Defining the Straightest Path: Theory and Computation of Geodesics in Differential Geometry DRAFT

Piseth Tang
a1906208

November 19, 2025

1   Abstract (Summary)

This report grew out of the final draft that I had written for my capstone course, directed by Professor Raymond Vozzo (my favorite educator during my undergraduate years). However, I aim for it to be a better version and an extension of the original one by refining the structure of my writing to be more cohesive and the story more compelling.

2   Introduction

2.1   Geodesics

Many centuries ago, one of Western civilization's most influential thinkers, Aristotle, developed a theory of motion, which has laid the physical foundations for later thinkers [1]. In this collection of treatises, he describes different kinds of motions and one of them being "uniform motion." He stated that a uniform motion is a motion on a straight line or a great circle. However, this is not entirely accurate as later physicists and scientists refined the notion some more. And from all of that, modern theoretical physicist would tell us that uniform motions are motions along geodesics. Geodesics are a generalization of the concept of a "straight line" from a flat plane to a curved surface. The study of these "shortest curves" was revitalized with foundational work by mathematicians like Bernoulli, Euler, and Gauss [6]. Beyond the theory, let us imagine ourselves on an airplane going from Adelaide to Singapore. We would naturally assume that that our airplane would travel the shortest possible path between the two countries. And we would be right to think so. This "shortest possible path" is a path on a "great circle" (informally, a circle with the largest diameter on a sphere) and we will see later that any great circles on a sphere are geodesics. Therefore, every time we are on airplane, we travel along a geodesic by solving a system of ODEs (modulo wind). Other applications are in physics, especially theoretical physics, where particles in space travel along geodesics to preserve energy; in robotics where motions of robots are computed along geodesics, and many more [2]. This report aims to setup and attack the geodesic problem on general surfaces.

Definition 2.1 (Geodesics). A curve $\gamma : (\alpha, \beta) \to \mathbb{R}^2$ on a surface $S \subseteq \mathbb{R}^3$ is called a *geodesics* if $\ddot{\gamma}$ is zero or perpendicular to the tangent plane of S at the point $\gamma(t)$, i.e., parallel to its unit formal, **for all** values of the parameter t.

Remark. Whenever a definition is introduced, it helps to try to come up with different examples, of varying difficulty and abstractions, to illuminate it further and chip away at its *actual* meaning(s). Although we may not be able to always do this and may require more powerful results, the examples below are hand-picked to enforce the reader's understanding of the above definition from various angles before reading more complicated results. Below are examples that should be kept in our

mathematical toolbox as we shall see later that they can be used to deduce further (remaining) geodesics of more complicated surfaces.

2.2   Examples

1. Any (part of a) straight line on a surface is a geodesics.

   Proof. A straight line can be parameterized with constant speed by the equation
   $$\gamma(t) = \mathbf{a} + t\mathbf{b}$$
   where $\mathbf{a}$ and $\mathbf{b}$ are constant vectors. The velocity vector is found by differentiating with respect to t:
   $$\dot{\gamma}(t) = \mathbf{b}$$
   Differentiating a second time gives the acceleration vector:
   $$\ddot{\gamma}(t) = \mathbf{0}$$
   According to Definition 1.1, a curve is a geodesic if its acceleration vector is zero or parallel to the surface normal. Since the acceleration is the zero vector, the condition is satisfied. Therefore, any straight line on a surface is a geodesic. This immediately implies that all straight lines in a plane, as well as the rulings on cylinders, cones, and hyperboloids of one sheet, are all geodesics. □

2. From Example 1, we can deduce that all straight lines in a plane, hyperboloid of one sheet and are geodesics.

3. Probably the next simplest (but not simple) example of a geodesic can be found in a geometry different from Euclid's, namely spherical geometry. The analog to "straightest curves/lines" on a sphere are *great circles*. Before we

   Definition 2.2 (Tangent Vector Field)**.** ...

   Definition 2.3 (Parallel tangent vector field)**.** ...

   We will see later that it is, quite rigorously, the shortest possible curve on a sphere. But before that, we

   Proof. A great circle is the intersection of a sphere with a plane that passes through the sphere's center. Let's consider a sphere of radius R centered at the origin. We can parameterize any great circle on this sphere by choosing two orthogonal unit vectors, $\mathbf{a}$ and $\mathbf{b}$, that lie in the plane of the circle. A unit-speed parametrization is then given by:
   $$\gamma(s) = \cos(s/R)\mathbf{a} + R\sin(s/R)\mathbf{b}$$

where s is the arc length parameter. Differentiating with respect to s gives the velocity vector:

$$\dot{\gamma}(s) = -\sin(s/R)\mathbf{a} + \cos(s/R)\mathbf{b}$$

Differentiating a second time gives the acceleration vector:

$$\ddot{\gamma}(s) = -\frac{1}{R}\cos(s/R)\mathbf{a} - \frac{1}{R}\sin(s/R)\mathbf{b} = -\frac{1}{R^2}\gamma(s)$$

For a sphere centered at the origin, the unit normal vector $\mathbf{N}$ at a point $\mathbf{p}$ on its surface is parallel to the position vector of that point, i.e., $\mathbf{N}$ is parallel to $\mathbf{p} = \gamma(s)$. Our calculated acceleration vector, $\ddot{\gamma}(s) = -\frac{1}{R^2}\gamma(s)$, is a scalar multiple of the position vector $\gamma(s)$. Therefore, the acceleration is parallel to the surface normal at every point along the curve. By Definition 1.1, the great circle is a geodesic. $\qquad\square$

Remark. A perceptive reader may ask, "Since a straight line segment on any given surface is a geodesic, does that mean a straight line on a sphere is also a geodesic?". We would say that straight line segments do not even *exist* on a sphere! A little thinking may induce an image of a straight line segment protruding out of a sphere! It cannot be on the sphere, however if we orthogonally project it onto the sphere, then it will be great circle assuming that it passes through the *center* of the sphere in the first place.

3   Existence and Uniqueness of Geodesics

As the name of this section suggests or implies, there is a "differential equations" flavor to it and the reader is right to think so. Beginning with the uniqueness, as with most uniqueness theorems in the theory of ODEs, we are well-equipped to grapple with

While simple examples of geodesics can be identified via the definition, finding all the geodesics on an arbitrary surface *requires* a more powerful, analytical approach. This is achieved by solving a system of second-order differential equations known as the geodesic equations. These equations are derived from the fundamental definition of a geodesic and are expressed in terms of the coefficients of the surface's first fundamental form.

Theorem 3.1 (Geodesic Equations). A curve $\gamma(t) = \sigma(u(t), v(t))$ on a surface patch σ is a geodesic if and only if its component functions $u(t)$ and $v(t)$ satisfy the following pair of differential equations:

$$\frac{d}{dt}(E\dot{u} + F\dot{v}) = \frac{1}{2}(E_u\dot{u}^2 + 2F_u\dot{u}\dot{v} + G_u\dot{v}^2)$$
$$\frac{d}{dt}(F\dot{u} + G\dot{v}) = \frac{1}{2}(E_v\dot{u}^2 + 2F_v\dot{u}\dot{v} + G_v\dot{v}^2)$$

where E, F, G are the coefficients of the first fundamental form of σ, and subscripts denote partial derivatives (e.g., $E_u = \frac{\partial E}{\partial u}$).

Proof. This proof demonstrates the equivalence between the geometric definition of a geodesic and the analytical *geodesic equations* (9.2).

The core strategy is to take the geometric definition of a geodesic and show that it simplifies into the two differential equations presented in the theorem. The proof concludes by stating that a "similar" derivation starting from the second condition, $\ddot{\gamma} \cdot \sigma_v = 0$, yields the second geodesic equation. For full detail, refer to [7][Theorem 9.2.1, p. 220]

□

These equations are generally non-linear and difficult to solve analytically. However, for surfaces with a high degree of symmetry, such as a circular cylinder, a complete solution can be found.

3.1   Examples (and sanity check):

1. Having developed the theorem above, we can verify it on examples we have already known to be geodesics. Beginning with:

   (a) Straight lines. By using the coefficients of its First Fundamental Form, $E = G = 1$ and $F = 0$, hence we have from the two geodesic equations $\ddot{u} = \ddot{v} = 0$ and only straight lines satisfy this condition (why?).

   (b) Great circles. Similarly,

2. Geodesics on a Circular Cylinder. Let's find the geodesics on a circular cylinder of radius a using the standard parametrization $\sigma(u, v) = (a \cos u, a \sin u, v)$.

First, we compute the coefficients of the first fundamental form. The tangent vectors are:

$$\sigma_u = (-a \sin u, a \cos u, 0)$$
$$\sigma_v = (0, 0, 1)$$

The coefficients are then:

$$E = \sigma_u \cdot \sigma_u = a^2 \sin^2 u + a^2 \cos^2 u = a^2$$
$$F = \sigma_u \cdot \sigma_v = 0$$
$$G = \sigma_v \cdot \sigma_v = 1$$

Notice that E, F, and G are all constants. This means all of their partial derivatives ($E_u, E_v, F_u, \dots$) are zero. Substituting these into the geodesic equations gives a significant simplification:

$$\frac{d}{dt}(a^2 \dot{u} + 0\dot{v}) = 0 \quad \Longrightarrow \quad \frac{d}{dt}(a^2 \dot{u}) = 0$$
$$\frac{d}{dt}(0\dot{u} + 1\dot{v}) = 0 \quad \Longrightarrow \quad \frac{d}{dt}(\dot{v}) = 0$$

Integrating these two equations with respect to t yields:

$$a^2 \dot{u} = c_1$$
$$\dot{v} = c_2$$

where c_1 and c_2 are constants. From this, we can find the relationship between v and u along the geodesic curve using the chain rule:

$$\frac{dv}{du} = \frac{\dot{v}}{\dot{u}} = \frac{c_2}{c_1/a^2} = D_1 \quad \text{(a constant)}$$

Integrating this simple differential equation gives the final result:

$$v = D_1 u + D_2$$

This equation describes all possible geodesics on the cylinder:

- If $c_1 = 0$, then $\dot{u} = 0$ and u is constant. These are the vertical straight-line rulings on the cylinder.

- If $c_2 = 0$, then $\dot{v} = 0$ and v is constant. These are the horizontal circular cross-sections.

- If neither are zero, the path is a **circular helix**, which spirals around the cylinder at a constant angle to the vertical rulings.

TODO: Add images of the 3 families of geodesics

3.2   Examples: Geodesics on surfaces of revolution

Possibly, the simplest non-trivial example of geodesics are the ones on a torus
[4][Widget]. The surface of a torus with major radius a and minor radius b (where
$a > b$) can be parameterized by $u, v \in [0, 2\pi]$. The parameter u runs around the
circumference of the tube, while v runs around the main axis of the torus.

1. Parameterization

The parameterization $\vec{r}(u, v) \rightarrow \mathbb{R}^3$ is given by:

$$x = (a + b \sin u) \cos v \tag{1}$$
$$y = (a + b \sin u) \sin v \tag{2}$$
$$z = b \cos u \tag{3}$$

where $0 \leq u \leq 2\pi$ and $0 \leq v \leq 2\pi$.

2. Squared Differential Length

The squared differential length ds^2 (which is used to measure distances on the sur-
face) is given by:

$$ds^2 = b^2 du^2 + (a + b \sin u)^2 dv^2 \tag{4}$$

3. Geodesic Equations

The geodesic equations, which describe the shortest paths on the surface, are given
by the following system of differential equations:

$$\ddot{u} = \frac{1}{b}(a + b \sin u) \cos u \dot{v}^2 \tag{5}$$

$$\ddot{v} = -\frac{2b \cos u}{a + b \sin u} \dot{u}\dot{v} \tag{6}$$

As noted in the source, these equations do not have singularities as long as $a > b > 0$,
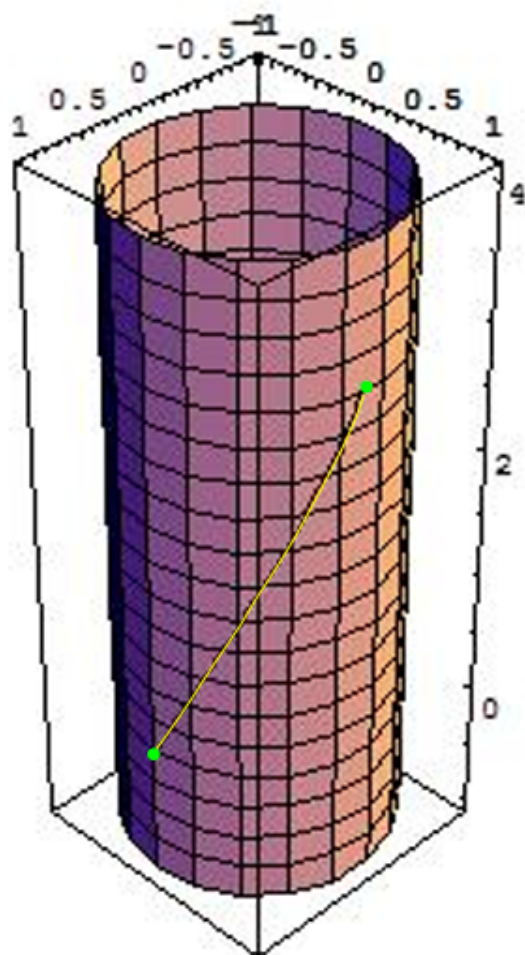since the term $(a + b \sin u)$ is always positive.

Figure 1: A straight line in the plane (left) is mapped by a local isometry to a geodesic helix on the cylinder (right). This illustrates the different types of geodesics. Reference to be added.

4   Geodesics as the Shortest Paths

Our journey began with the intuition that a geodesic is the straightest possible path on a curved surface. This idea is formally linked to the familiar concept of "shortest distance" by a fundamental theorem in differential geometry. We know from elementary (grade-school) geometry that on a plane, *the shortest path between 2 fixed points is the straight line segment*. However, has one encountered a rigorous proof?

Proof. Let $\gamma : I \subset_{open} \mathbb{R} \to \mathbb{R}^n (n \in \mathbb{N})$ be a curve with two fixed points $p, q \in \gamma(I)$. Then $p = \gamma(a)$ and $q = \gamma(b)$ for some $a < b \in I$. Let $u \in \mathbb{R}^n$ be a unit vector. We will prove it in 3 stages.

By the Fundamental Theorem of Calculus, the properties of integrals, and the Cauchy-Schwarz inequality, we have:

$$
\begin{aligned}
(\mathbf{q} - \mathbf{p}) \cdot \mathbf{u} &= (\gamma(b) - \gamma(a)) \cdot \mathbf{u} \\
&= \left(\int_a^b \dot{\gamma}(t) dt \right) \cdot \mathbf{u} \\
&= \int_a^b (\dot{\gamma}(t) \cdot \mathbf{u}) dt \\
&\leq \int_a^b |\dot{\gamma}(t) \cdot \mathbf{u}| dt && (\text{Since } x \leq |x|) \\
&\leq \int_a^b \|\dot{\gamma}(t)\| \|\mathbf{u}\| dt && (\text{Cauchy-Schwarz}) \\
&= \int_a^b \|\dot{\gamma}(t)\| dt && (\text{Since } \|\mathbf{u}\| = 1) \\
&= L(\gamma)
\end{aligned}
$$

We have now shown that $(\mathbf{q} - \mathbf{p}) \cdot \mathbf{u} \leq L(\gamma)$ for any unit vector $\mathbf{u}$.

Now, we choose the specific unit vector $\mathbf{u} = \frac{\mathbf{q} - \mathbf{p}}{\|\mathbf{q} - \mathbf{p}\|}$. (If $\mathbf{p} = \mathbf{q}$, the inequality $0 \leq L(\gamma)$ is trivial).

Substituting this $\mathbf{u}$ into our inequality:

$$
(\mathbf{q} - \mathbf{p}) \cdot \left(\frac{\mathbf{q} - \mathbf{p}}{\|\mathbf{q} - \mathbf{p}\|} \right) \leq L(\gamma)
$$

$$
\frac{(\mathbf{q} - \mathbf{p}) \cdot (\mathbf{q} - \mathbf{p})}{\|\mathbf{q} - \mathbf{p}\|} \leq L(\gamma)
$$

$$
\frac{\|\mathbf{q} - \mathbf{p}\|^2}{\|\mathbf{q} - \mathbf{p}\|} \leq L(\gamma)
$$

This simplifies to the final result:

$$
\|\mathbf{q} - \mathbf{p}\| \leq L(\gamma)
$$

$\square$

Although less trivially, great circles on a sphere are the shortest curve connecting non-antipodal points!

Proof. Refer to [7, p.155, Proposition 6.5.1].   □

4.1   The Converse: Are All Geodesics Shortest Paths?

An important subtlety is that the converse of the theorem is not always true: a curve can be a geodesic without being the shortest path between two points. An example is found on a sphere. Given two points P and Q on a sphere (that are not diametrically opposite), they lie on a unique great circle. This great circle is a geodesic, but it is divided by the points into two arcs: a short arc and a long arc. Both arcs satisfy the definition of a geodesic, as their acceleration vectors always point toward the center of the sphere. However, only the **short arc** is the true shortest path between P and Q. The long arc is a geodesic that is a local, but not global, minimum of length.
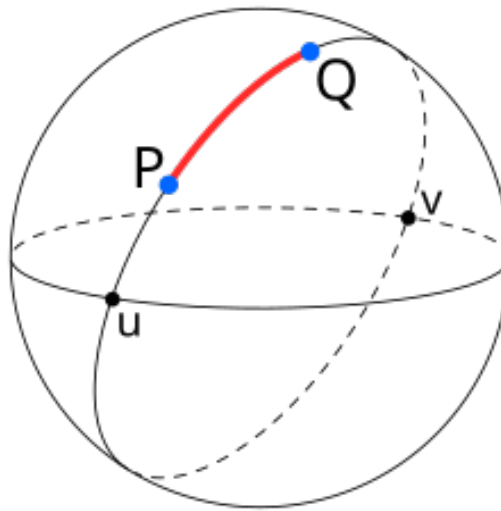

Figure 2: Both the short arc and the long arc between points P and Q on a sphere are geodesics, but only the short arc is the shortest path. ([2, Great circles])

4.2   Existence of Shortest Paths

Furthermore, a shortest path between two points on a surface may not always exist. Consider the surface S formed by the xy-plane with the origin removed. To travel from $P = (-1, 0)$ to $Q = (1, 0)$, the intuitive shortest path is the straight line segment connecting them. However, this path is not on the surface because it passes through the origin. Any curve on the surface S that connects P and Q must go around the origin. For any such path, it's always possible to find a slightly shorter one by moving closer to the origin. Therefore, there is no single "shortest" path, only a greatest lower bound (an infimum) on the set of possible path lengths between P and Q. [7, p. 239 - 241]

5   Computational Methods for Geodesics

The analytical approach of solving the geodesic equations works well for highly symmetric surfaces like the plane, sphere, or cylinder. However, for the free-form parametric surfaces commonly found in engineering design (CAD/CAM), robotics, and computer graphics, these equations become far too complex to solve by analytical methods. To find geodesics on these surfaces, we must resort to computational methods.

5.1   The Boundary Value Problem

When seeking the shortest path between two fixed points, say P_A and P_B, on a surface, the problem is no longer about just following an initial direction. Instead, we have a known start and a known end. The task is to find the path between them that satisfies the geodesic ordinary differential equations (ODEs). This is a classic numerical analysis problem known as a **two-point boundary value problem (BVP)**.

While several techniques exist for solving BVPs, a common and robust approach for geodesics is the **relaxation method**, which is based on a finite difference discretization.

5.2   The Relaxation Method

The core idea of the relaxation method is to transform the continuous differential equation into a system of algebraic equations that can be solved by a compute. The process involves three main steps [6]:

1. **Discretize the path:** The unknown geodesic curve is approximated by a finite sequence of m points, $Y_1, Y_2, \ldots, Y_m$, where $Y_1 = P_A$ and $Y_m = P_B$. The initial guess for this path can be as simple as a straight line between the points in the surface's parameter domain.

2. **Approximate the derivatives:** The derivatives in the geodesic equations (e.g., $\frac{du}{ds}, \frac{d^2u}{ds^2}$) are replaced by **finite difference** approximations. For instance, a first derivative $\frac{dy}{ds}$ at point k can be approximated by $\frac{Y_{k+1}-Y_{k-1}}{s_{k+1}-s_{k-1}}$. This converts the system of ODEs into a large, non-linear system of algebraic equations involving the unknown coordinates of the points Y_k.

3. **Iterate to a solution:** This system of algebraic equations is then solved iteratively. Starting with the initial guess, a numerical solver like Newton's method is used to adjust the positions of the intermediate points $Y_2, \ldots, Y_{m-1}$. Each iteration brings the path closer to satisfying the finite difference equations, effectively "relaxing" the initial guess into the final, true geodesic path.

This method is powerful and reliable, but it's not the only way to compute geodesics. Other approaches include the "heat method", which relates geodesic distance to the physics of heat diffusion on a surface, and various algorithms from computational geometry designed for specific domains like simple polygons. These advanced methods are essential for modern applications in shape analysis and robotics.

References

[1] Aristotle. *Physics*. Translated by R. P. Hardie and R. K. Gaye. The Internet Classics Archive, MIT. (Accessed: 3 November 2025).

[2] Wikipedia contributors. "Geodesic." *Wikipedia, The Free Encyclopedia*. (Accessed: 3 November 2025).

[3] Hagen, Hans, and Hotz, Ingrid. "Variational Modeling Methods for Visualization." In: *Visualization Handbook*, edited by Charles D. Hansen and Chris R. Johnson, Springer, 2004, pp. 381–392.

[4] Masson, Paul. *Visualizing Geodesics on Surfaces*. Analytic Physics. Link to webpage (Accessed: 3 November 2025).

[5] Cleve, Jonas, and Mulzer, Wolfgang. "An Experimental Study of Algorithms for Geodesic Shortest Paths in the Constant-Workspace Model." Technical Report, Institut für Informatik, Freie Universität Berlin.

[6] Computation of Shortest Paths on Free-Form Parametric Surfaces

[7] Pressley, Andrew. *Elementary Differential Geometry*. Springer-Verlag London, 2001.

[8] Tawfeek, Andrew. "An Introduction to Geodesics: The Shortest Distance Between Two Points." arXiv preprint, arXiv:2007.02864v2 [math.FA], 2020. `https://arxiv.org/abs/2007.02864v2`