = 0$ for all t, contradicting regularity unless the curve is a single point. Hence, for a regular curve, $\mathbf{L}_0 \neq 0$.

Thus, every regular solution of the system lies on a great circle.

Conversely, Let C be a great circle on S^2. Then $C = \Pi \cap S^2$ for some plane Π through the origin. Choose any C^2 regular parametrization $\gamma(t)$ of C. Since C is a geodesic on S^2, its acceleration vector as a curve in $\mathbb{R}^3$ must be normal to the surface (Levi-Civita condition). For S^2, the normal direction at $\gamma(t)$ is exactly $\pm\gamma(t)$. Hence $\mathbf{a}(t) = \mu(t)\gamma(t)$ for some scalar function $\mu(t)$. From the decomposition (**??**), this implies the coefficients of $\mathbf{e}_\theta$ and $\mathbf{e}_\varphi$ vanish, yielding precisely (7.1) and (7.2).

Thus, any regular parametrization of a great circle satisfies the system. Moreover, any solution of the system gives a great circle parametrization. This establishes the equivalence.

7.3 The punctured plane revisited

We return to the punctured plane

$$M = \mathbb{R}^2 \setminus \{0\}$$

with the induced Euclidean metric. In local coordinates away from the origin, the metric is just the standard Euclidean one, so geodesics are locally straight lines. However, global properties of these geodesics can be more complicated.

If p and q lie on the same ray emanating from the origin (say $q = \lambda p$ for some $\lambda > 0$), then the straight segment from p to q does not hit the origin and is a geodesic in M that minimizes length between its endpoints. On the other hand, if $q = -p$ as in Example 1.2, then any straight segment from p to q passes through 0 and is not contained in M. There are many geodesics passing near the origin, but none of them connect p to q in a way that realizes the distance $d_g(p, q)$.

This illustrates that the existence of minimizing geodesics between arbitrary pairs of points depends on global conditions like completeness. The punctured plane is incomplete, and the failure of geodesics to realize distances between some points is one manifestation of this incompleteness.

8 Pseduo-Riemannian manifolds

So far, we have restricted our attention only to Riemannian manifolds. However, many important applications of Riemannian geometry involve *pseudo-Riemannian metrics*; for example the theory of General Relativity takes place on a pseudo-Riemannian manifold.

Definition 8.1. Let V be a vector space. A symmetric 2-tensor $g : V \otimes V \to \mathbb{R}$ is *nondegenerate* if for all nonzero $v \in V$, there exists $w \in V$ such that $g(v, w) \neq 0$.

Non-degenerate 2-tensors are a generalization of inner products in the following way. The matrix representation of the Euclidean inner product $\bar{g}$ is the identity matrix $\bar{g}_{ij} = \delta_{ij}$. Any inner product can be transformed into the Euclidean one by changing to an orthonormal basis such that the isomorphism to $\mathbb{R}^n$ given by that basis preserves the inner product; existence of such a basis is easy to prove. Now instead of considering the identity matrix, consider the diagonal matrix whose entries are ± 1. A nondegenerate symmetric 2-tensor can be transformed under some change of basis to such a diagonal matrix.

The striking fact is that the number r of $+1$s and s of -1s in such a diagonal matrix representation of a nondegenerate form is *invariant* under change of basis. Thus we refer to the pair (r, s) as the *signature* of a nondegenerate form.

Definition 8.2. A *pseudo-Riemannian manifold* is a manifold M together with a smooth 2-tensor field $g \in \mathcal{T}^2(M)$ such that each g_p is symmetric and nondegenerate, and such that the signature is the same everywhere on M.

Much of the theory we have developed carries over to the situation for pseudo-Riemannian manifolds. Our derivation of the geodesic equation made reference very rarely to the inner product properties of the metric, and we only used symmetry and that $\langle -, 0\rangle = 0$. Both follow by definition for a pseudo-Riemannian metric, and the latter for any bilinear map. Thus the geodesic equation follows for more general pseudo-Riemannian manifolds.

However, not much more of the theory carries over. For example, $\|\dot{\gamma}(t)\| > 0$ for all t on Riemannian manifolds, but on pseudo-Riemannian manifolds we can have $\|\dot{\gamma}(t)\| = 0$, or even $\|\dot{\gamma}(t)\| < 0$ everywhere. As such the arc-length reparameterization can fail. There are many more examples; for example the Hopf-Rinow theorem (see [Lee18] Theorem 6.19), which has a number of useful corollaries.

Even still, geodesics on pseudo-Riemannian manifolds are of extreme importance. As noted earlier, General Relativity, which is our current best theory of gravity, takes place on a pseudo-Riemannian manifold, and the equation of motion for a particle under the influence of gravity is precisely the geodesic equation. The metric can then be derived using Einstein's equation $G_{\mu\nu} = 8\pi G_N T_{\mu\nu}$.

Conclusion

In this paper we have developed the variational approach to geodesics on Riemannian manifolds. Starting from the definition of a Riemannian metric, we introduced the length and energy functionals on the space of curves, computed the first variation of the energy, and derived the geodesic equation as the Euler–Lagrange equation for critical points of the energy with fixed endpoints. We saw that geodesics can be characterized intrinsically as curves with zero covariant acceleration, and that in local coordinates they satisfy a second-order system of ordinary differential equations involving the Christoffel symbols.

We also discussed the relationship between energy and length, showing that constant-speed geodesics are critical points of the length functional and are locally minimizing curves. Finally, we computed geodesics in several important examples, including Euclidean space, the round sphere, and the punctured plane, illustrating how both local and global geometric features affect the behavior of geodesics.

Many further directions are possible: the study of Jacobi fields and conjugate points, Morse theory on loop spaces, comparison theorems relating curvature to the behavior of geodesics, and the role of geodesics in physical theories such as general relativity. The variational methods and basic geometric ideas developed here provide a foundation for exploring these more advanced topics.

References

[Lee12] John M. Lee. *Introduction to Smooth Manifolds*. Graduate Texts in Mathematics. Springer New York, NY, 2 edition, 2012.

[Lee18] John M. Lee. *Introuction to Riemannian Manifolds*. Graduate Texts in Mathematics. Springer, 2 edition, 2018.

[Spi79] Michael Spivak. *A Comprehensive Introduction to Differential Geometry*. 1979.