9
Geodesics

Geodesics are the curves in a surface that a bug living in the surface would perceive to be straight. For example, the shortest path between two points in a surface is always a geodesic. We shall actually begin by giving a quite different definition of geodesics, since this definition is easier to work with. We give various methods of finding geodesics on surfaces, before finally making contact with the idea of shortest paths towards the end of the chapter.

9.1 Definition and basic properties

If we drive along a 'straight' road, we do not have to turn the wheel of our car to the right or left (this is what we mean by 'straight'!). However, the road is not, in fact, a straight line as the surface of the earth is, to a good approximation, a sphere and there can be no straight line on the surface of a sphere. If the road is represented by a curve $\boldsymbol{\gamma}$, its acceleration $\ddot{\boldsymbol{\gamma}}$ will be non-zero, but we perceive the curve as being straight because the *tangential component* of $\ddot{\boldsymbol{\gamma}}$ is zero, in other words because $\ddot{\boldsymbol{\gamma}}$ is perpendicular to the surface. This suggests

Definition 9.1.1

A curve $\boldsymbol{\gamma}$ on a surface $\mathcal{S}$ is called a *geodesic* if $\ddot{\boldsymbol{\gamma}}(t)$ is zero or perpendicular to the tangent plane of the surface at the point $\boldsymbol{\gamma}(t)$, i.e., parallel to its unit normal, for all values of the parameter t.

Andrew Pressley, *Elementary Differential Geometry: Second Edition*, Springer Undergraduate Mathematics Series, DOI 10.1007/978-1-84882-891-9_9,

Equivalently, γ is a geodesic if and only if its tangent vector $\dot{\gamma}$ is *parallel* along γ (see Section 7.4).

Note that this definition makes sense for any surface, orientable or not.

There is an interesting mechanical interpretation of geodesics: a particle moving on the surface, and subject to no forces except a force acting perpendicular to the surface that keeps the particle on the surface, would move along a geodesic. This is because Newton's second law of motion states that the force on the particle is parallel to its acceleration $\ddot{\gamma}$, which would therefore be perpendicular to the surface.

We begin our study of geodesics by noting that there is essentially no choice in their parametrization.

Proposition 9.1.2

Any geodesic has constant speed.

Proof

Let $\gamma(t)$ be a geodesic on a surface $\mathcal{S}$. Then, denoting d/dt by a dot,

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

Since γ is a geodesic, $\ddot{\gamma}$ is perpendicular to the tangent plane and is therefore perpendicular to the tangent vector $\dot{\gamma}$. So $\ddot{\gamma} \cdot \dot{\gamma} = 0$ and the last equation shows that $\parallel \dot{\gamma} \parallel$ is constant. □

It follows from this proposition that a unit-speed reparametrization of a geodesic γ is still a geodesic. For, if $\parallel \dot{\gamma} \parallel = \lambda$, then $\tilde{\gamma}(t) = \gamma(t/\lambda)$ is a unit-speed reparametrization of γ and $\frac{d^2\tilde{\gamma}}{dt^2} = \frac{1}{\lambda^2}\frac{d^2\gamma}{dt^2}$ is parallel to $\ddot{\gamma}$, and hence is also perpendicular to the surface. Thus, we can always restrict to unit-speed geodesics if we wish. In general, however, a reparametrization of a geodesic will not be a geodesic (see Exercise 9.1.2).

We observe next that there is an equivalent definition of a geodesic expressed in terms of the geodesic curvature κ_g (see Section 7.3). Of course, this is why κ_g is called the geodesic curvature!

Proposition 9.1.3

A unit-speed curve on a surface is a geodesic if and only if its geodesic curvature is zero everywhere.

Proof

Let $\boldsymbol{\gamma}$ be a unit-speed curve on the surface $\mathcal{S}$, and let $\mathbf{p} \in \mathcal{S}$. Let $\boldsymbol{\sigma}$ be a surface patch of $\mathcal{S}$ with $\mathbf{p}$ in its image, and let $\mathbf{N}$ be the standard unit normal of $\boldsymbol{\sigma}$, so that

$$\kappa_g = \ddot{\boldsymbol{\gamma}} \cdot (\mathbf{N} \times \dot{\boldsymbol{\gamma}}) \tag{9.1}$$

(changing $\boldsymbol{\sigma}$ may change the sign of $\mathbf{N}$, and hence that of κ_g, but that is not relevant to the present discussion). If $\ddot{\boldsymbol{\gamma}}$ is parallel to $\mathbf{N}$, it is obviously perpendicular to $\mathbf{N} \times \dot{\boldsymbol{\gamma}}$, so by Eq. 9.1, $\kappa_g = 0$.

Conversely, suppose that $\kappa_g = 0$. Then, $\ddot{\boldsymbol{\gamma}}$ is perpendicular to $\mathbf{N} \times \dot{\boldsymbol{\gamma}}$. But then, since $\dot{\boldsymbol{\gamma}}$, $\mathbf{N}$ and $\mathbf{N} \times \dot{\boldsymbol{\gamma}}$ are perpendicular unit vectors in $\mathbb{R}^3$ (see the discussion in Section 7.3), and since $\ddot{\boldsymbol{\gamma}}$ is perpendicular to $\dot{\boldsymbol{\gamma}}$, it follows that $\ddot{\boldsymbol{\gamma}}$ is parallel to $\mathbf{N}$. □

The following result gives the simplest examples of geodesics.

Proposition 9.1.4

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

By this, we mean that every straight line can be parametrized so that it is a geodesic. A similar remark applies to other geodesics we consider and whose parametrization is not specified (see Exercise 9.1.2).

Proof

This is obvious, for any straight line has a (constant speed) parametrization of the form

$$\boldsymbol{\gamma}(t) = \mathbf{a} + \mathbf{b}t,$$

where $\mathbf{a}$ and $\mathbf{b}$ are constant vectors, and clearly $\ddot{\boldsymbol{\gamma}} = \mathbf{0}$. □

Example 9.1.5

All straight lines in the plane are geodesics, as are the rulings of any ruled surface, such as those of a (generalized) cylinder or a (generalized) cone, or the straight lines on a hyperboloid of one sheet.

The next result is almost as simple:

Proposition 9.1.6

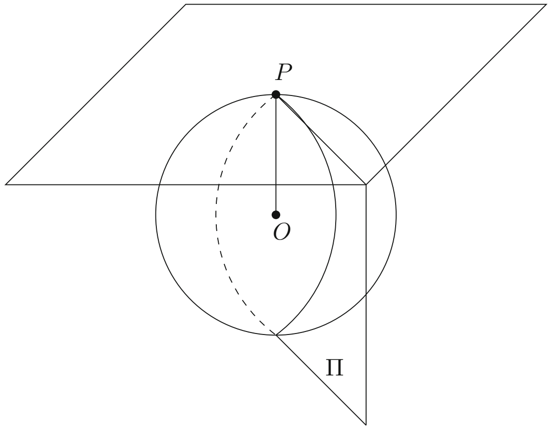
Any normal section of a surface is a geodesic.

Proof

Recall from Section 7.3 that a normal section of a surface $\mathcal{S}$ is the intersection $\mathcal{C}$ of $\mathcal{S}$ with a plane Π, such that Π is perpendicular to the surface at each point of $\mathcal{C}$. We showed in Corollary 7.3.4 that $\kappa_g = 0$ for a normal section, and so the result follows from Proposition 9.1.3. □

Example 9.1.7

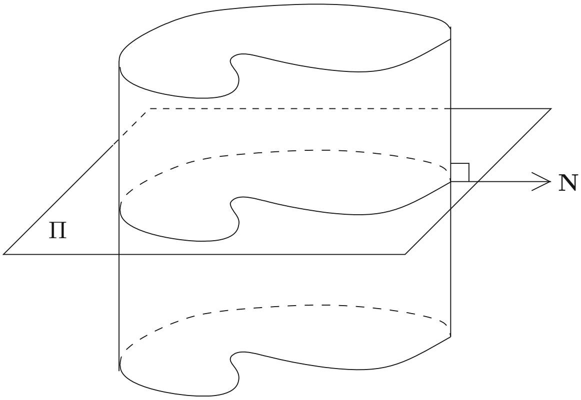
All great circles on a sphere are geodesics. For a great circle is the intersection of



the sphere with a plane Π passing through the centre O of the sphere, and so if P is a point of the great circle, the straight line through O and P lies in Π and is perpendicular to the tangent plane of the sphere at P. Hence, Π is perpendicular to the tangent plane at P.

Example 9.1.8

The intersection of a generalized cylinder with a plane Π perpendicular to the rulings of the cylinder is a geodesic. For it is clear that the unit normal $\mathbf{N}$ is perpendicular to the rulings. It follows that $\mathbf{N}$ is parallel to Π, and hence that Π is perpendicular to the tangent plane.



EXERCISES

9.1.1 Describe four different geodesics on the hyperboloid of one sheet

$$x^2 + y^2 - z^2 = 1$$

passing through the point $(1, 0, 0)$.

9.1.2 A (regular) curve $\boldsymbol{\gamma}$ with nowhere vanishing curvature on a surface $\mathcal{S}$ is called a *pre-geodesic* on $\mathcal{S}$ if some reparametrization of $\boldsymbol{\gamma}$ is a geodesic on $\mathcal{S}$ (recall that a reparametrization of a geodesic is not usually a geodesic). Show that:

(i) A curve $\boldsymbol{\gamma}$ is a pre-geodesic if and only if $\ddot{\boldsymbol{\gamma}} \cdot (\mathbf{N} \times \dot{\boldsymbol{\gamma}}) = 0$ everywhere on $\boldsymbol{\gamma}$ (in the notation of the proof of Proposition 9.1.3).

(ii) Any reparametrization of a pre-geodesic is a pre-geodesic.

(iii) Any constant speed reparametrization of a pre-geodesic is a geodesic.

(iv) A pre-geodesic is a geodesic if and only if it has constant speed.

9.1.3 Consider the tube of radius $a > 0$ around a unit-speed curve $\boldsymbol{\gamma}$ in $\mathbb{R}^3$ defined in Exercise 4.2.7:

$$\boldsymbol{\sigma}(s, \theta) = \boldsymbol{\gamma}(s) + a(\cos\theta\, \mathbf{n}(s) + \sin\theta\, \mathbf{b}(s)).$$

Show that the parameter curves on the tube obtained by fixing the value of s are circular geodesics on $\boldsymbol{\sigma}$.

9.1.4 Let $\boldsymbol{\gamma}(t)$ be a geodesic on an ellipsoid $\mathcal{S}$ (see Theorem 5.2.2(i)). Let $2R(t)$ be the length of the diameter of $\mathcal{S}$ parallel to $\dot{\boldsymbol{\gamma}}(t)$, and let $S(t)$ be the distance from the centre of $\mathcal{S}$ to the tangent plane $T_{\boldsymbol{\gamma}(t)}\mathcal{S}$. Show that the curvature of $\boldsymbol{\gamma}$ is $S(t)/R(t)^2$, and that the product $R(t)S(t)$ is independent of t.

9.1.5 Show that a geodesic with nowhere vanishing curvature is a plane curve if and only if it is a line of curvature.

9.1.6 Let $\mathcal{S}_1$ and $\mathcal{S}_2$ be two surfaces that intersect in a curve $\mathcal{C}$, and let $\boldsymbol{\gamma}$ be a unit-speed parametrization of $\mathcal{C}$.

(i) Show that if $\boldsymbol{\gamma}$ is a geodesic on *both* $\mathcal{S}_1$ and $\mathcal{S}_2$ and if the curvature of $\boldsymbol{\gamma}$ is nowhere zero, then $\mathcal{S}_1$ ad $\mathcal{S}_2$ touch along $\boldsymbol{\gamma}$ (i.e., they have the same tangent plane at each point of $\mathcal{C}$). Give an example of this situation.

(ii) Show that if $\mathcal{S}_1$ and $\mathcal{S}_2$ intersect orthogonally at each point of $\mathcal{C}$, then $\boldsymbol{\gamma}$ is a geodesic on $\mathcal{S}_1$ if and only if $\dot{\mathbf{N}}_2$ is parallel to $\mathbf{N}_1$ at each point of $\mathcal{C}$ (where $\mathbf{N}_1$ and $\mathbf{N}_2$ are unit normals of $\mathcal{S}_1$ and $\mathcal{S}_2$). Show also that, in this case, $\boldsymbol{\gamma}$ is a geodesic on *both* $\mathcal{S}_1$ and $\mathcal{S}_2$ if and only if $\mathcal{C}$ is part of a straight line.

9.2 Geodesic equations

Unfortunately, Propositions 9.1.4 and 9.1.6 are not usually sufficient to determine all the geodesics on a given surface. For that, we need the following result:

Theorem 9.2.1

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

$$\begin{aligned}\frac{d}{dt}(E\dot{u} + F\dot{v}) &= \frac{1}{2}(E_u\dot{u}^2 + 2F_u\dot{u}\dot{v} + G_u\dot{v}^2),\\ \frac{d}{dt}(F\dot{u} + G\dot{v}) &= \frac{1}{2}(E_v\dot{u}^2 + 2F_v\dot{u}\dot{v} + G_v\dot{v}^2),\end{aligned} \tag{9.2}$$

where $Edu^2 + 2Fdudv + Gdv^2$ is the first fundamental form of $\boldsymbol{\sigma}$.

The differential equations (9.2) are called the *geodesic equations*.

Proof

Since $\{\boldsymbol{\sigma}_u, \boldsymbol{\sigma}_v\}$ is a basis of the tangent plane of $\boldsymbol{\sigma}$, $\boldsymbol{\gamma}$ is a geodesic if and only if $\ddot{\boldsymbol{\gamma}}$ is perpendicular to $\boldsymbol{\sigma}_u$ and $\boldsymbol{\sigma}_v$. Since $\dot{\boldsymbol{\gamma}} = \dot{u}\boldsymbol{\sigma}_u + \dot{v}\boldsymbol{\sigma}_v$, this is equivalent to

$$\left(\frac{d}{dt}(\dot{u}\boldsymbol{\sigma}_u + \dot{v}\boldsymbol{\sigma}_v)\right) \cdot \boldsymbol{\sigma}_u = 0 \quad \text{and} \quad \left(\frac{d}{dt}(\dot{u}\boldsymbol{\sigma}_u + \dot{v}\boldsymbol{\sigma}_v)\right) \cdot \boldsymbol{\sigma}_v = 0. \tag{9.3}$$

We show that these two equations are equivalent to the two geodesic equations.

The left-hand side of the first equation in (9.3) is equal to

$$\begin{aligned} &\frac{d}{dt}\left((\dot{u}\boldsymbol{\sigma}_u + \dot{v}\boldsymbol{\sigma}_v) \cdot \boldsymbol{\sigma}_u\right) - (\dot{u}\boldsymbol{\sigma}_u + \dot{v}\boldsymbol{\sigma}_v) \cdot \frac{d\boldsymbol{\sigma}_u}{dt} \\ &= \frac{d}{dt}(E\dot{u} + F\dot{v}) - (\dot{u}\boldsymbol{\sigma}_u + \dot{v}\boldsymbol{\sigma}_v) \cdot (\dot{u}\boldsymbol{\sigma}_{uu} + \dot{v}\boldsymbol{\sigma}_{uv}) \qquad (9.4) \\ &= \frac{d}{dt}(E\dot{u} + F\dot{v}) - (\dot{u}^2(\boldsymbol{\sigma}_u \cdot \boldsymbol{\sigma}_{uu}) + \dot{u}\dot{v}(\boldsymbol{\sigma}_u \cdot \boldsymbol{\sigma}_{uv} + \boldsymbol{\sigma}_v \cdot \boldsymbol{\sigma}_{uu}) + \dot{v}^2(\boldsymbol{\sigma}_v \cdot \boldsymbol{\sigma}_{uv})). \end{aligned}$$

Now,

$$E_u = (\boldsymbol{\sigma}_u \cdot \boldsymbol{\sigma}_u)_u = \boldsymbol{\sigma}_{uu} \cdot \boldsymbol{\sigma}_u + \boldsymbol{\sigma}_u \cdot \boldsymbol{\sigma}_{uu} = 2\boldsymbol{\sigma}_u \cdot \boldsymbol{\sigma}_{uu},$$

so $\boldsymbol{\sigma}_u \cdot \boldsymbol{\sigma}_{uu} = \frac{1}{2}E_u$. Similarly, $\boldsymbol{\sigma}_v \cdot \boldsymbol{\sigma}_{uv} = \frac{1}{2}G_u$. Finally,

$$\boldsymbol{\sigma}_u \cdot \boldsymbol{\sigma}_{uv} + \boldsymbol{\sigma}_v \cdot \boldsymbol{\sigma}_{uu} = (\boldsymbol{\sigma}_u \cdot \boldsymbol{\sigma}_v)_u = F_u.$$

Substituting these values into (9.4) gives

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

This shows that the first equation in (9.3) is equivalent to the first geodesic equation in (9.2). Similarly for the other equations. □

The geodesic equations are non-linear differential equations, and are usually difficult or impossible to solve explicitly. The following example is one case in which this can be done. Another is given in Exercise 9.2.3.

Example 9.2.2

We determine the geodesics on the unit sphere S^2 by solving the geodesic equations. For the usual parametrization by latitude θ and longitude φ,

$$\boldsymbol{\sigma}(\theta, \varphi) = (\cos\theta\cos\varphi, \cos\theta\sin\varphi, \sin\theta),$$

we found in Example 6.1.3 that the first fundamental form is

$$d\theta^2 + \cos^2\theta\, d\varphi^2.$$

We might as well restrict ourselves to unit-speed curves $\boldsymbol{\gamma}(t) = \boldsymbol{\sigma}(\theta(t), \varphi(t))$, so that

$$\dot{\theta}^2 + \dot{\varphi}^2 \cos^2\theta = 1,$$

and if $\boldsymbol{\gamma}$ is a geodesic the second equation in (9.2) gives

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

so that

$$\dot{\varphi}\cos^2\theta = \Omega,$$

where Ω is a constant. If $\Omega = 0$, then $\dot{\varphi} = 0$ and so φ is constant and $\boldsymbol{\gamma}$ is part of a meridian. We assume that $\dot{\varphi} \neq 0$ from now on.

The unit-speed condition gives

$$\dot{\theta}^2 = 1 - \frac{\Omega^2}{\cos^2\theta},$$

so along the geodesic we have

$$\left(\frac{d\theta}{d\varphi}\right)^2 = \frac{\dot{\theta}^2}{\dot{\varphi}^2} = \cos^2\theta(\Omega^{-2}\cos^2\theta - 1),$$

and hence

$$\pm(\varphi - \varphi_0) = \int \frac{d\theta}{\cos\theta\sqrt{\Omega^{-2}\cos^2\theta - 1}},$$

where φ_0 is a constant. The integral can be evaluated by making the substitution $u = \tan\theta$. This gives

$$\pm(\varphi - \varphi_0) = \int \frac{du}{\sqrt{\Omega^{-2} - 1 - u^2}} = \sin^{-1}\left(\frac{u}{\sqrt{\Omega^{-2} - 1}}\right),$$

and hence

$$\tan\theta = \pm\sqrt{\Omega^{-2} - 1}\sin(\varphi - \varphi_0).$$

This implies that the coordinates $x = \cos\theta\cos\varphi$, $y = \cos\theta\sin\varphi$ and $z = \sin\theta$ of $\boldsymbol{\gamma}(t)$ satisfy the equation

$$z = ax + by,$$

where $a = \mp\sqrt{\Omega^{-2} - 1}\sin\varphi_0$, and $b = \pm\sqrt{\Omega^{-2} - 1}\cos\varphi_0$. This shows that $\boldsymbol{\gamma}$ is contained in the intersection of S^2 with a plane passing through the origin.

Hence, in all cases, $\boldsymbol{\gamma}$ is part of a great circle.

The geodesic equations can be expressed in a different, but equivalent, form which is sometimes more useful than that in Theorem 9.2.1.

Proposition 9.2.3

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

$$\ddot{u} + \Gamma_{11}^1 \dot{u}^2 + 2\Gamma_{12}^1 \dot{u}\dot{v} + \Gamma_{22}^1 \dot{v}^2 = 0$$
$$\ddot{v} + \Gamma_{11}^2 \dot{u}^2 + 2\Gamma_{12}^2 \dot{u}\dot{v} + \Gamma_{22}^2 \dot{v}^2 = 0.$$

Proof

As we noted after Definition 9.1.1, $\boldsymbol{\gamma}$ is a geodesic if and only if $\dot{\boldsymbol{\gamma}}$ is parallel along $\boldsymbol{\gamma}$. Since $\dot{\boldsymbol{\gamma}} = \dot{u}\boldsymbol{\sigma}_u + \dot{v}\boldsymbol{\sigma}_v$, the equations in the statement of the proposition follow from Proposition 7.4.5. □

It can of course be verified directly that the differential equations in Proposition 9.2.3 are equivalent to those in Theorem 9.2.1 (see Exercise 9.2.6).

Proposition 9.2.3 makes it obvious that the geodesic equations are second-order ordinary differential equations for the functions $u(t)$ and $v(t)$. Even though we may be unable in many situations to solve these equations explicitly, the general theory of ordinary differential equations provides valuable information about their solutions. This leads to the following result, which tells us exactly 'how many' geodesics there are.

Proposition 9.2.4

Let $\mathbf{p}$ be a point of a surface $\mathcal{S}$, and let $\mathbf{t}$ be a unit tangent vector to $\mathcal{S}$ at $\mathbf{p}$. Then, there exists a unique unit-speed geodesic $\boldsymbol{\gamma}$ on $\mathcal{S}$ which passes through $\mathbf{p}$ and has tangent vector $\mathbf{t}$ there.

In short, *there is a unique geodesic through any given point of a surface in any given tangent direction.*

Proof

The geodesic equations in Proposition 9.2.3 are of the form

$$\ddot{u} = f(u, v, \dot{u}, \dot{v}), \quad \ddot{v} = g(u, v, \dot{u}, \dot{v}), \tag{9.5}$$

where f and g are smooth functions of the four variables $u, v, \dot{u}$ and $\dot{v}$. It is proved in the theory of ordinary differential equations that, for any given constants a, b, c, and d, and any value t_0 of t, there is a solution of Eqs. 9.5 such that

$$u(t_0) = a,\ v(t_0) = b,\ \dot{u}(t_0) = c,\ \dot{v}(t_0) = d, \tag{9.6}$$

and such that $u(t)$ and $v(t)$ are defined and smooth for all t satisfying $|t-t_0| < \epsilon$, where ϵ is some positive number. Moreover, any two solutions of Eqs. 9.5 satisfying (9.6) agree for all values of t such that $|t - t_0| < \epsilon'$, where ϵ' is some positive number $\leq \epsilon$.

We now apply these facts to the geodesic equations. Suppose that $\mathbf{p}$ lies in a patch $\boldsymbol{\sigma}(u, v)$ of $\mathcal{S}$, say $\mathbf{p} = \boldsymbol{\sigma}(a, b)$, and that $\mathbf{t} = c\boldsymbol{\sigma}_u + d\boldsymbol{\sigma}_v$, where a, b, c, and d are scalars and the derivatives are evaluated at $u = a$, $v = b$. A unit-speed curve $\boldsymbol{\gamma}(t) = \boldsymbol{\sigma}(u(t), v(t))$ passes through $\mathbf{p}$ at $t = t_0$ if and only if $u(t_0) = a$, $v(t_0) = b$, and has tangent vector $\mathbf{t}$ there if and only if

$$c\boldsymbol{\sigma}_u + d\boldsymbol{\sigma}_v = \mathbf{t} = \dot{\boldsymbol{\gamma}}(t_0) = \dot{u}(t_0)\boldsymbol{\sigma}_u + \dot{v}(t_0)\boldsymbol{\sigma}_v,$$

i.e., $\dot{u}(t_0) = c$, $\dot{v}(t_0) = d$. Thus, finding a (unit-speed) geodesic $\boldsymbol{\gamma}$ passing through $\mathbf{p}$ at $t = t_0$ and having tangent vector $\mathbf{t}$ is equivalent to solving the geodesic equations subject to the initial conditions (9.6). But we have said above that this problem has a unique solution. □

Example 9.2.5

We already know that all straight lines in a plane are geodesics. Since there is a straight line in the plane through any given point of the plane in any given direction parallel to the plane, it follows from Proposition 9.2.4 that there are no other geodesics.

Example 9.2.6

Similarly, on a sphere, the great circles are the only geodesics, for there is clearly a great circle passing through any given point of the sphere in any given direction tangent to the sphere. (If $\mathbf{p}$ is the point and $\mathbf{t}$ the tangent direction, let Π be the plane passing through the origin parallel to $\mathbf{p}$ and $\mathbf{t}$ (i.e., with normal $\mathbf{p} \times \mathbf{t}$); then take the intersection of the sphere with Π.)

The following consequence of Theorem 9.2.1 can also be used in some cases to find geodesics without solving the geodesic equations.

Corollary 9.2.7

Any local isometry between two surfaces takes the geodesics of one surface to the geodesics of the other.

Proof

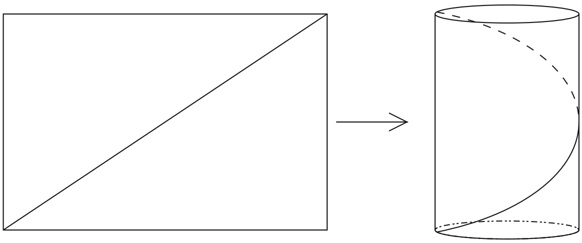
Let $\mathcal{S}_1$ and $\mathcal{S}_2$ be the two surfaces, let $f : \mathcal{S}_1 \to \mathcal{S}_2$ be the local isometry, and let $\boldsymbol{\gamma}_1$ be a geodesic on $\mathcal{S}_1$. Let $\mathbf{p}$ be a point on $\boldsymbol{\gamma}_1$ and let $\boldsymbol{\sigma}(u, v)$ be a surface patch of $\mathcal{S}_1$ with $\mathbf{p}$ in its image. Then, the part of $\boldsymbol{\gamma}_1$ lying in the patch $\boldsymbol{\sigma}$ is of the form $\boldsymbol{\gamma}_1(t) = \boldsymbol{\sigma}(u(t), v(t))$ with $a < t < b$, say, where the smooth functions u and v satisfy the geodesic equations (9.2), with E, F and G being the coefficients of the first fundamental form of $\boldsymbol{\sigma}$. By Corollary6.2.3, $f \circ \boldsymbol{\sigma}$ is a patch of $\mathcal{S}_2$ with the *same* first fundamental form as $\boldsymbol{\sigma}$. Hence, by Theorem 9.2.1, $\boldsymbol{\gamma}_2(t) = f(\boldsymbol{\sigma}(u(t), v(t)))$, with $a < t < b$, is a geodesic on $\mathcal{S}_2$. This implies that $\ddot{\boldsymbol{\gamma}}_2$ is perpendicular to $\mathcal{S}_2$ at $f(\mathbf{p})$. As this is true for all $\mathbf{p}$, $\boldsymbol{\gamma}_2$ is a geodesic on $\mathcal{S}_2$. □

Example 9.2.8

On the unit cylinder $\mathcal{S}$ given by $x^2 + y^2 = 1$, we know that the circles obtained by intersecting $\mathcal{S}$ with planes parallel to the xy-plane are geodesics (since they are normal sections). We also know that the straight lines on $\mathcal{S}$ parallel to the z-axis are geodesics. However, these are certainly not the only geodesics, for there is only one geodesic of each of the two types passing through each point of $\mathcal{S}$ (whereas we know that there is a geodesic passing through each point in *any given tangent direction*).

To find the missing geodesics, we recall that $\mathcal{S}$ is locally isometric to the plane (see Example 6.2.4). In fact, the local isometry takes the point $(u, v, 0)$ of the xy-plane to the point $(\cos u, \sin u, v) \in \mathcal{S}$. By Corollary 9.2.7, this map takes geodesics on the plane (i.e., straight lines) to geodesics on $\mathcal{S}$, and vice versa. So to find all the geodesics on $\mathcal{S}$, we have only to find the images under the local isometry of all the straight lines in the plane. Any line not parallel to the y-axis has equation $y = mx+c$, where m and c are constants. Parametrizing this line by $x = u$, $y = mu + c$, we see that its image is the curve

$$\boldsymbol{\gamma}(u) = (\cos u, \sin u, mu + c)$$



on $\mathcal{S}$. Comparing with Example 2.1.3, we see that this is a *circular helix* of radius one and pitch $2\pi|m|$ (adding c to the z-coordinate just translates the helix vertically). Note that if $m = 0$, we get the circular geodesics that we already know. Finally, any straight line in the xy-plane parallel to the y-axis is mapped by the local isometry to a straight line on $\mathcal{S}$ parallel to the z-axis, giving the other family of geodesics that we already know.

EXERCISES

9.2.1 Show that, if $\mathbf{p}$ and $\mathbf{q}$ are distinct points of the unit cylinder, there are either two or infinitely many geodesics on the cylinder with end-points $\mathbf{p}$ and $\mathbf{q}$ (and which do not otherwise pass through $\mathbf{p}$ or $\mathbf{q}$). Which pairs $\mathbf{p}, \mathbf{q}$ have the former property?

9.2.2 Use Corollary 9.2.7 to find all the geodesics on a circular cone.

9.2.3 Find the geodesics on the unit cylinder by solving the geodesic equations.

9.2.4 Consider the following three properties that a curve $\boldsymbol{\gamma}$ on a surface may have:

(i) $\boldsymbol{\gamma}$ has constant speed.

(ii) $\boldsymbol{\gamma}$ satisfies the first of the geodesic equations (9.2).

(iii) $\boldsymbol{\gamma}$ satisfies the second of the geodesic equations (9.2).

Show, without using Theorem 9.2.1, that (ii) and (iii) together imply (i). Show also that if (i) holds and if $\boldsymbol{\gamma}$ is not a parameter curve, then (ii) and (iii) are equivalent.

9.2.5 Let $\boldsymbol{\gamma}(t)$ be a unit-speed curve on the helicoid

$$\boldsymbol{\sigma}(u, v) = (u \cos v, u \sin v, v)$$

(Exercise 4.2.6). Show that

$$\dot{u}^2 + (1 + u^2)\dot{v}^2 = 1$$

(a dot denotes d/dt). Show also that, if $\boldsymbol{\gamma}$ is a geodesic on $\boldsymbol{\sigma}$, then

$$\dot{v} = \frac{a}{1 + u^2},$$

where a is a constant. Find the geodesics corresponding to $a = 0$ and $a = 1$.

Suppose that a geodesic $\boldsymbol{\gamma}$ on $\boldsymbol{\sigma}$ intersects a ruling at a point $\mathbf{p}$ a distance $D > 0$ from the z-axis, and that the angle between $\boldsymbol{\gamma}$ and the ruling at $\mathbf{p}$ is α, where $0 < \alpha < \pi/2$. Show that the geodesic intersects the z-axis if $D > \cot\alpha$, but that if $D < \cot\alpha$ its smallest distance from the z-axis is $\sqrt{D^2\sin^2\alpha - \cos^2\alpha}$. Find the equation of the geodesic if $D = \cot\alpha$.

9.2.6 Verify directly that the differential equations in Proposition 9.2.3 are equivalent to the geodesic equations in Theorem 9.2.1.

9.3 Geodesics on surfaces of revolution

It turns out that, although the geodesic equations for a surface of revolution cannot usually be solved explicitly, they can be used to get a good *qualitative* understanding of the geodesics on such a surface.

We parametrize the surface of revolution in the usual way

$$\boldsymbol{\sigma}(u, v) = (f(u)\cos v, f(u)\sin v, g(u)),$$

where we assume that $f > 0$ and $\left(\frac{df}{du}\right)^2 + \left(\frac{dg}{du}\right)^2 = 1$ (see Example 5.3.2 – we used a dot there to denote d/du, but now a dot is reserved for d/dt, where t is the parameter along a geodesic). We found in Example 6.1.3 that the first fundamental form of $\boldsymbol{\sigma}$ is $du^2 + f(u)^2dv^2$. Referring to Eq. 9.2,

$$\ddot{u} = f(u)\frac{df}{du}\dot{v}^2, \quad \frac{d}{dt}(f(u)^2\dot{v}) = 0. \tag{9.7}$$

We might as well consider unit-speed geodesics, so that

$$\dot{u}^2 + f(u)^2\dot{v}^2 = 1. \tag{9.8}$$

From this, we make the following easy deductions:

Proposition 9.3.1

On the surface of revolution

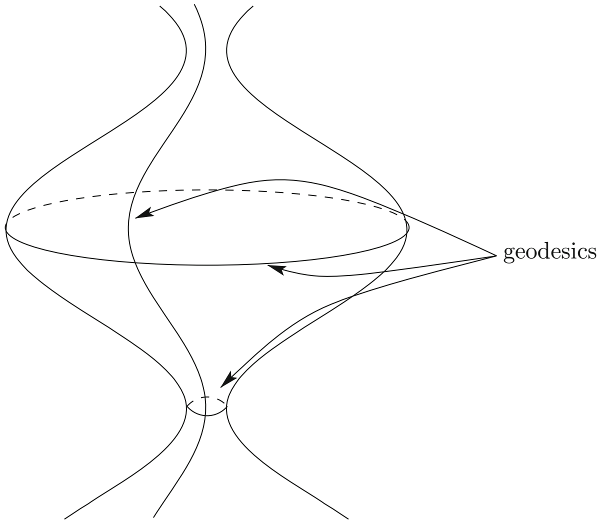
$$\boldsymbol{\sigma}(u, v) = (f(u)\cos v, f(u)\sin v, g(u)),$$

(i) Every meridian is a geodesic.

(ii) A parallel $u = u_0$ (say) is a geodesic if and only if $df/du = 0$ when $u = u_0$, i.e., u_0 is a stationary point of f.

Proof

On a meridian, we have v = constant so the second equation in (9.7) is obviously satisfied. Equation 9.8 gives $\dot{u} = \pm 1$, so $\dot{u}$ is constant and the first equation in (9.7) is also satisfied.



For (ii), note that if $u = u_0$ is constant, then by Eq. 9.8, $\dot{v} = \pm 1/f(u_0)$ is non-zero, and so the first equation in (9.7) holds only if $df/du = 0$. Conversely, if $df/du = 0$ when $u = u_0$, the first equation in (9.7) obviously holds, and the second holds because $\dot{v} = \pm 1/f(u_0)$ and $f(u) = f(u_0)$ are constant. $\square$

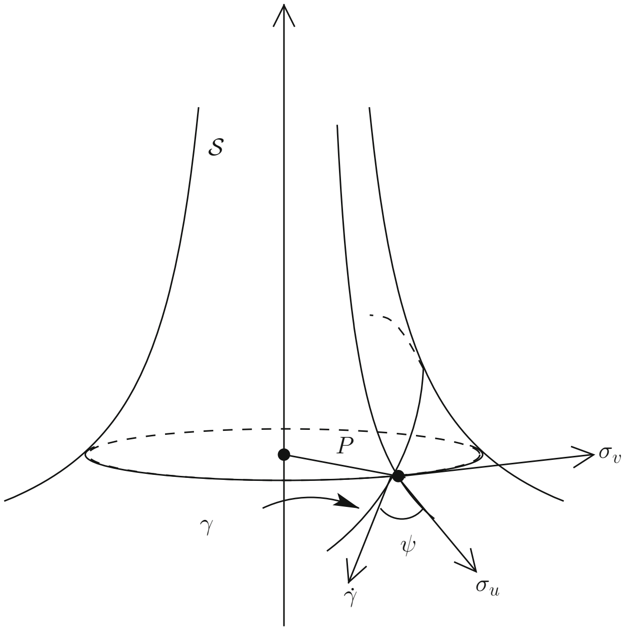
Of course, this proposition only gives some of the geodesics on a surface of revolution. The following result is very helpful in understanding the remaining geodesics.

Proposition 9.3.2 (Clairaut's Theorem)

Let $\boldsymbol{\gamma}$ be a unit-speed curve on a surface of revolution $\mathcal{S}$, let $\rho : \mathcal{S} \to \mathbb{R}$ be the distance of a point of $\mathcal{S}$ from the axis of rotation, and let ψ be the angle between $\dot{\boldsymbol{\gamma}}$ and the meridians of $\mathcal{S}$. If $\boldsymbol{\gamma}$ is a geodesic, then $\rho \sin \psi$ is constant along $\boldsymbol{\gamma}$. Conversely, if $\rho \sin \psi$ is constant along $\boldsymbol{\gamma}$, and if no part of $\boldsymbol{\gamma}$ is part of some parallel of $\mathcal{S}$, then $\boldsymbol{\gamma}$ is a geodesic.

By a 'part' of $\boldsymbol{\gamma}$ we mean $\boldsymbol{\gamma}(J)$, where J is an open interval. The hypothesis there cannot be relaxed, for on a parallel $\psi = \pi/2$, and so $\rho \sin \psi$

is certainly constant. But parallels are not geodesics in general, as Proposition 9.3.1(ii) shows.



Proof

Parametrizing $\mathcal{S}$ as in Proposition 9.3.1, we have $\rho = f(u)$. Note that $\boldsymbol{\sigma}_u$ and $\rho^{-1}\boldsymbol{\sigma}_v$ are unit vectors tangent to the meridians and parallels, respectively, and that they are perpendicular since $F = 0$. Assuming that $\boldsymbol{\gamma}(t) = \boldsymbol{\sigma}(u(t), v(t))$ is unit-speed, we have

$$\dot{\boldsymbol{\gamma}} = \cos\psi\, \boldsymbol{\sigma}_u + \rho^{-1} \sin\psi\, \boldsymbol{\sigma}_v$$

(this equation actually serves to define the sign of ψ, which is left ambiguous in the statement of Clairaut's Theorem). Hence,

$$\boldsymbol{\sigma}_u \times \dot{\boldsymbol{\gamma}} = \rho^{-1} \sin\psi\, \boldsymbol{\sigma}_u \times \boldsymbol{\sigma}_v.$$

Since $\dot{\boldsymbol{\gamma}} = \dot{u}\boldsymbol{\sigma}_u + \dot{v}\boldsymbol{\sigma}_v$, this gives

$$\dot{v}\, \boldsymbol{\sigma}_u \times \boldsymbol{\sigma}_v = \rho^{-1} \sin\psi\, \boldsymbol{\sigma}_u \times \boldsymbol{\sigma}_v.$$

Hence, $\rho\dot{v} = \sin\psi$ and so

$$\rho \sin\psi = \rho^2 \dot{v}.$$

But the second equation in (9.7) shows that this is a constant, say Ω, along the geodesic.

For the converse, if $\rho \sin \psi$ is a constant Ω along a unit-speed curve $\boldsymbol{\gamma}$ in $\mathcal{S}$, the above argument shows that the second equation in (9.7) is satisfied, and we must show that the first equation in (9.7) is satisfied too. Since

$$\dot{v} = \frac{\sin \psi}{\rho} = \frac{\Omega}{\rho^2}, \tag{9.9}$$

Eq. 9.8 gives

$$\dot{u}^2 = 1 - \frac{\Omega^2}{\rho^2}. \tag{9.10}$$

Differentiating both sides with respect to t gives

$$2\dot{u}\ddot{u} = \frac{2\Omega^2}{\rho^3}\dot{\rho} = \frac{2\Omega^2}{\rho^3}\frac{d\rho}{du}\dot{u},$$

$$\therefore \quad \dot{u}\left(\ddot{u} - \rho\frac{d\rho}{du}\dot{v}^2\right) = 0.$$

If the term in brackets does not vanish at some point of the curve, say at $\boldsymbol{\gamma}(t_0) = \boldsymbol{\sigma}(u_0, v_0)$, there will be a number $\epsilon > 0$ such that it does not vanish for $|t - t_0| < \epsilon$. But then $\dot{u} = 0$ for $|t - t_0| < \epsilon$, and so $\boldsymbol{\gamma}$ coincides with the parallel $u = u_0$ when $|t - t_0| < \epsilon$, contrary to our assumption. Hence, the term in brackets must vanish everywhere on $\boldsymbol{\gamma}$, i.e.,

$$\ddot{u} = \rho\frac{d\rho}{du}\dot{v}^2,$$

showing that the first equation in (9.7) is indeed satisfied. □

Clairaut's Theorem has a simple mechanical interpretation. Recall that the geodesics on a surface $\mathcal{S}$ are the curves traced on $\mathcal{S}$ by a particle subject to no forces except a force normal to $\mathcal{S}$ that constrains it to move on $\mathcal{S}$. When $\mathcal{S}$ is a surface of revolution, the force at a point $\mathbf{p} \in \mathcal{S}$ lies in the plane containing the axis of revolution and $\mathbf{p}$, and so has no moment about the axis. It follows that the angular momentum Ω of the particle about the axis is constant. But, if the particle moves along a unit-speed geodesic, the component of its velocity along the parallel through $\mathbf{p}$ is $\sin \psi$, so its angular momentum about the axis is proportional to $\rho \sin \psi$.

Example 9.3.3

We use Clairaut's theorem to determine the geodesics on the pseudosphere:

$$\boldsymbol{\sigma}(u, v) = (e^u \cos v, e^u \sin v, \sqrt{1 - e^{2u}} - \cosh^{-1}(e^{-u})).$$

We found in Section 8.3 that its first fundamental form is

$$du^2 + e^{2u} dv^2.$$

It is convenient to reparametrize by setting $w = e^{-u}$. The reparametrized surface is

$$\tilde{\boldsymbol{\sigma}}(v, w) = \left(\frac{1}{w} \cos v, \frac{1}{w} \sin v, \sqrt{1 - \frac{1}{w^2}} - \cosh^{-1} w \right),$$

and its first fundamental form is

$$\frac{dv^2 + dw^2}{w^2}. \tag{9.11}$$

We must have $w > 1$ for $\tilde{\boldsymbol{\sigma}}$ to be well defined and smooth.

If $\boldsymbol{\gamma}(t) = \tilde{\boldsymbol{\sigma}}(v(t), w(t))$ is a unit-speed geodesic, the unit-speed condition gives

$$\dot{v}^2 + \dot{w}^2 = w^2, \tag{9.12}$$

and Clairaut's theorem gives

$$\frac{1}{w} \sin \psi = \frac{1}{w^2} \dot{v} = \Omega, \tag{9.13}$$

where Ω is a constant, since $\rho = 1/w$. Thus, $\dot{v} = \Omega w^2$. If $\Omega = 0$, we get a meridian $v =$ constant. Assuming now that $\Omega \neq 0$ and substituting in Eq. 9.12 gives

$$\dot{w} = \pm w \sqrt{1 - \Omega^2 w^2}.$$

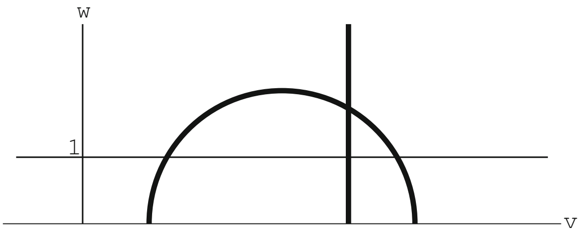
Hence, along the geodesic,

$$\frac{dv}{dw} = \frac{\dot{v}}{\dot{w}} = \pm \frac{\Omega w}{\sqrt{1 - \Omega^2 w^2}},$$

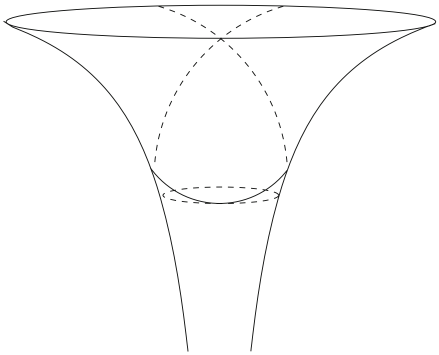
$$\therefore \quad (v - v_0) = \mp \frac{1}{\Omega} \sqrt{1 - \Omega^2 w^2}, \tag{9.14}$$

$$\therefore \quad (v - v_0)^2 + w^2 = \frac{1}{\Omega^2},$$

where v_0 is a constant. So the geodesics are the images under $\tilde{\boldsymbol{\sigma}}$ of the parts of the circles in the vw-plane given by Eq. 9.14 and lying in the region $w > 1$. Note that these circles all have centre on the v-axis, and so intersect the v-axis perpendicularly. The meridians correspond to straight lines perpendicular to the v-axis.



The corresponding geodesics on the pseudosphere itself are shown below. Note that the geodesics cannot be extended indefinitely, in one direction in the case of the meridians and in both directions for the others. This is because the geodesics 'run into' the circular edge of the pseudosphere in the xy-plane. A bug walking at constant speed along such a geodesic would reach the edge in a finite time, and thus would suffer the fate feared by ancient mariners of falling off the edge of the world. In terms of the vw-plane, the reason for this is that the line $w = 1$ is a boundary of the region that corresponds to the pseudosphere and the straight lines and semicircles that correspond to the geodesics cross this line.



Clairaut's theorem can often be used to determine the *qualitative* behaviour of the geodesics on a surface $\mathcal{S}$, when solving the geodesic differential equations explicitly may be difficult or impossible. Note first that, in general, there are two geodesics passing through any given point $\mathbf{p} \in \mathcal{S}$ with a given angular momentum Ω, for $\dot{v}$ is determined by Eq. 9.9 and $\dot{u}$ up to sign by Eq. 9.10. In fact, one geodesic is obtained from the other by reflecting in the plane through $\mathbf{p}$ containing the of rotation (which changes Ω to $-\Omega$) followed by changing the parameter t of the geodesic to $-t$ (which changes the angular momentum back to Ω again).

The discussion in the preceding paragraph shows that we may as well assume that $\Omega > 0$, which we do from now on. Then, Eq. 9.10 shows that *the geodesic is confined to the part of $\mathcal{S}$ which is at a distance $\geq \Omega$ from the axis.*

If all of $\mathcal{S}$ is a distance $> \Omega$ from the axis, the geodesic will cross every parallel of $\mathcal{S}$. For otherwise, u would be bounded above or below on $\mathcal{S}$, say the former. Let u_0 be the least upper bound of u on the geodesic, and let $\Omega + 2\epsilon$, where $\epsilon > 0$, be the radius of the parallel $u = u_0$. If u is sufficiently close to u_0, the radius of the corresponding parallel will be $\geq \Omega + \epsilon$, and on the part of the geodesic lying in this region we shall have

$$|\dot{u}| \geq \sqrt{1 - \left(\frac{\Omega}{\Omega + \epsilon}\right)^2} > 0$$

by Eq. 9.10. But this clearly implies that the geodesic will cross $u = u_0$, contradicting our assumption.

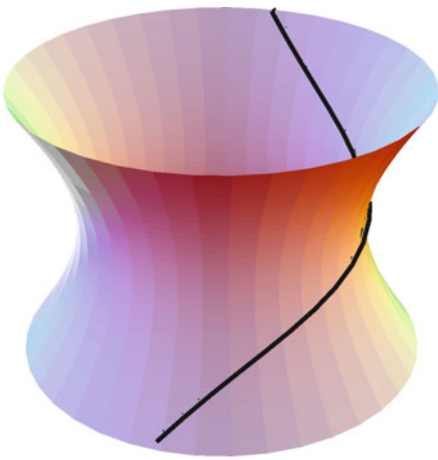
Thus, the interesting case is that in which part of $\mathcal{S}$ is within a distance Ω of the axis. The discussion of this case will be clearer if we consider a concrete example whose geodesics nevertheless exhibit essentially all possible forms of behaviour.

Example 9.3.4

We consider the hyperboloid of one sheet obtained by rotating the hyperbola

$$x^2 - z^2 = 1, \quad x > 0,$$

in the xz-plane around the z-axis. Since all of the surface is at a distance ≥ 1 from the z-axis, we have seen above that, if $0 \leq \Omega < 1$, a geodesic with angular momentum Ω crosses every parallel of the hyperboloid and so extends from $z = -\infty$ to $z = \infty$.



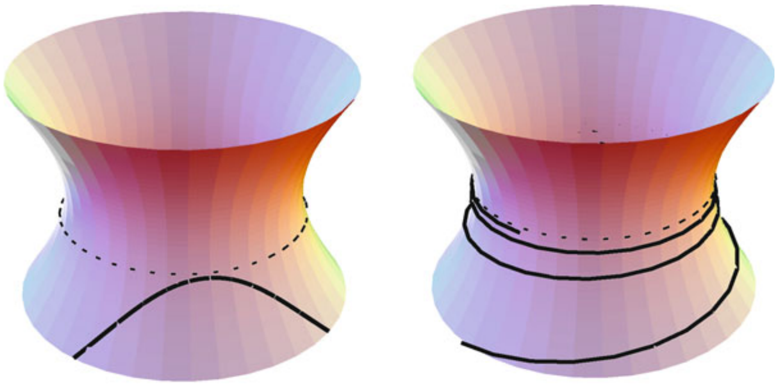
$0 < \Omega < 1$

Suppose now that $\Omega > 1$. Then the geodesic is confined to one of the two regions

$$z \geq \sqrt{\Omega^2 - 1}, \quad z \leq -\sqrt{\Omega^2 - 1},$$

which are bounded by circles Γ^+ and Γ^-, respectively, of radius Ω. Let $\mathbf{p}$ be a point on Γ^-, and consider the geodesic $\mathcal{C}$ that passes through $\mathbf{p}$ and is tangent to Γ^- there. Then, $\psi = \pi/2$ and $\rho = \Omega$ at $\mathbf{p}$, so $\mathcal{C}$ has angular momentum Ω. Now $\mathcal{C}$ cannot be contained in Γ^-, since Γ^- is not a geodesic (by Proposition 9.3.1(ii)), so $\mathcal{C}$ must head into the region below Γ^- as it leaves $\mathbf{p}$. Moreover, $\mathcal{C}$ must be symmetric about $\mathbf{p}$, since reflection in the plane through $\mathbf{p}$ containing the z-axis takes $\mathcal{C}$ to another geodesic that also passes through $\mathbf{p}$ and is tangent to Γ^- there, and so must coincide with $\mathcal{C}$ by the uniqueness part of Corollary 9.2.4. Since $\dot{u} \neq 0$ in the region below Γ^- by Eq. 9.10, the geodesic crosses every parallel below Γ^- and $z \to -\infty$ as $t \to \pm\infty$.

Suppose now that $\tilde{\mathcal{C}}$ is *any* geodesic with angular momentum $\Omega > 1$ in the region below Γ^-. Then a suitable rotation around the z-axis will cause $\tilde{\mathcal{C}}$ to intersect $\mathcal{C}$, say at $\mathbf{q}$, and so to coincide with it (possibly after reflecting in the plane through $\mathbf{q}$ containing the z-axis and changing t to $-t$). We have therefore described the behaviour of every geodesic with angular momentum $\Omega > 1$ that is confined to the region below Γ^-. Of course, the geodesics with angular momentum $\Omega > 1$ in the region above Γ^+ are obtained by reflecting those below Γ^- in the xy-plane.



$\Omega > 1$ $\qquad$ $\Omega = 1$

Suppose finally that $\Omega = 1$. Let $\mathcal{C}$ be a geodesic with angular momentum 1 passing through a point $\mathbf{p}$. If $\mathbf{p}$ is on the waist Γ of the hyperboloid (i.e., the unit circle in the xy-plane), which is a geodesic by Proposition 9.3.1(ii), then $\rho = 1$ at $\mathbf{p}$ and so $\psi = \pi/2$ and $\mathcal{C}$ is tangent to Γ at $\mathbf{p}$. It must therefore coincide with Γ. If, on the other hand, $\mathbf{p}$ is in the region below Γ, then $0 < \psi < \pi/2$

at $\mathbf{p}$, so as it leaves $\mathbf{p}$ in one direction, $\mathcal{C}$ approaches Γ. It must in fact get arbitrarily close to Γ. For if it were to stay always below a parallel $\tilde{\Gamma}$ of radius $1 + \epsilon$, say (with $\epsilon > 0$), then we would have

$$|\dot{u}| \geq \sqrt{1 - \left(\frac{1}{1+\epsilon}\right)^2}$$

everywhere along $\mathcal{C}$ by Eq. 9.10, which clearly implies that $\mathcal{C}$ must cross every parallel, contradicting our assumption. So, if $\Omega = 1$, the geodesic spirals around the hyperboloid approaching, and getting arbitrarily close to, Γ but never quite reaching it.

EXERCISES

9.3.1 There is another way to see that all the meridians, and the parallels corresponding to the stationary points of f, are geodesics on a surface of revolution considered in this section. What is it?

9.3.2 Describe qualitatively the geodesics on:

(i) A spheroid, obtained by rotating an ellipse around one of its axes.

(ii) A torus (Exercise 4.2.5).

9.3.3 Show that a geodesic on the pseudosphere with non-zero angular momentum Ω intersects itself if and only if $\Omega < (1 + \pi^2)^{-1/2}$. How many self-intersections are there in that case?

9.3.4 Show that if we reparametrize the pseudosphere as in Exercise 8.3.1(ii), the geodesics on the pseudosphere correspond to segments of straight lines and circles in the parameter plane that intersect the boundary of the disc orthogonally. Deduce that, in the parametrization of Exercise 8.3.1(iii), the geodesics correspond to segments of *straight lines* in the parameter plane. We shall see in Section 10.4 that there are very few surfaces that have parametrizations with this property.

9.4 Geodesics as shortest paths

Everyone knows that the straight line segment joining two points $\mathbf{p}$ and $\mathbf{q}$ in a plane is the shortest path between $\mathbf{p}$ and $\mathbf{q}$ (see Exercise 1.2.4). It is almost as well known that great circles are the shortest paths on a sphere (Proposition 6.5.1). And we have seen that the straight lines are the geodesics in a plane, and the great circles are the geodesics on a sphere.

To see the connection between geodesics and shortest paths on an arbitrary surface $\mathcal{S}$, we consider a unit-speed curve $\boldsymbol{\gamma}$ on $\mathcal{S}$ passing through two fixed points $\mathbf{p}, \mathbf{q} \in \mathcal{S}$. If $\boldsymbol{\gamma}$ is a shortest path on $\mathcal{S}$ from $\mathbf{p}$ to $\mathbf{q}$, then the part of $\boldsymbol{\gamma}$ contained in any surface patch $\boldsymbol{\sigma}$ of $\mathcal{S}$ must be the shortest path between any two of its points. For if $\mathbf{p}'$ and $\mathbf{q}'$ are any two points of $\boldsymbol{\gamma}$ in (the image of) $\boldsymbol{\sigma}$, and if there were a shorter path in $\boldsymbol{\sigma}$ from $\mathbf{p}'$ to $\mathbf{q}'$ than $\boldsymbol{\gamma}$, we could replace the part of $\boldsymbol{\gamma}$ between $\mathbf{p}'$ and $\mathbf{q}'$ by this shorter path, thus giving a shorter path from $\mathbf{p}$ to $\mathbf{q}$ in $\mathcal{S}$.

We may therefore consider a path $\boldsymbol{\gamma}$ entirely contained in a surface patch $\boldsymbol{\sigma}$. To test whether $\boldsymbol{\gamma}$ has smaller length than any other path in $\boldsymbol{\sigma}$ passing through two fixed points $\mathbf{p}, \mathbf{q}$ on $\boldsymbol{\sigma}$; we embed $\boldsymbol{\gamma}$ in a smooth family of curves on $\boldsymbol{\sigma}$ passing through $\mathbf{p}$ and $\mathbf{q}$. By such a family, we mean a curve $\boldsymbol{\gamma}^\tau$ on $\boldsymbol{\sigma}$, for each τ in an open interval $(-\delta, \delta)$, such that

(i) there is an $\epsilon > 0$ such that $\boldsymbol{\gamma}^\tau(t)$ is defined for all $t \in (-\epsilon, \epsilon)$ and all $\tau \in (-\delta, \delta)$;

(ii) for some a, b with $-\epsilon < a < b < \epsilon$, we have

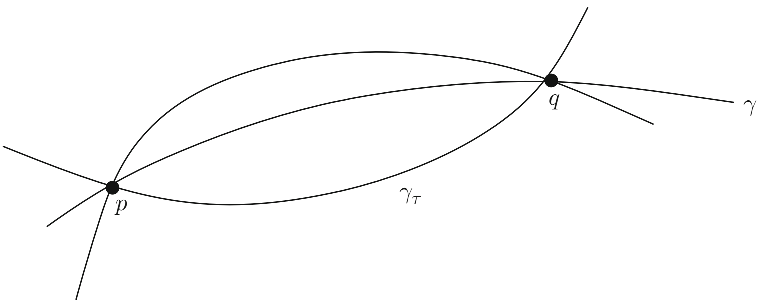
$$\boldsymbol{\gamma}^\tau(a) = \mathbf{p} \quad \text{and} \quad \boldsymbol{\gamma}^\tau(b) = \mathbf{q} \quad \text{for all} \quad \tau \in (-\delta, \delta);$$

(iii) the map from the rectangle $(-\delta, \delta) \times (-\epsilon, \epsilon)$ into $\mathbb{R}^3$ given by

$$(\tau, t) \mapsto \boldsymbol{\gamma}^\tau(t)$$

is smooth;

(iv) $\boldsymbol{\gamma}^0 = \boldsymbol{\gamma}$.



The length of the part of $\boldsymbol{\gamma}^\tau$ between $\mathbf{p}$ and $\mathbf{q}$ is

$$\mathcal{L}(\tau) = \int_a^b \parallel \dot{\boldsymbol{\gamma}}^\tau \parallel \, dt,$$

where a dot denotes d/dt.

Theorem 9.4.1

With the above notation, the unit-speed curve $\boldsymbol{\gamma}$ is a geodesic if and only if

$$\frac{d}{d\tau}\mathcal{L}(\tau) = 0 \quad \text{when } \tau = 0$$

for all families of curves $\boldsymbol{\gamma}^\tau$ with $\boldsymbol{\gamma}^0 = \boldsymbol{\gamma}$.

Note that although we assumed that $\boldsymbol{\gamma} = \boldsymbol{\gamma}^0$ is unit-speed, we *cannot assume that* $\boldsymbol{\gamma}^\tau$ *is unit-speed if* $\tau \neq 0$, as this would imply that the length of the segment of $\boldsymbol{\gamma}^\tau$ corresponding to $a \leq t \leq b$ is independent of τ.

Proof

We use the formula for 'differentiating under the integral sign': if $f(\tau, t)$ is smooth,

$$\frac{d}{d\tau}\int f(\tau, t)dt = \int \frac{\partial f}{\partial \tau}dt.$$

Thus,

$$\begin{aligned}
\frac{d}{d\tau}\mathcal{L}(\tau) &= \frac{d}{d\tau}\int_a^b \parallel \dot{\boldsymbol{\gamma}}^\tau \parallel \, dt \\
&= \frac{d}{d\tau}\int_a^b (E\dot{u}^2 + 2F\dot{u}\dot{v} + G\dot{v}^2)^{1/2}\, dt \\
&= \int_a^b \frac{\partial}{\partial \tau}(g(\tau, t)^{1/2})\, dt \qquad (9.15) \\
&= \frac{1}{2}\int_a^b g(\tau, t)^{-1/2}\frac{\partial g}{\partial \tau}\, dt,
\end{aligned}$$

where

$$g(\tau, t) = E\dot{u}^2 + 2F\dot{u}\dot{v} + G\dot{v}^2$$

and a dot denotes d/dt. Now,

$$\begin{aligned}
\frac{\partial g}{\partial \tau} &= \frac{\partial E}{\partial \tau}\dot{u}^2 + 2\frac{\partial F}{\partial \tau}\dot{u}\dot{v} + \frac{\partial G}{\partial \tau}\dot{v}^2 + 2E\dot{u}\frac{\partial \dot{u}}{\partial \tau} + 2F\left(\frac{\partial \dot{u}}{\partial \tau}\dot{v} + \dot{u}\frac{\partial \dot{v}}{\partial \tau}\right) + 2G\dot{v}\frac{\partial \dot{v}}{\partial \tau} \\
&= \left(E_u\frac{\partial u}{\partial \tau} + E_v\frac{\partial v}{\partial \tau}\right)\dot{u}^2 + 2\left(F_u\frac{\partial u}{\partial \tau} + F_v\frac{\partial v}{\partial \tau}\right)\dot{u}\dot{v} + \left(G_u\frac{\partial u}{\partial \tau} + G_v\frac{\partial v}{\partial \tau}\right)\dot{v}^2 \\
&\quad + 2E\dot{u}\frac{\partial^2 u}{\partial \tau \partial t} + 2F\left(\frac{\partial^2 u}{\partial \tau \partial t}\dot{v} + \dot{u}\frac{\partial^2 v}{\partial \tau \partial t}\right) + 2G\dot{v}\frac{\partial^2 v}{\partial \tau \partial t} \\
&= (E_u\dot{u}^2 + 2F_u\dot{u}\dot{v} + G_u\dot{v}^2)\frac{\partial u}{\partial \tau} + (E_v\dot{u}^2 + 2F_v\dot{u}\dot{v} + G_v\dot{v}^2)\frac{\partial v}{\partial \tau} \\
&\quad + 2(E\dot{u} + F\dot{v})\frac{\partial^2 u}{\partial \tau \partial t} + 2(F\dot{u} + G\dot{v})\frac{\partial^2 v}{\partial \tau \partial t}.
\end{aligned}$$

The contribution to the integral in Eq. 9.15 coming from the terms involving the second partial derivatives is

$$\begin{aligned}
&\int_a^b g^{-1/2}\left\{(E\dot{u}+F\dot{v})\frac{\partial^2 u}{\partial\tau\partial t} +(F\dot{u}+G\dot{v})\frac{\partial^2 v}{\partial\tau\partial t}\right\} dt \\
&\quad = g^{-1/2}\left\{(E\dot{u}+F\dot{v})\frac{\partial u}{\partial\tau}+(F\dot{u}+G\dot{v})\frac{\partial v}{\partial\tau}\right\}\Bigg|_{t=a}^{t=b} \\
&\qquad -\int_a^b\left(\frac{\partial}{\partial t}\left\{g^{-1/2}(E\dot{u}+F\dot{v})\right\}\frac{\partial u}{\partial\tau}+\frac{\partial}{\partial t}\left\{g^{-1/2}(F\dot{u}+G\dot{v})\right\}\frac{\partial v}{\partial\tau}\right) dt,
\end{aligned} \tag{9.16}$$

using integration by parts. Now, since $\boldsymbol{\gamma}^\tau(a)$ and $\boldsymbol{\gamma}^\tau(b)$ are independent of τ (being equal to $\mathbf{p}$ and $\mathbf{q}$, respectively), we have

$$\frac{\partial\boldsymbol{\gamma}^\tau}{\partial\tau}=\mathbf{0}\quad\text{when } t=a \text{ or } b.$$

Since

$$\frac{\partial\boldsymbol{\gamma}^\tau}{\partial\tau}=\frac{\partial u}{\partial\tau}\boldsymbol{\sigma}_u+\frac{\partial v}{\partial\tau}\boldsymbol{\sigma}_v,$$

we see that

$$\frac{\partial u}{\partial\tau}=\frac{\partial v}{\partial\tau}=0\quad\text{when } t=a \text{ or } b.$$

Hence, the first term on the right-hand side of Eq. 9.16 is zero. Inserting the remaining terms in Eq. 9.16 back into Eq. 9.15, we get

$$\frac{d}{d\tau}\mathcal{L}(\tau)=\int_a^b\left(U\frac{\partial u}{\partial\tau}+V\frac{\partial v}{\partial\tau}\right) dt, \tag{9.17}$$

where

$$\begin{aligned}
U(\tau,t)&=\frac{1}{2}g^{-1/2}(E_u\dot{u}^2+2F_u\dot{u}\dot{v}+G_u\dot{v}^2)-\frac{d}{dt}\left\{g^{-1/2}(E\dot{u}+F\dot{v})\right\},\\
V(\tau,t)&=\frac{1}{2}g^{-1/2}(E_v\dot{u}^2+2F_v\dot{u}\dot{v}+G_v\dot{v}^2)-\frac{d}{dt}\left\{g^{-1/2}(F\dot{u}+G\dot{v})\right\}.
\end{aligned} \tag{9.18}$$

Now $\boldsymbol{\gamma}^0=\boldsymbol{\gamma}$ is unit-speed, so since $\parallel\dot{\boldsymbol{\gamma}}^\tau\parallel^2=g(\tau,t)$, we have $g(\tau,t)=1$ for all t when $\tau=0$. Comparing Eq. 9.18 with the geodesic equations in (9.2), we see that, if $\boldsymbol{\gamma}$ is a geodesic, then $U=V=0$ when $\tau=0$, and hence by Eq. 9.17,

$$\frac{d}{d\tau}\mathcal{L}(\tau)=0\quad\text{when } \tau=0.$$

For the converse, we have to show that, if

$$\int_a^b\left(U\frac{\partial u}{\partial\tau}+V\frac{\partial v}{\partial\tau}\right) dt=0\quad\text{when } \tau=0 \tag{9.19}$$

for *all* families of curves γ^τ, then $U = V = 0$ when $\tau = 0$ (since this will prove that γ satisfies the geodesic equations). Assume, then, that condition (9.19) holds, and suppose, for example, that $U \neq 0$ when $\tau = 0$. We will show that this leads to a contradiction.

Since $U \neq 0$ when $\tau = 0$, there is some $t_0 \in (a, b)$ such that $U(0, t_0) \neq 0$, say $U(0, t_0) > 0$. Since U is a continuous function, there exists $\eta > 0$ such that

$$U(0,t) > 0 \quad \text{if } t \in (t_0 - \eta, t_0 + \eta).$$

Let ϕ be a smooth function such that

$$\phi(t) > 0 \quad \text{if } t \in (t_0 - \eta, t_0 + \eta) \quad \text{and} \quad \phi(t) = 0 \quad \text{if } t \notin (t_0 - \eta, t_0 + \eta). \tag{9.20}$$

(The construction of such a function ϕ is outlined in Exercise 9.4.3.) Suppose that $\gamma(t) = \sigma(u(t), v(t))$, and consider the family of curves $\gamma^\tau(t) = \sigma(u(\tau, t), v(\tau, t))$, where

$$u(\tau, t) = u(t) + \tau\phi(t), \quad v(\tau, t) = v(t).$$

Then, $\partial u/\partial\tau = \phi$ and $\partial v/\partial\tau = 0$ for all τ and t, so Eq. 9.19 gives

$$0 = \int_a^b \left(U\frac{\partial u}{\partial \tau} + V\frac{\partial v}{\partial \tau} \right)\bigg|_{\tau=0} dt = \int_{t_0-\eta}^{t_0+\eta} U(0,t)\phi(t)\,dt. \tag{9.21}$$

But $U(0, t)$ and $\phi(t)$ are both > 0 for all $t \in (t_0 - \eta, t_0 + \eta)$, so the integral on the right-hand side of Eq. 9.21 is > 0. This contradiction proves that we must have $U(0, t) = 0$ for all $t \in (a, b)$. One proves similarly that $V(0, t) = 0$ for all $t \in (a, b)$. Together, these results prove that γ satisfies the geodesic equations. □

It is worth making several comments on Theorem 9.4.1 to be clear about what it implies, and also what it does not imply.

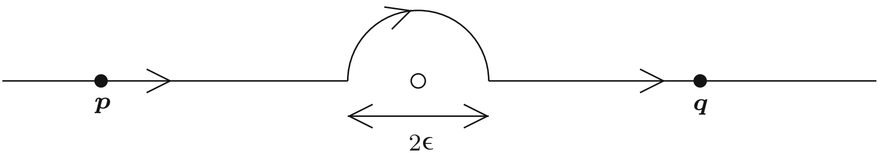
First, if γ is a shortest path on σ from **p** to **q**, then $\mathcal{L}(\tau)$ must have an absolute minimum when $\tau = 0$. This implies that $\frac{d}{d\tau}\mathcal{L}(\tau) = 0$ when $\tau = 0$, and hence by Theorem 9.4.1 that γ is a geodesic.

Second, if γ is a geodesic on σ passing through **p** and **q**, then $\mathcal{L}(\tau)$ has a stationary point (extremum) when $\tau = 0$, but this need not be an absolute minimum, or even a local minimum, so γ need not be a shortest path from **p** to **q**. For example, if **p** and **q** are two nearby points on a sphere, the short great circle arc joining **p** and **q** is the shortest path from **p** to **q** (this is not quite obvious – see below), but the long great circle arc joining **p** and **q** is also a geodesic – see the diagram preceding Proposition 6.5.1.

Third, in general, a shortest path joining two points on a surface may not exist. For example, consider the surface $\mathcal{S}$ consisting of the xy-plane with the origin removed. This is a perfectly good surface, but there is *no* shortest path on the surface from the point $\mathbf{p} = (-1, 0)$ to the point $\mathbf{q} = (1, 0)$. Of course, the shortest path should be the straight line segment joining the two points, but this does not lie entirely on the surface, since it passes through the origin which is not part of the surface. For a 'real life' analogy, imagine trying to walk from $\mathbf{p}$ to $\mathbf{q}$ but finding that there is a deep hole in the ground at the origin. The solution might be to walk in a straight line as long as possible, and then skirt around the hole at the last minute, say taking something like the route shown below. This path consists of two straight line segments of length $1 - \epsilon$, together with a semicircle of radius ϵ, so its total length is

$$2(1 - \epsilon) + \pi\epsilon = 2 + (\pi - 2)\epsilon.$$

Of course, this is greater than the straight line distance 2, but it can be made as close as we like to 2 by taking ϵ sufficiently small. In the language of real analysis, the greatest lower bound of the lengths of curves on the surface joining $\mathbf{p}$ and $\mathbf{q}$ is 2, but there is no curve from $\mathbf{p}$ to $\mathbf{q}$ *in the surface* whose length is equal to this lower bound.



Finally, it can be proved that if a surface $\mathcal{S}$ is a *closed* subset of $\mathbb{R}^3$ (i.e., if the set of points of $\mathbb{R}^3$ that are *not* in $\mathcal{S}$ is an open subset of $\mathbb{R}^3$), and if there is *some* path in $\mathcal{S}$ joining any two points of $\mathcal{S}$, then there is always a shortest path joining any two points of $\mathcal{S}$. For example, a plane is a closed subset of $\mathbb{R}^3$, and so there is a shortest path joining any two points. This path must be a straight line, for by the first remark above it is a geodesic, and we know that the only geodesics on a plane are the straight lines. Similarly, a sphere is a closed subset of $\mathbb{R}^3$, and it follows that the short great circle arc joining two points on the sphere is the shortest path joining them. But the surface $\mathcal{S}$ considered above is *not* a closed subset of $\mathbb{R}^3$, for $(0, 0) \notin \mathcal{S}$, but any open ball containing $(0, 0)$ must clearly contain points of $\mathcal{S}$, and so the set of points not in $\mathcal{S}$ is not open.

Another property of surfaces that are closed subsets of $\mathbb{R}^3$ (that we shall also not prove) is that geodesics on such surfaces can be extended indefinitely, i.e., they can be defined on the whole of $\mathbb{R}$. This is clear for straight lines in the plane, for example, and for great circles on the sphere (although in the latter case the geodesics 'close up' after an increment in the unit-speed parameter equal to the circumference of the sphere). But, for the straight line $\gamma(t) = (t - 1, 0)$ on the surface $\mathcal{S}$ defined above, which passes through $\mathbf{p}$ when $t = 0$, the largest interval containing $t = 0$ on which it is defined as a curve *in the surface* is $(-\infty, 1)$. We encountered a less artificial example of this 'incompleteness' in

Example 9.3.3: the pseudosphere considered there fails to be a closed subset of $\mathbb{R}^3$ because the points of its boundary circle in the xy-plane are not in the surface.

EXERCISES

9.4.1 The geodesics on a circular (half) cone were determined in Exercise 9.2.2. Interpreting 'line' as 'geodesic', which of the following (true) statements in plane Euclidean geometry are true for the cone?

(i) There is a line passing through any two points.

(ii) There is a unique line passing through any two distinct points.

(iii) Any two distinct lines intersect in at most one point.

(iv) There are lines that do not intersect each other.

(v) Any line can be continued indefinitely.

(vi) A line defines the shortest distance between any two of its points.

(vii) A line cannot intersect itself transversely (i.e., with two non-parallel tangent vectors at the point of intersection).

9.4.2 Show that the long great circle arc on S^2 joining the points $\mathbf{p} = (1, 0, 0)$ and $\mathbf{q} = (0, 1, 0)$ is not even a *local* minimum of the length function $\mathcal{L}$ (see the remarks following the proof of Theorem 9.4.1).

9.4.3 Construct a smooth function with the properties in (9.20) in the following steps:

(i) Show that, for all integers n (positive and negative), $t^n e^{-1/t^2}$ tends to 0 as t tends to 0.

(ii) Deduce from (i) that the function

$$\theta(t) = \begin{cases} e^{-1/t^2} & \text{if } t \geq 0, \\ 0 & \text{if } t \leq 0 \end{cases}$$

is smooth everywhere.

(iii) Show that the function

$$\psi(t) = \theta(1 + t)\theta(1 - t)$$

is smooth everywhere, that $\psi(t) > 0$ if $-1 < t < 1$, and that $\psi(t) = 0$ otherwise.

(iv) Show that the function

$$\phi(t) = \psi\left(\frac{t - t_0}{\eta}\right)$$

has the properties we want.

9.5 Geodesic coordinates

The existence of geodesics on a surface $\mathcal{S}$ allows us to construct a very useful atlas for $\mathcal{S}$. For this, let $\mathbf{p} \in \mathcal{S}$ and let $\boldsymbol{\gamma}$, with parameter v say, be a unit-speed geodesic on $\mathcal{S}$ with $\boldsymbol{\gamma}(0) = \mathbf{p}$. For any value of v, let $\tilde{\boldsymbol{\gamma}}^v$, with parameter u, say, be a unit-speed geodesic such that $\tilde{\boldsymbol{\gamma}}^v(0) = \boldsymbol{\gamma}(v)$ and which is perpendicular to $\boldsymbol{\gamma}$ at $\boldsymbol{\gamma}(v)$ ($\tilde{\boldsymbol{\gamma}}^v$ is unique up to the reparametrization $u \mapsto -u$). We define $\boldsymbol{\sigma}(u, v) = \tilde{\boldsymbol{\gamma}}^v(u)$.

Proposition 9.5.1

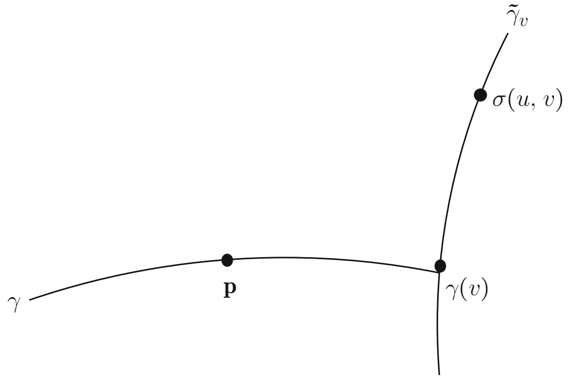
With the above notation, there is an open subset U of $\mathbb{R}^2$ containing $(0, 0)$ such that $\boldsymbol{\sigma} : U \to \mathbb{R}^3$ is an allowable surface patch of $\mathcal{S}$. Moreover, the first fundamental form of $\boldsymbol{\sigma}$ is

$$du^2 + G(u, v)dv^2,$$

where G is a smooth function on U such that

$$G(0, v) = 1, \quad G_u(0, v) = 0,$$

whenever $(0, v) \in U$.



Proof

The proof that $\boldsymbol{\sigma}$ is (for a suitable open set U) an allowable surface patch makes use of the inverse function theorem (see Section 5.6).

Note first that, for any value of v,

$$\boldsymbol{\sigma}_u(0,v) = \frac{d}{du}\tilde{\boldsymbol{\gamma}}^v(u)\bigg|_{u=0}, \quad \boldsymbol{\sigma}_v(0,v) = \frac{d}{dv}\tilde{\boldsymbol{\gamma}}^v(0) = \frac{d}{dv}\boldsymbol{\gamma}(v),$$

and that these are perpendicular unit vectors by construction. If

$$\boldsymbol{\sigma}(u,v) = (f(u,v), g(u,v), h(u,v)),$$

it follows that the Jacobian matrix

$$\begin{pmatrix} f_u & f_v \\ g_u & g_v \\ h_u & h_v \end{pmatrix}$$

has rank 2 when $u = v = 0$. Hence, at least one of its three 2×2 submatrices is invertible at $(0,0)$, say

$$\begin{pmatrix} f_u & f_v \\ g_u & g_v \end{pmatrix}. \tag{9.22}$$

By the Inverse Function Theorem 5.6.1, there is an open subset U of $\mathbb{R}^2$ such that the map given by

$$F(u,v) = (f(u,v), g(u,v))$$

is a bijection from U to an open subset $F(U)$ of $\mathbb{R}^2$, and such that its inverse map $F(U) \to U$ is also smooth. The matrix (9.22) is then invertible for all $(u,v) \in U$, and so $\boldsymbol{\sigma}_u$ and $\boldsymbol{\sigma}_v$ are linearly independent for $(u,v) \in U$. It follows that $\boldsymbol{\sigma} : U \to \mathbb{R}^3$ is a surface patch.

As to the first fundamental form of $\boldsymbol{\sigma}$, note first that

$$E = \| \boldsymbol{\sigma}_u \|^2 = \left\| \frac{d}{du}\tilde{\boldsymbol{\gamma}}^v(u) \right\|^2 = 1$$

because $\tilde{\boldsymbol{\gamma}}^v$ is a unit-speed curve. Next, we apply the second of the geodesic equations (9.2) to $\tilde{\boldsymbol{\gamma}}^v$. The unit-speed parameter is u and v is constant, so we get $F_u = 0$. But when $u = 0$, we have already seen that $\boldsymbol{\sigma}_u$ and $\boldsymbol{\sigma}_v$ are perpendicular, so $F = 0$. It follows that $F = 0$ everywhere. Hence, the first fundamental form of $\boldsymbol{\sigma}$ is

$$du^2 + G(u,v)dv^2.$$

We have

$$G(0,v) = \| \boldsymbol{\sigma}_v(0,v) \|^2 = \left\| \frac{d\boldsymbol{\gamma}}{dv} \right\|^2 = 1$$

because $\boldsymbol{\gamma}$ is unit-speed. Finally, from the first geodesic equation in (9.2) applied to the geodesic $\boldsymbol{\gamma}$, for which $u = 0$ and v is the unit-speed parameter, we get $G_u(0, v) = 0$. □

A surface patch $\boldsymbol{\sigma}$ constructed as above is called a *geodesic patch*, and u and v are called *geodesic coordinates*.

Example 9.5.2

If $\mathbf{p}$ is a point on the equator of the unit sphere S^2, take $\boldsymbol{\gamma}$ to be the equator with parameter the longitude φ, and let $\tilde{\boldsymbol{\gamma}}^{\varphi}$ be the meridian parametrized by latitude θ and passing through the point on the equator with longitude φ. The corresponding geodesic patch is the usual latitude-longitude patch, for which the first fundamental form is

$$d\theta^2 + \cos^2\theta\, d\varphi^2,$$

in accordance with Proposition 9.5.1.

We give an application of geodesic coordinates in the proof of Theorem 10.3.1.

EXERCISES

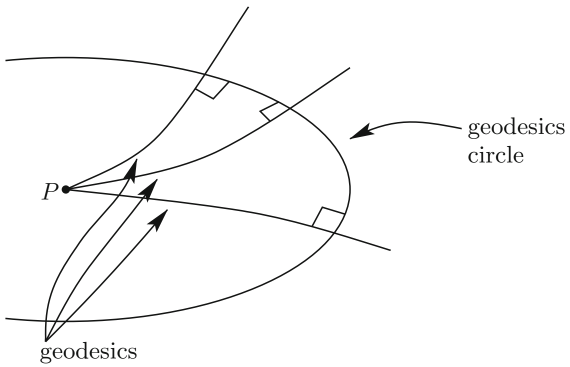
9.5.1 Let P be a point of a surface $\mathcal{S}$ and let $\mathbf{v}$ be a unit tangent vector to $\mathcal{S}$ at P. Let $\boldsymbol{\gamma}^{\theta}(r)$ be the unit-speed geodesic on $\mathcal{S}$ passing through P when $r = 0$ and such that the oriented angle $\widehat{\mathbf{v}\,\frac{d\boldsymbol{\gamma}^{\theta}}{dr}} = \theta$. It can be shown that $\boldsymbol{\sigma}(r, \theta) = \boldsymbol{\gamma}^{\theta}(r)$ is smooth for $-\epsilon < r < \epsilon$ and all values of θ, where ϵ is some positive number, and that it is an allowable surface patch for $\mathcal{S}$ defined for $0 < r < \epsilon$ and for θ in any open interval of length $\leq 2\pi$. This is called a *geodesic polar patch* on $\mathcal{S}$.

Show that, if $0 < R < \epsilon$,

$$\int_0^R \left\| \frac{d\boldsymbol{\gamma}^{\theta}}{dr} \right\|^2 dr = R.$$

By differentiating both sides with respect to θ, prove that

$$\boldsymbol{\sigma}_r \cdot \boldsymbol{\sigma}_{\theta} = 0.$$



This is called *Gauss' Lemma* – geometrically, it means that the parameter curve $r = R$, called the *geodesic circle* with centre P and radius R, is perpendicular to each of its radii, i.e., the geodesics passing through P. Deduce that the first fundamental form of $\boldsymbol{\sigma}$ is

$$dr^2 + G(r, \theta)d\theta^2,$$

for some smooth function $G(r, \theta)$.

9.5.2 Let P and Q be two points on a surface $\mathcal{S}$, and assume that there is a geodesic polar patch with centre P as in Exercise 9.5.1 that also contains Q; suppose that Q is the point $\boldsymbol{\sigma}(R, \alpha)$, where $0 < R < \epsilon$, $0 \le \alpha < 2\pi$. Show in the following steps that the geodesic $\boldsymbol{\gamma}^\alpha(t) = \boldsymbol{\sigma}(t, \alpha)$ is (up to reparametrization) the unique *shortest* curve on $\mathcal{S}$ joining P and Q.

(i) Let $\boldsymbol{\gamma}(t) = \boldsymbol{\sigma}(f(t), g(t))$ be any curve in the patch $\boldsymbol{\sigma}$ joining P and Q. We assume that $\boldsymbol{\gamma}$ passes through P when $t = 0$ and through Q when $t = R$ (this can always be achieved by a suitable reparametrization). Show that the length of the part of $\boldsymbol{\gamma}$ between P and Q is $\ge R$, and that R is the length of the part of $\boldsymbol{\gamma}^\alpha$ between P and Q.

(ii) Show that, if $\boldsymbol{\gamma}$ is *any* curve on $\mathcal{S}$ joining P and Q (not necessarily staying inside the patch $\boldsymbol{\sigma}$), the length of the part of $\boldsymbol{\gamma}$ between P and Q is $\ge R$.

(iii) Show that, if the part of a curve $\boldsymbol{\gamma}$ on $\mathcal{S}$ joining P to Q has length R, then $\boldsymbol{\gamma}$ is a reparametrization of $\boldsymbol{\gamma}^\alpha$.