

Geodesics*

(Com S 477/577 Notes)

Yan-Bin Jia

Oct 27, 2020

Geodesics are the curves in a surface that make turns just to stay on the surface and never move sideways. A bug living in the surface and following such a curve would perceive it to be straight. A geodesic is a generalization of the notion of a “straight line” from a plane to a surface, on which it represents in some sense the shortest path between two points. We will begin with a definition of geodesics, then present various method for finding geodesics on surfaces, and later reveal their relationships to shortest paths.

The term *geodesic* comes from the science of geodesy, which is concerned with measurements of the earth’s surface [1, p. 163]. F. W. Bessel (1784–1846) was involved with determining the shape of the earth as an ellipsoid of rotation. C. G. Jacobi (1804–1851) studied the “shortest curves” on an ellipsoid of rotation which he referred to as “geodesic curves”. The term “shortest curves” had earlier been used by Johannes Bernoulli (1667–1748) and Carl-Friedrich Gauss (1777–1855).

1 Definition

A curve $\gamma(t)$ on a surface \mathcal{S} is called a *geodesic* if at every point $\gamma(t)$ the acceleration $\ddot{\gamma}(t)$ is either zero or parallel to its unit normal $\hat{\mathbf{n}}$.

EXAMPLE 1. Strictly following the definition, a straight line $\gamma(t) = \mathbf{a}t + \mathbf{b}$ is a geodesic since $\ddot{\gamma}(t) = \mathbf{0}$. However, a reparametrization of the line as $\delta(u) = \mathbf{a} \tan u + \mathbf{b}$ over $u \in (-\frac{\pi}{2}, \frac{\pi}{2})$ is not a geodesic since $\ddot{\delta}(u) = (\mathbf{a}/\cos^2 u)' = -(2 \sin u / \cos^3 u)\mathbf{a} \neq 0$ unless $u = 0$. The fact is that every line can be parametrized so that it is a geodesic.

Here is an interesting interpretation of geodesics from mechanics. Suppose a particle is moving on the surface under a force perpendicular to the surface to maintain contact. Its trajectory would be a geodesic, because Newton’s second law states that the particle’s acceleration $\ddot{\gamma}$ is parallel to the force, hence always perpendicular to the surface.

Proposition 1 *A geodesic $\gamma(t)$ on a surface S has constant speed.*

Proof We have

$$\frac{d}{dt} \|\dot{\gamma}\|^2 = \frac{d}{dt} (\dot{\gamma} \cdot \dot{\gamma}) = 2\ddot{\gamma} \cdot \dot{\gamma}.$$

*The material is adapted from the book *Elementary Differential Geometry* by Andrew Pressley, Springer-Verlag, 2001.

Since γ is geodesic, $\ddot{\gamma}$ is perpendicular to the tangent plane which contains $\dot{\gamma}$. Hence $\ddot{\gamma} \cdot \dot{\gamma} = 0$. Subsequently, $d\|\dot{\gamma}\|^2/dt = 0$. Therefore, the speed $\|\dot{\gamma}\|$ is constant. \square

Proposition 1 suggests that the unit-speed parametrization of a geodesic is still a geodesic, since the acceleration is just scaled by a non-zero constant factor. So we can always consider unit-speed geodesics only if needed.

Proposition 2 *A curve on a surface is a geodesic if and only if its geodesic curvature is zero everywhere.*

Proof Let γ be a unit-speed geodesic in a patch σ of the surface, and N the unit normal of σ . The geodesic curvature is

$$\kappa_g = \ddot{\gamma} \cdot (N \times \dot{\gamma}).$$

If $\ddot{\gamma}$ is zero, then the above implies that $\kappa_g = 0$. Otherwise, by definition $\ddot{\gamma}$ is parallel to N . It is thus perpendicular to $N \times \dot{\gamma}$, thereby $\kappa_g = 0$.

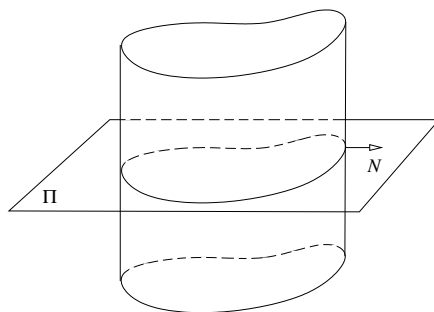
Conversely, suppose $\kappa_g = 0$. Then $\ddot{\gamma} \perp N \times \dot{\gamma}$ if $\ddot{\gamma} \neq 0$. Meanwhile, from $\dot{\gamma} \times \dot{\gamma} = 1$ we obtain that $\ddot{\gamma} \perp \dot{\gamma}$. Therefore, $\ddot{\gamma}$ is parallel to $\dot{\gamma} \times (N \times \dot{\gamma})$; in other words, it is parallel to N . \square

By Proposition 2, we can claim that all straight lines are geodesics. Other simple geodesics include the rulings of any ruled surface, such as the generators of a (generalized) cylinder or cone.

Suppose a curve on a surface is its intersection with a plane that happens to be perpendicular to the tangent plane at every point on the curve. Since the curve lies in a normal plane, its curvature κ equals the normal curvature κ_n everywhere. From $\kappa = \sqrt{\kappa_n^2 + \kappa_g^2}$, we infer that $\kappa_g = 0$ everywhere.

EXAMPLE 2. A great circle on a sphere is its intersection with a plane Π passing through the center O of the sphere. Every point \mathbf{p} on the great circle defines a vector with $O\mathbf{p}$ which is perpendicular to the tangent plane at \mathbf{p} . Thus, the great circle is a geodesic.

EXAMPLE 3. The intersection of a generalized cylinder with a plane Π perpendicular to the rulings of the cylinder is a geodesic. Clearly, the unit normal at such an intersection point is perpendicular to the rulings, and hence contained in Π . So, Π is perpendicular to the tangent plane at the point.



2 Geodesic Equations

We have seen two examples of geodesics. To determine all the geodesics on a given surface, we need to solve differential equations stated in the following theorem.

Theorem 3 A curve γ on a surface \mathcal{S} is a geodesic if and only if for any part $\gamma(t) = \sigma(u(t), v(t))$ contained in a surface patch σ of \mathcal{S} , the following two equations are satisfied:

$$\frac{d}{dt}(E\dot{u} + F\dot{v}) = \frac{1}{2}(E_u\dot{u}^2 + 2F_u\dot{u}\dot{v} + G_u\dot{v}^2), \quad (1)$$

$$\frac{d}{dt}(F\dot{u} + G\dot{v}) = \frac{1}{2}(E_v\dot{u}^2 + 2F_v\dot{u}\dot{v} + G_v\dot{v}^2), \quad (2)$$

where $Edu^2 + 2Fdudv + Gdv^2$ is the first fundamental form of σ .

Proof The tangent plane is spanned by σ_u and σ_v . By definition the curve γ is a geodesic if and only if $\ddot{\gamma} \cdot \sigma_u = \ddot{\gamma} \cdot \sigma_v = 0$. Since $\dot{\gamma} = \dot{u}\sigma_u + \dot{v}\sigma_v$, $\ddot{\gamma} \cdot \sigma_u = 0$ becomes

$$\left(\frac{d}{dt}(\dot{u}\sigma_u + \dot{v}\sigma_v) \right) \cdot \sigma_u = 0.$$

We rewrite the left hand side of the above equation:

$$\begin{aligned} \left(\frac{d}{dt}(\dot{u}\sigma_u + \dot{v}\sigma_v) \right) \cdot \sigma_u &= \frac{d}{dt}((\dot{u}\sigma_u + \dot{v}\sigma_v) \cdot \sigma_u) - (\dot{u}\sigma_u + \dot{v}\sigma_v) \cdot \frac{d\sigma_u}{dt} \\ &= \frac{d}{dt}(E\dot{u} + F\dot{v}) - (\dot{u}\sigma_u + \dot{v}\sigma_v) \cdot (\dot{u}\sigma_{uu} + \dot{v}\sigma_{uv}) \\ &= \frac{d}{dt}(E\dot{u} + F\dot{v}) - \left(\dot{u}^2(\sigma_u \cdot \sigma_{uu}) + \dot{u}\dot{v}(\sigma_u \cdot \sigma_{uv} + \sigma_v \cdot \sigma_{uu}) \right. \\ &\quad \left. + \dot{v}^2(\sigma_v \cdot \sigma_{uv}) \right). \end{aligned} \quad (3)$$

We have that

$$\begin{aligned} \sigma_u \cdot \sigma_{uu} &= \frac{1}{2} \frac{\partial}{\partial u} (\sigma_u \cdot \sigma_u) = \frac{1}{2} E_u, \\ \sigma_v \cdot \sigma_{uv} &= \frac{1}{2} G_u, \\ \sigma_u \cdot \sigma_{uv} + \sigma_v \cdot \sigma_{uu} &= F_u. \end{aligned}$$

Substituting them into (3), we obtain

$$\left(\frac{d}{dt}(\dot{u}\sigma_u + \dot{v}\sigma_v) \right) \cdot \sigma_u = \frac{d}{dt}(E\dot{u} + F\dot{v}) - \frac{1}{2}(E_u\dot{u}^2 + 2F_u\dot{u}\dot{v} + G_u\dot{v}^2).$$

This establishes the first differential equation (1). Similarly, equation (2) can be established from

$$\left(\frac{d}{dt}(\dot{u}\sigma_u + \dot{v}\sigma_v) \right) \cdot \sigma_v = 0.$$

□

The two equations in Theorem 3 are called the *geodesic equations*. They are nonlinear and solvable analytically on rare occasions only. The following is an example where a close-form solution can be found.

EXAMPLE 4. Let us find the geodesics on the unit sphere S^2 by solving equations (1) and (2). Consider the patch under the parametrization

$$\sigma(\theta, \phi) = (\cos \theta \cos \phi, \cos \theta \sin \phi, \sin \theta).$$

In Example 1 of the notes titled “Surface Curves and Fundamental Forms”, the first fundamental form is found to be $d\theta^2 + \cos^2 \theta d\phi^2$, with $E = 1, F = 0$, and $G = \cos^2 \theta$. We restrict to unit-speed curves $\gamma(t) = \sigma(\theta(t), \phi(t))$ so that

$$E\dot{\theta}^2 + 2F\dot{\theta}\dot{\phi} + G\dot{\phi}^2 = \dot{\theta}^2 + \dot{\phi}^2 \cos^2 \theta = 1. \quad (4)$$

If γ is a geodesic, equation (2) is satisfied. Here, the equation reduces to

$$\frac{d}{dt}(\dot{\phi} \cos^2 \theta) = 0, \quad (5)$$

is satisfied. Hence

$$\dot{\phi} \cos^2 \theta = C, \quad (6)$$

for some constant C . There are two cases.

- $C = 0$ Then $\dot{\phi} = 0$. In this case, ϕ is constant and γ is part of a meridian.
- $C \neq 0$ Substituting (6) into the unit-speed condition (4), we have

$$\dot{\theta}^2 = 1 - \frac{C^2}{\cos^2 \theta}.$$

Combining the above with (6), along the geodesic it holds that

$$\left(\frac{d\phi}{d\theta}\right)^2 = \frac{\dot{\phi}^2}{\dot{\theta}^2} = \frac{1}{\cos^2 \theta \left(\frac{\cos^2 \theta}{C^2} - 1\right)}.$$

Integrate the derivative $d\phi/d\theta$:

$$\phi - \phi_0 = \pm \int \frac{d\theta}{\cos \theta \sqrt{\frac{\cos^2 \theta}{C^2} - 1}},$$

where ϕ_0 is a constant. The substitution $u = \tan \theta$ yields

$$\phi - \phi_0 = \pm \int \frac{du}{\sqrt{C^{-2} - 1 - u^2}} = \sin^{-1} \left(\frac{u}{\sqrt{C^{-2} - 1}} \right),$$

which leads to

$$\tan \theta = \pm \sqrt{C^{-2} - 1} \cdot \sin(\phi - \phi_0).$$

Multiply both sides of the above equation with $\cos \theta$:

$$\sin \theta = \pm \sqrt{C^{-2} - 1} (\cos \phi_0 \cos \theta \sin \phi - \sin \phi_0 \cos \theta \cos \phi).$$

Since $\sigma(\theta, \phi) = (xy, z)$, we have

$$z = \mp (\sin \phi_0 \sqrt{C^{-2} - 1})x \pm (\cos \phi_0 \sqrt{C^{-2} - 1})y.$$

Clearly, $z = 0$ when $x = y = 0$. Therefore, γ is contained in the intersection of S^2 with a plane through the center of the sphere. Hence it is part of a great circle.

In both cases, γ is part of a great circle.

EXAMPLE 5. We now find the geodesics of the circular cylinder

$$\sigma = (a \cos \phi, a \sin \phi, z).$$

First, we obtain the coefficients of the first fundamental form:

$$E = a^2, \quad F = 0, \quad G = 1.$$

Substitute these coefficients into (1) and (2):

$$\begin{aligned} \frac{d}{dt} \left(\frac{a^2 \dot{\phi}}{\sqrt{a^2 \dot{\phi}^2 + \dot{z}^2}} \right) &= 0, \\ \frac{d}{dt} \left(\frac{\dot{z}}{\sqrt{a^2 \dot{\phi}^2 + \dot{z}^2}} \right) &= 0, \end{aligned}$$

Integrate the two equations above:

$$\begin{aligned} \frac{a^2 \dot{\phi}}{\sqrt{a^2 \dot{\phi}^2 + \dot{z}^2}} &= C_1, \\ \frac{\dot{z}}{\sqrt{a^2 \dot{\phi}^2 + \dot{z}^2}} &= C_2. \end{aligned}$$

Dividing the last equation by the one before, we obtain

$$\frac{dz}{d\phi} = D_1,$$

which has the solution

$$z = D_1 \phi + D_2.$$

The geodesic between two points on a cylinder thus is a helix lying on the cylinder. Given two points $\sigma(\phi_0, z_0)$ and $\sigma(\phi_1, z_1)$, the helix is described as

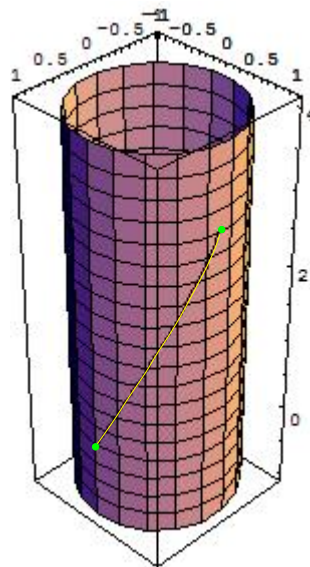
$$\alpha(\phi) = \left(a \cos \phi, a \sin \phi, \frac{z_0 - z_1}{\phi_0 - \phi_1} \phi + \frac{\phi_0 z_1 - \phi_1 z_0}{\phi_0 - \phi_1} \right).$$

The figure on the top right shows a geodesic $(\cos \phi, \sin \phi, 2\phi)$ connecting two points with $\phi_0 = 0$ and $\phi_1 = \frac{3}{2}$ on a cylinder $(\cos \phi, \sin \phi, z)$ which is plotted over the subdomain $[0, 2\pi] \times [-1, 4]$.

The following theorem states that a unique geodesic exists on a surface that passes through any of its point in any given tangent direction.¹

Theorem 4 *Let \mathbf{p} be a point on a surface S , and $\hat{\mathbf{t}}$ a unit tangent vector at \mathbf{p} . There exists a unique unit-speed geodesic γ on S which passes through \mathbf{p} with velocity $\gamma' = \hat{\mathbf{t}}$.*

A trivial example would be straight lines in a plane. At any point, in any direction, there is a unique straight line through the point. Theorem 4 says that there are no other geodesics. On a sphere, the great circles are the only geodesics. This is because, in any tangent direction at a point



¹For a proof of the theorem, we refer to [3, p. 178].

on the sphere, there is a great circle through the point which is the intersection of the sphere with the plane determined by the tangent line in that direction and the sphere's center.

Theorem 4 provides four initial values needed to solve the geodesic equations (1) and (2) for u and v as functions of the curve parameter. Suppose we need to construct a geodesic passing through a point \mathbf{p} on a surface patch $\boldsymbol{\sigma}(u, v)$ in the unit tangent direction $\hat{\mathbf{t}}$, with $\mathbf{p} = \boldsymbol{\sigma}(u_0, v_0)$. We consider a unit-speed parametrization $\boldsymbol{\gamma}(t) = \boldsymbol{\sigma}(u(t), v(t))$ such that

$$\begin{aligned} u(0) &= u_0, \\ v(0) &= v_0. \end{aligned}$$

The unit tangent at \mathbf{p} is $\boldsymbol{\sigma}_u(u_0, v_0)\dot{u}(0) + \boldsymbol{\sigma}_v(u_0, v_0)\dot{v}(0)$. Because $\boldsymbol{\sigma}_u$ and $\boldsymbol{\sigma}_v$ span the tangent plane at \mathbf{p} , the values of the two derivatives $\dot{u}(0)$ and $\dot{v}(0)$ can be determined from the equation

$$\hat{\mathbf{t}} = \boldsymbol{\sigma}_u(u_0, v_0)\dot{u}(0) + \boldsymbol{\sigma}_v(u_0, v_0)\dot{v}(0).$$

One of the three scalar equations in the above equation is redundant since $\hat{\mathbf{t}}$, $\boldsymbol{\sigma}_u(u_0, v_0)$, and $\boldsymbol{\sigma}_v(u_0, v_0)$ are co-planar. To eliminate the redundancy, we take the dot products of the equation with $\boldsymbol{\sigma}_u(u_0, v_0)$ and $\boldsymbol{\sigma}_v(u_0, v_0)$ separately, obtaining

$$\begin{pmatrix} E & F \\ F & G \end{pmatrix} \begin{pmatrix} \dot{u}(0) \\ \dot{v}(0) \end{pmatrix} = \begin{pmatrix} \hat{\mathbf{t}} \cdot \boldsymbol{\sigma}_u(u_0, v_0) \\ \hat{\mathbf{t}} \cdot \boldsymbol{\sigma}_v(u_0, v_0) \end{pmatrix}.$$

This gives us

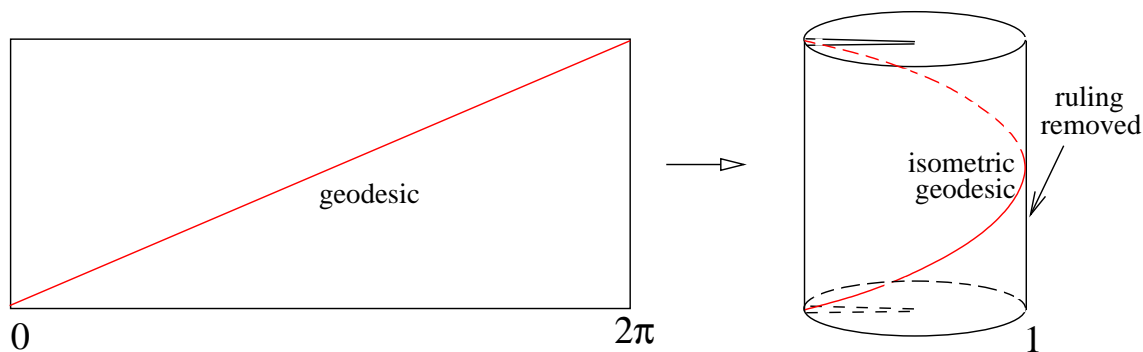
$$\begin{pmatrix} \dot{u}(0) \\ \dot{v}(0) \end{pmatrix} = \mathcal{F}_1^{-1} \begin{pmatrix} \hat{\mathbf{t}} \cdot \boldsymbol{\sigma}_u(u_0, v_0) \\ \hat{\mathbf{t}} \cdot \boldsymbol{\sigma}_v(u_0, v_0) \end{pmatrix},$$

where the first fundamental form matrix \mathcal{F}_1 is evaluated at (u_0, v_0) . The four initial values $u(0), v(0), \dot{u}(0), \dot{v}(0)$ uniquely specify a solution curve to the geodesic equations. In fact, this is the idea behind the proof of Theorem 4.

3 Preservation of Geodesics Under Isometry

Let S_1 and S_2 be two surfaces. A map $\mathbf{f} : S_1 \rightarrow S_2$ is called a *diffeomorphism* if it is smooth and bijective and its inverse map $\mathbf{f}^{-1} : S_2 \rightarrow S_1$ is also smooth. A diffeomorphism \mathbf{f} is called an *isometry* if it maps curves in S_1 to curves of the same length in S_2 . If an isometry exists, then S_1 and S_2 are *isometric*. A sufficient and necessary condition for \mathbf{f} to be an isometry is *any surface patch $\boldsymbol{\sigma}_1$ of S_1 and its image patch $\mathbf{f} \circ \boldsymbol{\sigma}_1$ of S_2 have the same first fundamental form*. The proof, given in [3, p. 102], relies on the fact that the length of a curve $\boldsymbol{\gamma}(t) = \boldsymbol{\sigma}(u(t), v(t))$ on a surface patch is $\int \sqrt{E\dot{u}^2 + 2F\dot{u}\dot{v} + G\dot{v}^2} dt$.

EXAMPLE 6. Let S_1 be the infinite strip in the xy -plane given by $0 < x < 2\pi$, and S_2 be the circular cylinder $x^2 + y^2 = 1$ with the ruling $x = 1, y = 0$ removed. Then S_1 is covered by a single patch $\boldsymbol{\sigma}_1(u, v) = (u, v, 0)$, and S_2 by the patch $\boldsymbol{\sigma}_2(u, v) = (\cos u, \sin u, v)$, where $0 < u < 2\pi$ for both cases. Consider the map $\mathbf{f} : (u, v, 0) \mapsto (\cos u, \sin u, v)$ from S_1 to S_2 . We can verify that both patches have the same first fundamental form $du^2 + dv^2$. Therefore \mathbf{f} is an isometry, and S_1 and S_2 are isometric.



Theorem 5 An isometry $f : S_1 \rightarrow S_2$ maps the geodesics of S_1 to the geodesics of S_2 .

Proof Let $\gamma(t)$ be a geodesic in S_1 . Suppose a part of γ lies in the patch $\sigma(u, v)$ in S_1 so that $\gamma(t) = \sigma(u(t), v(t))$. Then, u and v satisfy the geodesic equations (1) and (2). Now, consider the patch $f \circ \sigma$ of S_2 as the image of σ under the isometry. Because both patches have the same first fundamental form, the curve $f(\sigma(u(t), v(t)))$ on S_2 must also satisfy the geodesic equations. Thus $f \circ \gamma$ is a geodesic on S_2 by Theorem 3. \square

EXAMPLE 7. We know that on the circular cylinder $x^2 + y^2 = 1$, geodesics include cross sections where planes parallel to the xy -plane intersect the cylinder, as well as straight lines parallel to the z -axis. To find the missing geodesics, we make use of the isometry given in Example 5 which maps $(u, v, 0)$ to $(\cos u, \sin u, v)$. In the isometric strip in the xy -plane given by $0 < x < 2\pi$, any line not parallel to the y -axis has a parametrization $(u, mu + c)$. Its image on the cylinder is the curve

$$\gamma(u) = (\cos u, \sin u, mu + c).$$

This curve is a geodesic on the cylinder by Theorem 5. In the previous figure, a diagonal in the strip is drawn, along with its image geodesic curve on the isometric cylinder.

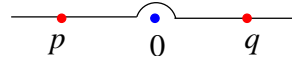
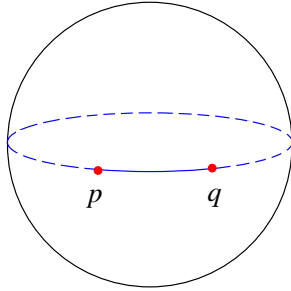
4 Geodesics vs. Shortest Paths

We all know that the shortest path between two points \mathbf{p} and \mathbf{q} in the plane is the line segment $\overline{\mathbf{p}\mathbf{q}}$. It is almost as well known that great circles are the shortest paths on a sphere. The following can be established.

Theorem 6 Let γ be a shortest path on a surface S connecting two points \mathbf{p} and \mathbf{q} . Then the part of γ contained in any surface patch σ of S must be a geodesic.

The converse of the statement in the theorem is not necessarily true, however. If γ is a geodesic on σ connecting \mathbf{p} and \mathbf{q} . Then γ need not be a shortest path between the two points. As an example, the great circle connecting two points \mathbf{p} and \mathbf{q} on a sphere is split into two circular arcs by the points. Both arcs are geodesics. Only the shorter one of the two is the shortest path joining \mathbf{p} and \mathbf{q} .

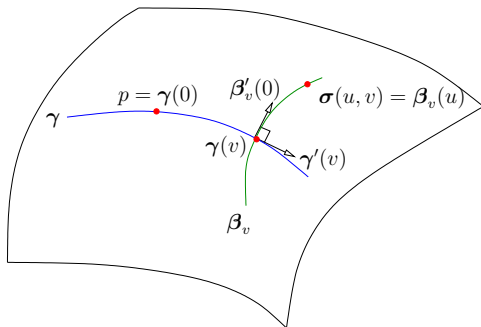
In general, a shortest path joining two points on a surface may not exist. For example, consider the surface P which is the xy -plane with its origin removed. There is no shortest path from the point $\mathbf{p} = (-1, 0)$ to the point $\mathbf{q} = (1, 0)$. Such a path would be the line segment, except it passes



through the origin and does not lie entirely on the surface. Any short path in P from p to q would walk in a straight line as long as possible, and then move around the origin, and continue in a straight line. It can always be improved on by moving a little closer to the origin before circling around it. Therefore, there is no shortest path connecting the two points.

If a surface S is a closed subset² of \mathbb{R}^3 , and if there is some path in S joining any two points p and q , then there always exists a shortest path joining the two points. For instance, a sphere is a closed subset of \mathbb{R}^3 , and the short great circle arc joining two points on the sphere is their shortest path. The surface P , the xy -plane with its origin removed, is not a closed subset of \mathbb{R}^3 because any open ball containing the origin must contain points of P (so the set of points not in P is not open).

5 Geodesic Coordinates



The existence of geodesics on a surface S gives a way to construct a useful atlas for S in practice. Let p be a point in S , and $\gamma(v)$ a unit-speed geodesic on S with $\gamma(0) = p$. For any value of v , there exists a unique unit-speed geodesic $\beta_v(u)$ such that $\beta_v(0) = \gamma(v)$ and $\beta'_v(0) \perp \gamma'(v)$. We define a patch $\sigma(u, v) = \beta_v(u)$ and call it a *geodesic patch*, and u and v *geodesic coordinates*.

EXAMPLE 8. Consider a point p on the equator of the unit sphere S^2 . Let the equator be parametrized with the longitude ϕ as $\gamma(\phi)$. Let β_ϕ be the meridian parametrized by the latitude θ and passing through the point on the equator with longitude ϕ . The corresponding geodesic patch is the usual one in latitude and longitude, for which the first fundamental form is $d\theta^2 + \cos^2 \theta d\phi^2$.

References

- [1] J. McCleary. *Geometry from a Differentiable Viewpoint*. Cambridge University Press, 1994.
- [2] B. O'Neill. *Elementary Differential Geometry*. Academic Press, Inc., 1966.
- [3] A. Pressley. *Elementary Differential Geometry*. Springer-Verlag London, 2001.

²In other words, the set $\mathbb{R}^3 - S$ is open.