6
The first fundamental form

Perhaps the first thing that a geometrically inclined bug living on a surface might wish to do is to measure the distance between two points of the surface. Of course, this will usually be different from the distance between these points as measured by an inhabitant of the ambient three-dimensional space, since the straight line segment which furnishes the shortest path between the points in $\mathbb{R}^3$ will generally not be contained in the surface. The object that allows one to compute lengths on a surface, and also angles and areas, is the *first fundamental form* of the surface.

6.1 Lengths of curves on surfaces

If our bug-geometer walks along a curve $\boldsymbol{\gamma}$ on a surface $\mathcal{S}$, the distance he travels is

$$\int \parallel \dot{\boldsymbol{\gamma}}(t) \parallel \, dt$$

(see Definition 1.2.1). To compute this he would need to be able to find the length of *tangent vectors* to the surface, such as $\dot{\boldsymbol{\gamma}}$, which in turn can be computed from the object in the following definition.

Andrew Pressley, *Elementary Differential Geometry: Second Edition*,
Springer Undergraduate Mathematics Series, DOI 10.1007/978-1-84882-891-9_6,

Definition 6.1.1

Let $\mathbf{p}$ be a point of a surface $\mathcal{S}$. The *first fundamental form* of $\mathcal{S}$ at $\mathbf{p}$ associates to tangent vectors $\mathbf{v}, \mathbf{w} \in T_{\mathbf{p}}\mathcal{S}$ the scalar

$$\langle \mathbf{v}, \mathbf{w} \rangle_{\mathbf{p},\mathcal{S}} = \mathbf{v} \cdot \mathbf{w}.$$

Thus, $\langle \mathbf{v}, \mathbf{w} \rangle_{\mathbf{p},\mathcal{S}}$ is just the dot product, but restricted to tangent vectors to $\mathcal{S}$ at $\mathbf{p}$. We shall usually omit one or both of the subscripts unless there is some danger of confusion as to which point or surface is intended.

The first fundamental form $\langle \ , \ \rangle$ is an example of an *inner product* (see Appendix 0): this follows immediately from the fact that the dot product defines an inner product on $\mathbb{R}^3$.

In traditional works on this subject, the first fundamental form looks slightly different. Suppose that $\boldsymbol{\sigma}(u, v)$ is a surface patch of $\mathcal{S}$. Then, any tangent vector to $\mathcal{S}$ at a point $\mathbf{p}$ in the image of $\boldsymbol{\sigma}$ can be expressed uniquely as a linear combination of $\boldsymbol{\sigma}_u$ and $\boldsymbol{\sigma}_v$. Define maps $du : T_{\mathbf{p}}\mathcal{S} \to \mathbb{R}$ and $dv : T_{\mathbf{p}}\mathcal{S} \to \mathbb{R}$ by

$$du(\mathbf{v}) = \lambda, \quad dv(\mathbf{v}) = \mu \qquad \text{if } \mathbf{v} = \lambda\boldsymbol{\sigma}_u + \mu\boldsymbol{\sigma}_v,$$

for some $\lambda, \mu \in \mathbb{R}$. It is easy to see that du and dv are *linear* maps. Then, using the fact that $\langle \ , \ \rangle$ is a symmetric bilinear form, we have

$$\langle \mathbf{v}, \mathbf{v} \rangle = \lambda^2 \langle \boldsymbol{\sigma}_u, \boldsymbol{\sigma}_u \rangle + 2\lambda\mu \langle \boldsymbol{\sigma}_u, \boldsymbol{\sigma}_v \rangle + \mu^2 \langle \boldsymbol{\sigma}_v, \boldsymbol{\sigma}_v \rangle.$$

Writing

$$E = \| \boldsymbol{\sigma}_u \|^2, \quad F = \boldsymbol{\sigma}_u \cdot \boldsymbol{\sigma}_v, \quad G = \| \boldsymbol{\sigma}_v \|^2,$$

this becomes

$$\langle \mathbf{v}, \mathbf{v} \rangle = E\lambda^2 + 2F\lambda\mu + G\mu^2 = Edu(\mathbf{v})^2 + 2Fdu(\mathbf{v})dv(\mathbf{v}) + Gdv(\mathbf{v})^2.$$

Traditionally, the expression

$$Edu^2 + 2Fdudv + Gdv^2$$

is called the first fundamental form of the surface patch $\boldsymbol{\sigma}(u, v)$. Note that the coefficients E, F, G and the linear maps du, dv depend on the choice of surface patch for $\mathcal{S}$ (see Exercise 6.1.4), but the first fundamental form itself depends only on $\mathcal{S}$ and $\mathbf{p}$.

If $\boldsymbol{\gamma}$ is a curve lying in the image of a surface patch $\boldsymbol{\sigma}$, we have

$$\boldsymbol{\gamma}(t) = \boldsymbol{\sigma}(u(t), v(t))$$

for some smooth functions $u(t)$ and $v(t)$. Then, denoting d/dt by a dot, we have $\dot{\boldsymbol{\gamma}} = \dot{u}\boldsymbol{\sigma}_u + \dot{v}\boldsymbol{\sigma}_v$ by the chain rule, so

$$\langle \dot{\boldsymbol{\gamma}}, \dot{\boldsymbol{\gamma}} \rangle = E\dot{u}^2 + 2F\dot{u}\dot{v} + G\dot{v}^2,$$

and the length of $\boldsymbol{\gamma}$ is given by

$$\int (E\dot{u}^2 + 2F\dot{u}\dot{v} + G\dot{v}^2)^{1/2} dt. \tag{6.1}$$

Example 6.1.2

For the plane

$$\boldsymbol{\sigma}(u, v) = \mathbf{a} + u\mathbf{p} + v\mathbf{q}$$

(see Example 4.1.2) with $\mathbf{p}$ and $\mathbf{q}$ being perpendicular unit vectors, we have $\boldsymbol{\sigma}_u = \mathbf{p}$, $\boldsymbol{\sigma}_v = \mathbf{q}$, so $E = \| \boldsymbol{\sigma}_u \|^2 = \| \mathbf{p} \|^2 = 1$, $F = \boldsymbol{\sigma}_u \cdot \boldsymbol{\sigma}_v = \mathbf{p} \cdot \mathbf{q} = 0$, $G = \| \boldsymbol{\sigma}_v \|^2 = \| \mathbf{q} \|^2 = 1$, and the first fundamental form is simply

$$du^2 + dv^2.$$

Example 6.1.3

Consider a surface of revolution

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

Recall from Example 5.3.2 that we can assume that $f(u) > 0$ for all values of u and that the profile curve $u \mapsto (f(u), 0, g(u))$ is unit-speed, i.e., $\dot{f}^2 + \dot{g}^2 = 1$ (a dot denoting d/du). Then:

$$\boldsymbol{\sigma}_u = (\dot{f}\cos v, \dot{f}\sin v, \dot{g}), \quad \boldsymbol{\sigma}_v = (-f\sin v, f\cos v, 0),$$

$$\therefore \quad E = \| \boldsymbol{\sigma}_u \|^2 = \dot{f}^2 + \dot{g}^2 = 1, \ F = \boldsymbol{\sigma}_u \cdot \boldsymbol{\sigma}_v = 0, \ G = \| \boldsymbol{\sigma}_v \|^2 = f^2.$$

So the first fundamental form is

$$du^2 + f(u)^2 dv^2.$$

A special case is the unit sphere S^2 in latitude-longitude coordinates (Example 4.1.4). We take $u = \theta$, $v = \varphi$, $f(\theta) = \cos\theta$, $g(\theta) = \sin\theta$, giving the first fundamental form of S^2 as

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

Example 6.1.4

We consider a generalized cylinder

$$\boldsymbol{\sigma}(u, v) = \boldsymbol{\gamma}(u) + v\mathbf{a}$$

defined in Example 5.3.1. As we saw in Exercise 5.3.3, we can assume that $\boldsymbol{\gamma}$ is unit-speed, $\mathbf{a}$ is a unit vector, and $\boldsymbol{\gamma}$ is contained in a plane perpendicular to $\mathbf{a}$. Then, denoting d/du by a dot, $\boldsymbol{\sigma}_u = \dot{\boldsymbol{\gamma}}$, $\boldsymbol{\sigma}_v = \mathbf{a}$, so $E = \| \boldsymbol{\sigma}_u \|^2 = \| \dot{\boldsymbol{\gamma}} \|^2 = 1$,

$F = \boldsymbol{\sigma}_u \cdot \boldsymbol{\sigma}_v = \dot{\boldsymbol{\gamma}} \cdot \mathbf{a} = 0$, $G = \parallel \boldsymbol{\sigma}_v \parallel^2 = \parallel \mathbf{a} \parallel^2 = 1$, and the first fundamental form of $\boldsymbol{\sigma}$ is

$$du^2 + dv^2.$$

Note that this is the *same* as the first fundamental form of the plane (Example 6.1.2). The geometrical reason for this coincidence will be revealed in the next section.

Example 6.1.5

We consider a generalized cone

$$\boldsymbol{\sigma}(u, v) = (1 + v)\boldsymbol{\gamma}(u) - v\mathbf{v}$$

(Example 5.3.1). Before computing its first fundamental form, we make some simplifications to $\boldsymbol{\sigma}$.

First, translating the surface by $\mathbf{v}$ (which does not change its first fundamental form – see Exercise 6.1.2), we get the surface patch $\boldsymbol{\sigma}_1 = \boldsymbol{\sigma} - \mathbf{v} = (1 + v)(\boldsymbol{\gamma} - \mathbf{v})$, so if we replace $\boldsymbol{\gamma}$ by $\boldsymbol{\gamma}_1 = \boldsymbol{\gamma} - \mathbf{v}$ we get $\boldsymbol{\sigma}_1 = (1 + v)\boldsymbol{\gamma}_1$. This means that we might as well assume that $\mathbf{v} = \mathbf{0}$ to begin with. Next, we saw in Example 5.3.1 that for $\boldsymbol{\sigma}$ to be a regular surface patch, $\boldsymbol{\gamma}$ must not pass through the origin, so we can define a new curve $\tilde{\boldsymbol{\gamma}}$ by $\tilde{\boldsymbol{\gamma}}(u) = \boldsymbol{\gamma}(u)/ \parallel \boldsymbol{\gamma}(u) \parallel$. Setting $\tilde{u} = u$, $\tilde{v} = (1 + v)/ \parallel \boldsymbol{\gamma}(u) \parallel$, we get a reparametrization $\tilde{\boldsymbol{\sigma}}(\tilde{u}, \tilde{v}) = \tilde{v}\tilde{\boldsymbol{\gamma}}(\tilde{u})$ of $\boldsymbol{\sigma}$ with $\parallel \tilde{\boldsymbol{\gamma}} \parallel = 1$. We can therefore assume to begin with that $\boldsymbol{\sigma}(u, v) = v\boldsymbol{\gamma}(u)$ with $\parallel \boldsymbol{\gamma}(u) \parallel = 1$ for all values of u (geometrically, this means that we can replace $\boldsymbol{\gamma}$ by the intersection of the cone with S^2). Finally, reparametrizing again, we can assume that $\boldsymbol{\gamma}$ is unit-speed, for we saw in Example 5.3.1 that for $\boldsymbol{\sigma}$ to be regular, $\boldsymbol{\gamma}$ must be regular.

With these assumptions, and with a dot denoting d/du, we have $\boldsymbol{\sigma}_u = v\dot{\boldsymbol{\gamma}}$, $\boldsymbol{\sigma}_v = \boldsymbol{\gamma}$, so $E = \parallel v\dot{\boldsymbol{\gamma}} \parallel^2 = v^2 \parallel \dot{\boldsymbol{\gamma}} \parallel^2 = v^2$, $F = v\dot{\boldsymbol{\gamma}} \cdot \boldsymbol{\gamma} = 0$ (since $\parallel \boldsymbol{\gamma} \parallel = 1$), $G = \parallel \boldsymbol{\gamma} \parallel^2 = 1$, and the first fundamental form is

$$v^2 du^2 + dv^2.$$

Note that, as for the generalized cylinder in Example 6.1.4, the first fundamental form of the generalized cone does not depend on the curve $\boldsymbol{\gamma}$.

EXERCISES

6.1.1 Calculate the first fundamental forms of the following surfaces:

(i) $\boldsymbol{\sigma}(u, v) = (\sinh u \sinh v, \sinh u \cosh v, \sinh u)$.

(ii) $\boldsymbol{\sigma}(u, v) = (u - v, u + v, u^2 + v^2)$.

(iii) $\boldsymbol{\sigma}(u,v) = (\cosh u, \sinh u, v)$.

(iv) $\boldsymbol{\sigma}(u,v) = (u, v, u^2 + v^2)$.

What kinds of surfaces are these?

6.1.2 Show that applying an isometry of $\mathbb{R}^3$ to a surface does not change its first fundamental form. What is the effect of a dilation (i.e., a map $\mathbb{R}^3 \to \mathbb{R}^3$ of the form $\mathbf{v} \mapsto a\mathbf{v}$ for some constant $a \neq 0$)?

6.1.3 Let $Edu^2+2Fdudv+Gdv^2$ be the first fundamental form of a surface patch $\boldsymbol{\sigma}(u,v)$ of a surface $\mathcal{S}$. Show that, if $\mathbf{p}$ is a point in the image of $\boldsymbol{\sigma}$ and $\mathbf{v}, \mathbf{w} \in T_{\mathbf{p}}\mathcal{S}$, then

$$\langle \mathbf{v}, \mathbf{w} \rangle = Edu(\mathbf{v})du(\mathbf{w}) + F(du(\mathbf{v})dv(\mathbf{w}) + du(\mathbf{w})dv(\mathbf{v})) + Gdv(\mathbf{w})dv(\mathbf{w}).$$

6.1.4 Suppose that a surface patch $\tilde{\boldsymbol{\sigma}}(\tilde{u}, \tilde{v})$ is a reparametrization of a surface patch $\boldsymbol{\sigma}(u,v)$, and let

$$\tilde{E}d\tilde{u}^2 + 2\tilde{F}d\tilde{u}d\tilde{v} + \tilde{G}d\tilde{v}^2 \quad \text{and} \quad Edu^2 + 2Fdudv + Gdv^2$$

be their first fundamental forms. Show that:

(i) $du = \frac{\partial u}{\partial \tilde{u}} d\tilde{u} + \frac{\partial u}{\partial \tilde{v}} d\tilde{v}, \quad dv = \frac{\partial v}{\partial \tilde{u}} d\tilde{u} + \frac{\partial v}{\partial \tilde{v}} d\tilde{v}$.

(ii) If

$$J = \begin{pmatrix} \dfrac{\partial u}{\partial \tilde{u}} & \dfrac{\partial u}{\partial \tilde{v}} \\ \dfrac{\partial v}{\partial \tilde{u}} & \dfrac{\partial v}{\partial \tilde{v}} \end{pmatrix}$$

is the Jacobian matrix of the reparametrization map $(\tilde{u}, \tilde{v}) \mapsto (u, v)$, and J^t is the transpose of J, then

$$\begin{pmatrix} \tilde{E} & \tilde{F} \\ \tilde{F} & \tilde{G} \end{pmatrix} = J^t \begin{pmatrix} E & F \\ F & G \end{pmatrix} J.$$

6.1.5 Show that the following are equivalent conditions on a surface patch $\boldsymbol{\sigma}(u,v)$ with first fundamental form $Edu^2 + 2Fdudv + Gdv^2$:

(i) $E_v = G_u = 0$.

(ii) $\boldsymbol{\sigma}_{uv}$ is parallel to the standard unit normal $\mathbf{N}$.

(iii) The opposite sides of any quadrilateral formed by parameter curves of $\boldsymbol{\sigma}$ have the same length (see the remarks following the proof of Proposition 4.4.2).

When these conditions are satisfied, the parameter curves of $\boldsymbol{\sigma}$ are said to form a *Chebyshev net*. Show that, in that case, $\boldsymbol{\sigma}$ has a reparametrization $\tilde{\boldsymbol{\sigma}}(\tilde{u}, \tilde{v})$ with first fundamental form

$$d\tilde{u}^2 + 2\cos\theta \, d\tilde{u}d\tilde{v} + d\tilde{v}^2,$$

where θ is a smooth function of $(\tilde{u}, \tilde{v})$. Show that θ is the angle between the parameter curves of $\tilde{\boldsymbol{\sigma}}$. Show further that, if we put $\hat{u} = \tilde{u} + \tilde{v}$, $\hat{v} = \tilde{u} - \tilde{v}$, the resulting reparametrization $\hat{\boldsymbol{\sigma}}(\hat{u}, \hat{v})$ of $\tilde{\boldsymbol{\sigma}}(\tilde{u}, \tilde{v})$ has first fundamental form

$$\cos^2\omega \, d\hat{u}^2 + \sin^2\omega \, d\hat{v}^2,$$

where $\omega = \theta/2$.

6.2 Isometries of surfaces

We observed in Example 6.1.4 that a plane and a generalized cylinder, when suitably parametrized, have the *same* first fundamental form. The geometric reason for this is not hard to see. A plane piece of paper can be 'wrapped' on a cylinder in the obvious way without crumpling the paper (see Example 4.3.2). If we draw a curve on the plane, then after wrapping it becomes a curve on the cylinder. Because there is no crumpling, the lengths of these two curves will be the same. Since the lengths are computed as the integral of (the square root of) the first fundamental form, it is plausible that the first fundamental forms of the two surfaces should be the same. Experiment suggests, on the other hand, that it is impossible to wrap a plane sheet of paper around a sphere without crumpling. Thus, we expect that a plane and a sphere do not have the same first fundamental form.

The following definition makes precise what it means to wrap one surface onto another without crumpling.

Definition 6.2.1

If $\mathcal{S}_1$ and $\mathcal{S}_2$ are surfaces, a smooth map $f : \mathcal{S}_1 \to \mathcal{S}_2$ is called a *local isometry* if it takes any curve in $\mathcal{S}_1$ to a curve of *the same length* in $\mathcal{S}_2$. If a local isometry $f : \mathcal{S}_1 \to \mathcal{S}_2$ exists, we say that $\mathcal{S}_1$ and $\mathcal{S}_2$ are *locally isometric*.

We shall see that every local isometry is a local diffeomorphism; a local isometry that is a diffeomorphism is called an *isometry*. It is obvious that any composite of local isometries is a local isometry, and that the inverse of any isometry is an isometry.

To express the condition for a local isometry in a more useful form, we need the following construction. Let $f : \mathcal{S}_1 \to \mathcal{S}_2$ be a smooth map and let $\mathbf{p} \in \mathcal{S}_1$. For $\mathbf{v}, \mathbf{w} \in T_{\mathbf{p}}\mathcal{S}_1$, define

$$f^*\langle \mathbf{v}, \mathbf{w} \rangle_{\mathbf{p}} = \langle D_{\mathbf{p}}f(\mathbf{v}), D_{\mathbf{p}}f(\mathbf{w}) \rangle_{f(\mathbf{p})}.$$

Then, $f^*\langle \ , \ \rangle_{\mathbf{p}}$ is a symmetric bilinear form on $T_{\mathbf{p}}\mathcal{S}_1$. Indeed, the symmetry is obvious and if $\lambda, \lambda' \in \mathbb{R}$, $\mathbf{v}, \mathbf{v}', \mathbf{w} \in T_{\mathbf{p}}$,

$$\begin{aligned} f^*\langle \lambda\mathbf{v} + \lambda'\mathbf{v}', \mathbf{w} \rangle_{\mathbf{p}} &= \langle D_{\mathbf{p}}f(\lambda\mathbf{v} + \lambda'\mathbf{v}'), D_{\mathbf{p}}f(\mathbf{w}) \rangle_{f(\mathbf{p})} \\ &= \langle \lambda D_{\mathbf{p}}f(\mathbf{v}) + \lambda' D_{\mathbf{p}}f(\mathbf{v}'), D_{\mathbf{p}}f(\mathbf{w}) \rangle_{f(\mathbf{p})} \\ &= \lambda \langle D_{\mathbf{p}}f(\mathbf{v}), D_{\mathbf{p}}f(\mathbf{w}) \rangle_{f(\mathbf{p})} + \lambda' \langle D_{\mathbf{p}}f(\mathbf{v}'), D_{\mathbf{p}}f(\mathbf{w}) \rangle_{f(\mathbf{p})} \\ &= \lambda f^*\langle \mathbf{v}, \mathbf{w} \rangle_{\mathbf{p}} + \lambda' f^*\langle \mathbf{v}', \mathbf{w} \rangle_{\mathbf{p}}. \end{aligned}$$

Theorem 6.2.2

A smooth map $f : \mathcal{S}_1 \to \mathcal{S}_2$ is a local isometry if and only if the symmetric bilinear forms $\langle \ , \ \rangle_{\mathbf{p}}$ and $f^*\langle \ . \ \rangle_{\mathbf{p}}$ on $T_{\mathbf{p}}\mathcal{S}_1$ are equal for all $\mathbf{p} \in \mathcal{S}_1$.

Proof

If $\boldsymbol{\gamma}_1$ is a curve on $\mathcal{S}_1$, the length of the part of $\boldsymbol{\gamma}_1$ with endpoints $\boldsymbol{\gamma}_1(t_0)$ and $\boldsymbol{\gamma}_1(t_1)$ is

$$\int_{t_0}^{t_1} \langle \dot{\boldsymbol{\gamma}}_1, \dot{\boldsymbol{\gamma}}_1 \rangle^{1/2} dt. \tag{6.2}$$

The length of the corresponding part of the curve $\boldsymbol{\gamma}_2 = f \circ \boldsymbol{\gamma}_1$ on $\mathcal{S}_2$ is

$$\int_{t_0}^{t_1} \langle \dot{\boldsymbol{\gamma}}_2, \dot{\boldsymbol{\gamma}}_2 \rangle^{1/2} dt = \int_{t_0}^{t_1} \langle Df(\dot{\boldsymbol{\gamma}}_1), Df(\dot{\boldsymbol{\gamma}}_1) \rangle^{1/2} dt = \int_{t_0}^{t_1} f^*\langle \dot{\boldsymbol{\gamma}}_1, \dot{\boldsymbol{\gamma}}_1 \rangle^{1/2} dt. \tag{6.3}$$

It is now obvious that, if the two symmetric bilinear forms in the statement of the theorem are equal, the curves $\boldsymbol{\gamma}_1$ and $f \circ \boldsymbol{\gamma}_1$ have the same length.

Conversely, suppose that the integrals in (6.2) and (6.3) are equal for all curves $\boldsymbol{\gamma}$ on $\mathcal{S}_1$. Then, the integrands must be the same for all $\boldsymbol{\gamma}$:

$$\langle \dot{\boldsymbol{\gamma}}, \dot{\boldsymbol{\gamma}} \rangle = f^*\langle \dot{\boldsymbol{\gamma}}, \dot{\boldsymbol{\gamma}} \rangle.$$

Since any tangent vector $\mathbf{v}$ to $\mathcal{S}_1$ is the tangent vector of a curve on $\mathcal{S}_1$, it follows that

$$\langle \mathbf{v}, \mathbf{v} \rangle = f^*\langle \mathbf{v}, \mathbf{v} \rangle \quad \text{for all } \mathbf{v}. \tag{6.4}$$

Since $\langle\ ,\ \rangle$ and $f^*\langle\ ,\ \rangle$ are symmetric bilinear forms, it follows from (6.4) that they are equal (see Appendix 0). $\square$

Thus, f is a local isometry if and only if

$$\langle D_{\mathbf{p}}f(\mathbf{v}), D_{\mathbf{p}}f(\mathbf{w})\rangle_{f(\mathbf{p})} = \langle \mathbf{v}, \mathbf{w}\rangle_{\mathbf{p}}$$

for all $\mathbf{p} \in \mathcal{S}_1$ and all $\mathbf{v}, \mathbf{w} \in T_{\mathbf{p}}\mathcal{S}_1$. This means that the linear map $D_{\mathbf{p}}f : T_{\mathbf{p}}\mathcal{S}_1 \to T_{f(\mathbf{p})}\mathcal{S}_2$ is an *isometry*, i.e., it preserves lengths (see Appendix 1). In short, f is a local isometry if and only if $D_{\mathbf{p}}f$ is an isometry for all $\mathbf{p} \in \mathcal{S}_1$.

It follows from this theorem that every local isometry is a local diffeomorphism. Indeed, let $f : \mathcal{S}_1 \to \mathcal{S}_2$ be a local isometry and let $\mathbf{p} \in \mathcal{S}_1$. If $D_{\mathbf{p}}f$ is not invertible, there is a non-zero tangent vector $\mathbf{v} \in T_{\mathbf{p}}\mathcal{S}_1$ such that $D_{\mathbf{p}}f(\mathbf{v}) = \mathbf{0}$. But this gives a contradiction: since f is a local isometry,

$$0 \neq \langle \mathbf{v}, \mathbf{v}\rangle_{\mathbf{p}} = \langle D_{\mathbf{p}}f(\mathbf{v}), D_{\mathbf{p}}f(\mathbf{v})\rangle_{f(\mathbf{p})} = \langle \mathbf{0}, \mathbf{0}\rangle_{\mathbf{p}} = 0.$$

Hence, $D_{\mathbf{p}}f$ is invertible, and so f is a local diffeomorphism (Proposition 4.4.6).

It will be useful to express Theorem 6.2.2 in terms of surface patches.

Corollary 6.2.3

A local diffeomorphism $f : \mathcal{S}_1 \to \mathcal{S}_2$ is a local isometry if and only if, for any surface patch $\boldsymbol{\sigma}_1$ of $\mathcal{S}_1$, the patches $\boldsymbol{\sigma}_1$ and $f \circ \boldsymbol{\sigma}_1$ of $\mathcal{S}_1$ and $\mathcal{S}_2$, respectively, have the same first fundamental form.

Proof

In view of the theorem, we have to show that the patches $\boldsymbol{\sigma}_1$ and $f \circ \boldsymbol{\sigma}_1 = \boldsymbol{\sigma}_2$, say, have the same first fundamental form if and only if the symmetric bilinear forms $\langle\ ,\ \rangle_{\mathbf{p}}$ and $f^*\langle\ ,\ \rangle_{\mathbf{p}}$ are equal for all $\mathbf{p}$ in the image of $\boldsymbol{\sigma}_1$.

The first fundamental form of $\boldsymbol{\sigma}_i$ $(i = 1, 2)$ is $E_i du^2 + 2F_i du dv + G_i dv^2$, where $E_i = \langle(\boldsymbol{\sigma}_i)_u, (\boldsymbol{\sigma}_i)_u\rangle$, $F_i = \langle(\boldsymbol{\sigma}_i)_u, (\boldsymbol{\sigma}_i)_v\rangle$, $G_i = \langle(\boldsymbol{\sigma}_i)_v, (\boldsymbol{\sigma}_i)_v\rangle$. We compute

$$\langle(\boldsymbol{\sigma}_2)_u, (\boldsymbol{\sigma}_2)_u\rangle = \langle Df((\boldsymbol{\sigma}_1)_u), Df((\boldsymbol{\sigma}_1)_u)\rangle = f^*\langle(\boldsymbol{\sigma}_1)_u, (\boldsymbol{\sigma}_1)_u\rangle.$$

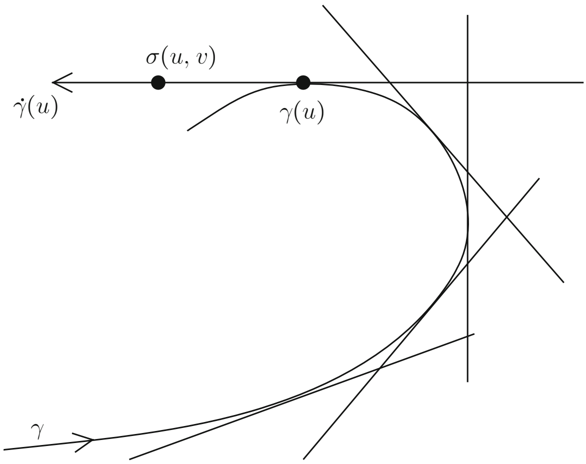
Thus, if $\langle\ ,\ \rangle = f^*\langle\ .\ \rangle$, then $E_1 = E_2$, and similarly $F_1 = F_2$ and $G_1 = G_2$. Conversely, if these last three equations hold, then $\langle \mathbf{v}, \mathbf{w}\rangle = f^*\langle \mathbf{v}, \mathbf{w}\rangle$ whenever the tangent vectors $\mathbf{v}, \mathbf{w}$ are of the form $(\boldsymbol{\sigma}_1)_u$ or $(\boldsymbol{\sigma}_1)_v$. The bilinearity property then implies that $\langle \mathbf{v}, \mathbf{w}\rangle = f^*\langle \mathbf{v}, \mathbf{w}\rangle$ for all $\mathbf{v}, \mathbf{w}$. $\square$

This proof actually shows that, if $\mathbf{p} \in \mathcal{S}_1$ is in the image of a surface patch $\boldsymbol{\sigma}_1$, then $\boldsymbol{\sigma}_1$ and $f \circ \boldsymbol{\sigma}_1$ have the same first fundamental form at $\mathbf{p}$ if and only if $D_{\mathbf{p}}f$ is an isometry; it follows that, if $\mathbf{p}$ is in the image of another surface patch $\boldsymbol{\sigma}_2$, then $\boldsymbol{\sigma}_1$ and $f \circ \boldsymbol{\sigma}_1$ have the same first fundamental form at $\mathbf{p}$ if and only if the same is true of $\boldsymbol{\sigma}_2$ and $f \circ \boldsymbol{\sigma}_2$.

Example 6.2.4

The map f from the yz-plane to the unit cylinder defined in Example 4.3.2 is a local isometry. For, if we use the surface patch $\boldsymbol{\sigma}_1(u,v) = (0,u,v)$ for the plane and $\boldsymbol{\sigma}_2(u,v) = (\cos u, \sin u, v)$ for the cylinder, then $f(\boldsymbol{\sigma}_1(u,v)) = \boldsymbol{\sigma}_2(u,v)$, and by Example 6.1.4 $\boldsymbol{\sigma}_1$ and $\boldsymbol{\sigma}_2$ have the same first fundamental form.

A similar argument shows that a generalized cone is locally isometric to a plane (see Example 6.2.1). It turns out that there is another class of surfaces that is locally isometric to a plane, called *tangent developables*. (In older works, a 'development' of one surface on another was the term used for a local isometry.) A tangent developable is the union of the tangent lines to a curve in $\mathbb{R}^3$ – the tangent line to a curve $\boldsymbol{\gamma}$ at a point $\boldsymbol{\gamma}(u)$ is the straight line passing through $\boldsymbol{\gamma}(u)$ and parallel to the tangent vector $\dot{\boldsymbol{\gamma}}(u)$.



We might as well assume that $\boldsymbol{\gamma}$ is unit-speed. The most general point on the tangent line at $\boldsymbol{\gamma}(u)$ is

$$\boldsymbol{\sigma}(u,v) = \boldsymbol{\gamma}(u) + v\dot{\boldsymbol{\gamma}}(u),$$

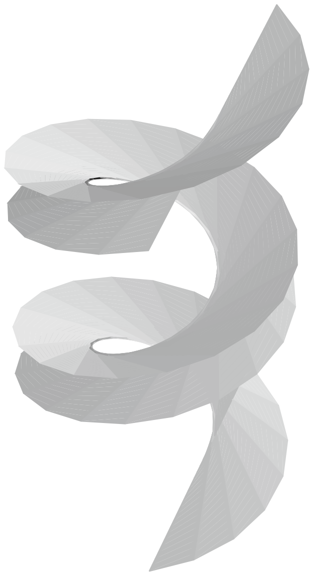
for some scalar v. Now

$$\boldsymbol{\sigma}_u \times \boldsymbol{\sigma}_v = (\dot{\boldsymbol{\gamma}} + v\ddot{\boldsymbol{\gamma}}) \times \dot{\boldsymbol{\gamma}} = v\ddot{\boldsymbol{\gamma}} \times \dot{\boldsymbol{\gamma}}.$$

For $\boldsymbol{\sigma}$ to be regular, it is thus necessary that $\ddot{\boldsymbol{\gamma}}$ is never zero, or in other words, the curvature $\kappa = \parallel \ddot{\boldsymbol{\gamma}} \parallel$ is > 0 at all points of $\boldsymbol{\gamma}$. Now, $\dot{\boldsymbol{\gamma}} = \mathbf{t}$, the unit tangent vector of $\boldsymbol{\gamma}$, and $\ddot{\boldsymbol{\gamma}} = \dot{\mathbf{t}} = \kappa\mathbf{n}$, where $\mathbf{n}$ is the principal normal to $\boldsymbol{\gamma}$, so

$$\boldsymbol{\sigma}_u \times \boldsymbol{\sigma}_v = \kappa v \mathbf{n} \times \mathbf{t} = -\kappa v \mathbf{b},$$

where $\mathbf{b}$ is the binormal of $\boldsymbol{\gamma}$. Thus, $\boldsymbol{\sigma}$ will be regular if $\kappa > 0$ everywhere and $v \neq 0$. The latter condition means that, for regularity, we must exclude the curve $\boldsymbol{\gamma}$ itself from the surface. Typically, the regions $v > 0$ and $v < 0$ of the tangent developable form two sheets which meet along a sharp edge formed by the curve $\boldsymbol{\gamma}$ where $v = 0$, as the following illustration of the tangent developable of a circular helix indicates (see Exercise 6.2.4):



Our interest in tangent developables stems from the following result.

Proposition 6.2.5

Any tangent developable is locally isometric to a plane.

Proof

We use the above notation, assuming that $\boldsymbol{\gamma}$ is unit-speed and that $\kappa > 0$. Now,

$$\begin{aligned} E &= \| \boldsymbol{\sigma}_u \|^2 = (\dot{\boldsymbol{\gamma}} + v\ddot{\boldsymbol{\gamma}}) \cdot (\dot{\boldsymbol{\gamma}} + v\ddot{\boldsymbol{\gamma}}) = \dot{\boldsymbol{\gamma}} \cdot \dot{\boldsymbol{\gamma}} + 2v\dot{\boldsymbol{\gamma}} \cdot \ddot{\boldsymbol{\gamma}} + v^2\ddot{\boldsymbol{\gamma}} \cdot \ddot{\boldsymbol{\gamma}} = 1 + v^2\kappa^2, \\ F &= \boldsymbol{\sigma}_u \cdot \boldsymbol{\sigma}_v = (\dot{\boldsymbol{\gamma}} + v\ddot{\boldsymbol{\gamma}}) \cdot \dot{\boldsymbol{\gamma}} = \dot{\boldsymbol{\gamma}} \cdot \dot{\boldsymbol{\gamma}} + v\dot{\boldsymbol{\gamma}} \cdot \ddot{\boldsymbol{\gamma}} = 1, \\ G &= \| \boldsymbol{\sigma}_v \|^2 = \dot{\boldsymbol{\gamma}} \cdot \dot{\boldsymbol{\gamma}} = 1, \end{aligned}$$

since $\dot{\boldsymbol{\gamma}} \cdot \dot{\boldsymbol{\gamma}} = 1$, $\dot{\boldsymbol{\gamma}} \cdot \ddot{\boldsymbol{\gamma}} = 0$, $\ddot{\boldsymbol{\gamma}} \cdot \ddot{\boldsymbol{\gamma}} = \kappa^2$. So the first fundamental form of the tangent developable is

$$(1 + v^2\kappa^2)du^2 + 2dudv + dv^2. \tag{6.5}$$

We are going to show that an open subset of the plane can be parametrized so that it has the same first fundamental form. This will prove the proposition.

By Theorem 2.2.5, there is a *plane* unit-speed curve $\tilde{\boldsymbol{\gamma}}$ whose curvature is κ (we can even assume that its signed curvature is κ). By the above calculations, the first fundamental form of the tangent developable of $\tilde{\boldsymbol{\gamma}}$ is also given by (6.5).

But since $\tilde{\gamma}$ is a plane curve, its tangent lines obviously fill out part of the plane in which $\tilde{\gamma}$ lies. □

There is a converse to Proposition 6.2.5: any sufficiently small open subset of a surface locally isometric to a plane *is* an open subset of a plane, a generalized cylinder, a generalized cone or a tangent developable. The proof of this will be given in Section 8.4.

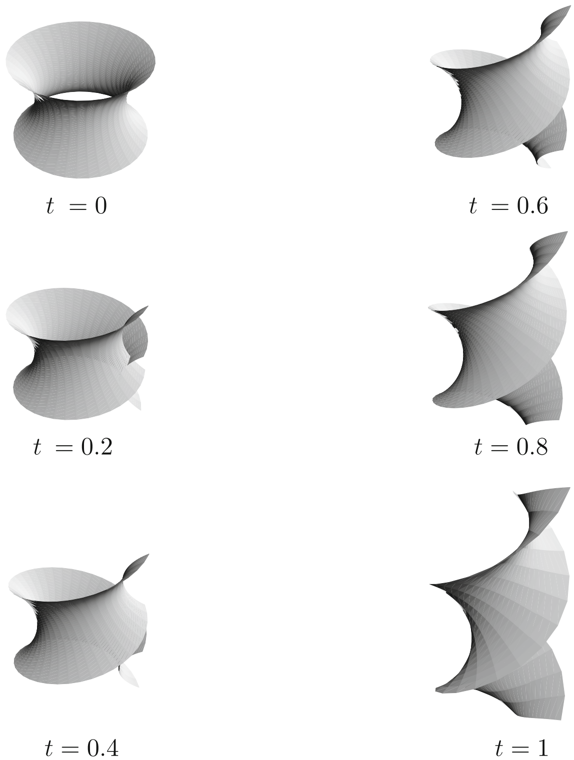
EXERCISES

6.2.1 By thinking about how a circular cone can be 'unwrapped' onto the plane, write down an isometry from

$$\boldsymbol{\sigma}(u,v) = (u\cos v, u\sin v, u), \qquad u > 0,\ 0 < v < 2\pi,$$

(a circular half-cone with a straight line removed) to an open subset of the xy-plane.

6.2.2 Is the map from the circular half-cone $x^2 + y^2 = z^2$, $z > 0$, to the xy-plane given by $(x, y, z) \mapsto (x, y, 0)$ a local isometry?



6.2.3 Consider the surface patches

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

parametrizing the catenoid (Exercise 5.3.1) and the helicoid (Exercise 4.2.6), respectively. Show that the map from the catenoid to the helicoid that takes $\boldsymbol{\sigma}(u,v)$ to $\tilde{\boldsymbol{\sigma}}(\sinh u, v)$ is a local isometry. Which curves on the helicoid correspond under this isometry to the parallels and meridians of the catenoid?
In fact, there is an *isometric deformation* of the catenoid into a helicoid. Let

$$\hat{\boldsymbol{\sigma}}(u,v) = (-\sinh u \sin v, \sinh u \cos v, -v).$$

This is the result of reflecting the helicoid $\tilde{\boldsymbol{\sigma}}$ in the xy-plane and then translating it by $\pi/2$ parallel to the z-axis. Define

$$\boldsymbol{\sigma}^t(u,v) = \cos t\, \boldsymbol{\sigma}(u,v) + \sin t\, \hat{\boldsymbol{\sigma}}(u,v),$$

so that $\boldsymbol{\sigma}^0(u,v) = \boldsymbol{\sigma}(u,v)$ and $\boldsymbol{\sigma}^{\pi/2}(u,v) = \hat{\boldsymbol{\sigma}}(u,v)$. Show that, for all values of t, the map $\boldsymbol{\sigma}(u,v) \mapsto \boldsymbol{\sigma}^t(u,v)$ is a local isometry. Show also that the tangent plane of $\boldsymbol{\sigma}^t$ at the point $\boldsymbol{\sigma}^t(u,v)$ depends only on u, v and not on t. The surfaces $\boldsymbol{\sigma}^t$ are shown above for several values of t. (The result of this exercise is 'explained' in Exercises 12.5.3 and 12.5.4.)

6.2.4 Show that the line of striction (Exercise 5.3.4) of the tangent developable of a unit-speed curve $\boldsymbol{\gamma}$ is $\boldsymbol{\gamma}$ itself. Show also that the intersection of this surface with the plane passing through a point $\boldsymbol{\gamma}(u_0)$ of the curve and perpendicular to it at that point is a curve of the form

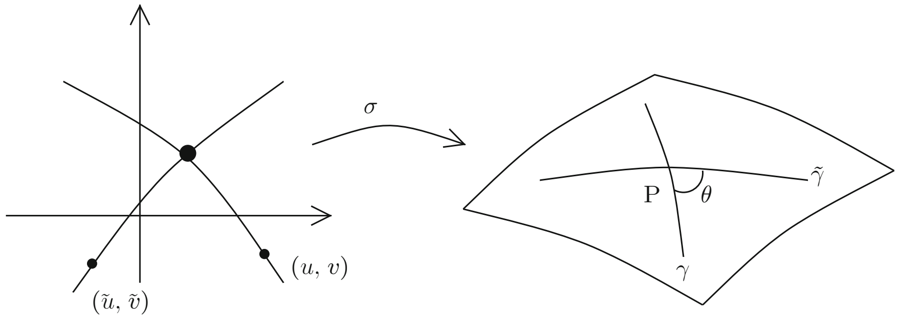
$$\boldsymbol{\Gamma}(v) = \boldsymbol{\gamma}(u_0) - \frac{1}{2}\kappa(u_0)v^2\mathbf{n}(u_0) + \frac{1}{3}\kappa(u_0)\tau(u_0)v^3\mathbf{b}(u_0)$$

if we neglect higher powers of v (we assume that the curvature $\kappa(u_0)$ and the torsion $\tau(u_0)$ of $\boldsymbol{\gamma}$ at $\boldsymbol{\gamma}(u_0)$ are both non-zero). Note that this curve has an ordinary cusp (Exercise 1.3.3) at $\boldsymbol{\gamma}(u_0)$, so the tangent developable has a sharp 'edge' where the two sheets $v > 0$ and $v < 0$ meet along $\boldsymbol{\gamma}$. This is evident for the tangent developable of a circular helix illustrated earlier in this section.

6.3 Conformal mappings of surfaces

Now that we understand how to measure lengths of curves on surfaces, it is natural to ask about angles. Suppose that two curves $\boldsymbol{\gamma}$ and $\tilde{\boldsymbol{\gamma}}$ on a surface $\mathcal{S}$ intersect at a point $\mathbf{p}$. The *angle* θ of intersection of $\boldsymbol{\gamma}$ and $\tilde{\boldsymbol{\gamma}}$ at $\mathbf{p}$ is defined to be the angle between the tangent vectors $\dot{\boldsymbol{\gamma}}$ and $\dot{\tilde{\boldsymbol{\gamma}}}$ (evaluated at $t = t_0$ and $t = \tilde{t}_0$, respectively). Using the dot product formula for the angle between vectors, we see that θ is given by

$$\cos\theta = \frac{\dot{\boldsymbol{\gamma}} \cdot \dot{\tilde{\boldsymbol{\gamma}}}}{\| \dot{\boldsymbol{\gamma}} \| \| \dot{\tilde{\boldsymbol{\gamma}}} \|} = \frac{\langle \dot{\boldsymbol{\gamma}}, \dot{\tilde{\boldsymbol{\gamma}}} \rangle}{\langle \dot{\boldsymbol{\gamma}}, \dot{\boldsymbol{\gamma}} \rangle^{1/2} \langle \dot{\tilde{\boldsymbol{\gamma}}}, \dot{\boldsymbol{\gamma}} \rangle^{1/2}}. \tag{6.6}$$



As usual, it will be useful to have an expression for this in terms of a surface patch. Suppose then that $\boldsymbol{\gamma}$ and $\tilde{\boldsymbol{\gamma}}$ lie in a surface patch $\boldsymbol{\sigma}$ of $\mathcal{S}$, so that $\boldsymbol{\gamma}(t) = \boldsymbol{\sigma}(u(t), v(t))$ and $\tilde{\boldsymbol{\gamma}}(t) = \boldsymbol{\sigma}(\tilde{u}(t), \tilde{v}(t))$ for some smooth functions $u, v, \tilde{u}$ and $\tilde{v}$. If $Edu^2 + 2Fdudv + Gdv^2$ is the first fundamental form of $\boldsymbol{\sigma}$, then by (6.6) we have

$$\cos\theta = \frac{E\dot{u}\dot{\tilde{u}} + F(\dot{u}\dot{\tilde{v}} + \dot{\tilde{u}}\dot{v}) + G\dot{v}\dot{\tilde{v}}}{(E\dot{u}^2 + 2F\dot{u}\dot{v} + G\dot{v}^2)^{1/2}(E\dot{\tilde{u}}^2 + 2F\dot{\tilde{u}}\dot{\tilde{v}} + G\dot{\tilde{v}}^2)^{1/2}}. \tag{6.7}$$

Example 6.3.1

The *parameter curves* on a surface patch $\boldsymbol{\sigma}(u, v)$ can be parametrized by

$$\boldsymbol{\gamma}(t) = \boldsymbol{\sigma}(u_0, t), \quad \tilde{\boldsymbol{\gamma}}(t) = \boldsymbol{\sigma}(t, v_0),$$

respectively, where u_0 is the constant value of u and v_0 is the constant value of v in the two cases. Thus,

$$\begin{aligned} &u(t) = u_0, \quad v(t) = t, \qquad \tilde{u}(t) = t, \quad \tilde{v}(t) = v_0, \\ \therefore \quad &\dot{u} = 0, \quad \dot{v} = 1, \qquad \dot{\tilde{u}} = 1, \quad \dot{\tilde{v}} = 0. \end{aligned}$$

These parameter curves intersect at the point $\boldsymbol{\sigma}(u_0, v_0)$ of the surface. By Eq. 6.7, their angle of intersection θ is given by

$$\cos\theta = \frac{F}{\sqrt{EG}},$$

where E, F and G are evaluated at (u_0, v_0). In particular, the parameter curves are orthogonal if and only if $F = 0$.

Corresponding to the Definition 6.2.1 of a local isometry, we have the following definition.

Definition 6.3.2

If $\mathcal{S}_1$ and $\mathcal{S}_2$ are surfaces, a *conformal map* $f : \mathcal{S}_1 \to \mathcal{S}_2$ is a local diffeomorphism such that, if $\boldsymbol{\gamma}_1$ and $\tilde{\boldsymbol{\gamma}}_1$ are any two curves on $\mathcal{S}_1$ that intersect, say at a point $\mathbf{p} \in \mathcal{S}_1$, and if $\boldsymbol{\gamma}_2$ and $\tilde{\boldsymbol{\gamma}}_2$ are their images under f, the angle of intersection of $\boldsymbol{\gamma}_1$ and $\tilde{\boldsymbol{\gamma}}_1$ at $\mathbf{p}$ is equal to the angle of intersection of $\boldsymbol{\gamma}_2$ and $\tilde{\boldsymbol{\gamma}}_2$ at $f(\mathbf{p})$.

In short, *f is conformal if and only if it preserves angles.* The reason this definition requires f to be a local diffeomorphism is contained in Exercise 4.4.4 – note that the angle between two intersecting curves is well defined only when both curves are regular.

It is obvious that any composite of conformal maps is conformal, and that the inverse of any conformal diffeomorphism is conformal.

As a special case, if $\boldsymbol{\sigma} : U \to \mathbb{R}^3$ is a surface, then $\boldsymbol{\sigma}$ may be viewed as a map from an open subset of the plane (namely U), parametrized by (u, v) in the usual way, and the image $\mathcal{S}$ of $\boldsymbol{\sigma}$, and we say that $\boldsymbol{\sigma}$ is a *conformal parametrization* or a *conformal surface patch* of $\mathcal{S}$ if this map between surfaces is conformal.

Theorem 6.3.3

A local diffeomorphism $f : \mathcal{S}_1 \to \mathcal{S}_2$ is conformal if and only if there is a function $\lambda : \mathcal{S}_1 \to \mathbb{R}$ such that

$$f^*\langle \mathbf{v}, \mathbf{w} \rangle_{\mathbf{p}} = \lambda(\mathbf{p})\langle \mathbf{v}, \mathbf{w} \rangle_{\mathbf{p}} \quad \text{for all } \mathbf{p} \in \mathcal{S}_1 \text{ and } \mathbf{v}, \mathbf{w} \in T_{\mathbf{p}}\mathcal{S}_1.$$

It is not hard to see that the function λ, if it exists, is necessarily smooth.

Proof

Let $\boldsymbol{\gamma}$ and $\tilde{\boldsymbol{\gamma}}$ be two curves on $\mathcal{S}_1$ that intersect at a point $\mathbf{p} \in \mathcal{S}_1$. The angle θ of intersection of the curves is given by Eq. 6.6. The corresponding angle of intersection of the curves $f \circ \boldsymbol{\gamma}$ and $f \circ \tilde{\boldsymbol{\gamma}}$ on $\mathcal{S}_2$ is obtained from the expression on the right-hand side of Eq. 6.6 by replacing $\dot{\boldsymbol{\gamma}}$ and $\dot{\tilde{\boldsymbol{\gamma}}}$ with $(f \circ \boldsymbol{\gamma})^{\cdot}$ and $(f \circ \tilde{\boldsymbol{\gamma}})^{\cdot}$, respectively. Now,

$$\langle (f \circ \boldsymbol{\gamma})^{\cdot}, (f \circ \tilde{\boldsymbol{\gamma}})^{\cdot} \rangle_{f(\mathbf{p})} = \langle D_{\mathbf{p}} f(\dot{\boldsymbol{\gamma}}), D_{\mathbf{p}} f(\dot{\tilde{\boldsymbol{\gamma}}}) \rangle_{f(\mathbf{p})} = f^* \langle \dot{\boldsymbol{\gamma}}, \dot{\tilde{\boldsymbol{\gamma}}} \rangle_{\mathbf{p}},$$

with similar expressions for $\langle (f \circ \boldsymbol{\gamma})^{\cdot}, f \circ \boldsymbol{\gamma})^{\cdot} \rangle_{f(\mathbf{p})}$ and $\langle (f \circ \tilde{\boldsymbol{\gamma}})^{\cdot}, f \circ \tilde{\boldsymbol{\gamma}})^{\cdot} \rangle_{f(\mathbf{p})}$. Thus, to compute the angle of intersection of the curves $f \circ \boldsymbol{\gamma}$ and $f \circ \tilde{\boldsymbol{\gamma}}$ on $\mathcal{S}_2$, we must replace $\langle\ ,\ \rangle$ in the numerator and denominator of the expression on the right-hand side of Eq. 6.6 by $f^* \langle\ ,\ \rangle$. It is now clear that, if $f^* \langle\ ,\ \rangle = \lambda \langle\ ,\ \rangle$, this replacement leaves the expression in Eq. 6.6 unchanged (since the factor λ cancels out) and so f is conformal.

For the converse, we must show that if

$$\frac{\langle \dot{\boldsymbol{\gamma}}, \dot{\tilde{\boldsymbol{\gamma}}} \rangle}{\langle \dot{\boldsymbol{\gamma}}, \dot{\boldsymbol{\gamma}} \rangle^{1/2} \langle \dot{\tilde{\boldsymbol{\gamma}}}, \dot{\tilde{\boldsymbol{\gamma}}} \rangle^{1/2}} = \frac{f^* \langle \dot{\boldsymbol{\gamma}}, \dot{\tilde{\boldsymbol{\gamma}}} \rangle}{f^* \langle \dot{\boldsymbol{\gamma}}, \dot{\boldsymbol{\gamma}} \rangle^{1/2} f^* \langle \dot{\tilde{\boldsymbol{\gamma}}}, \dot{\tilde{\boldsymbol{\gamma}}} \rangle^{1/2}} \tag{6.8}$$

for all pairs of intersecting curves $\boldsymbol{\gamma}$ and $\tilde{\boldsymbol{\gamma}}$ on $\mathcal{S}_1$, then $f^* \langle\ ,\ \rangle$ is proportional to $\langle\ ,\ \rangle$. Since every tangent vector to $\mathcal{S}_1$ is the tangent vector of a curve on $\mathcal{S}_1$, Eq. 6.8 implies that

$$\frac{\langle \mathbf{v}, \mathbf{w} \rangle}{\langle \mathbf{v}, \mathbf{v} \rangle^{1/2} \langle \mathbf{w}, \mathbf{w} \rangle^{1/2}} = \frac{f^* \langle \mathbf{v}, \mathbf{w} \rangle}{f^* \langle \mathbf{v}, \mathbf{v} \rangle^{1/2} f^* \langle \mathbf{w}, \mathbf{w} \rangle^{1/2}} \tag{6.9}$$

for all tangent vectors $\mathbf{v}, \mathbf{w}$ to $\mathcal{S}_1$.

Choose an orthonormal basis $\{\mathbf{v}_1, \mathbf{v}_2\}$ of the tangent plane to $\mathcal{S}_1$ with respect to its first fundamental form $\langle\ ,\ \rangle$. Let

$$\lambda = f^* \langle \mathbf{v}_1, \mathbf{v}_1 \rangle, \quad \mu = f^* \langle \mathbf{v}_1, \mathbf{v}_2 \rangle, \quad \nu = f^* \langle \mathbf{v}_2, \mathbf{v}_2 \rangle.$$

We apply Eq. 6.9 with $\mathbf{v} = \mathbf{v}_1$ and $\mathbf{w} = \cos\theta\, \mathbf{v}_1 + \sin\theta\, \mathbf{v}_2$, where $\theta \in \mathbb{R}$. This gives

$$\cos\theta = \frac{\lambda \cos\theta + \mu \sin\theta}{\sqrt{\lambda(\lambda \cos^2\theta + 2\mu \sin\theta \cos\theta + \nu \sin^2\theta)}}.$$

Taking $\theta = \pi/2$ gives $\mu = 0$, which implies that

$$\lambda = \lambda \cos^2\theta + \nu \sin^2\theta \quad \text{for all } \theta \in \mathbb{R}.$$

Hence, $\lambda = \nu$. This implies that $f^* \langle \mathbf{v}, \mathbf{w} \rangle = \lambda \langle \mathbf{v}, \mathbf{w} \rangle$ whenever $\mathbf{v}$ and $\mathbf{w}$ are basis vectors. Since both sides are bilinear forms, it follows that $f^* \langle\ ,\ \rangle = \lambda \langle\ ,\ \rangle$. □

Reinterpreting this result in terms of surface patches gives

Corollary 6.3.4

A local diffeomorphism $f : \mathcal{S}_1 \to \mathcal{S}_2$ is conformal if and only if, for any surface patch $\boldsymbol{\sigma}$ of $\mathcal{S}_1$, the first fundamental forms of the patches $\boldsymbol{\sigma}$ of $\mathcal{S}_1$ and $f \circ \boldsymbol{\sigma}$ of $\mathcal{S}_2$ are proportional.

In particular, a surface patch $\boldsymbol{\sigma}(u, v)$ is conformal if and only if its first fundamental form is $\lambda(du^2 + dv^2)$ for some smooth function $\lambda(u, v)$.

Example 6.3.5

We consider the unit sphere S^2. If $\mathbf{q}$ is any point of S^2 other than the north pole $\mathbf{n} = (0, 0, 1)$, the straight line joining $\mathbf{n}$ and $\mathbf{q}$ intersects the xy-plane at some point $\mathbf{p}$, say. The map that takes $\mathbf{q}$ to $\mathbf{p}$ is called *stereographic projection* from S^2 to the plane, and we denote it by Π. We are going to show that Π is conformal.

Let $\mathbf{p} = (u, v, 0)$, $\mathbf{q} = (x, y, z)$. Since $\mathbf{p}, \mathbf{q}, \mathbf{n}$ lie on a straight line, there is a scalar ρ such that

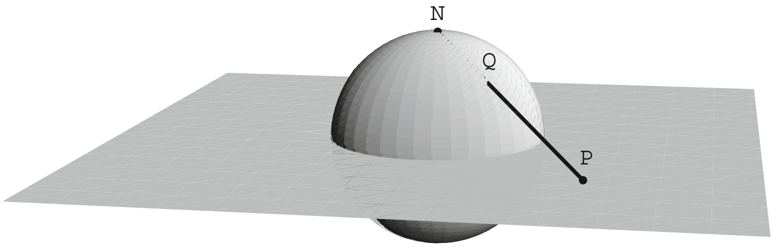
$$\mathbf{q} - \mathbf{n} = \rho(\mathbf{p} - \mathbf{n}),$$

and hence

$$(x, y, z) = (0, 0, 1) + \rho((u, v, 0) - (0, 0, 1)) = (\rho u, \rho v, 1 - \rho). \tag{6.10}$$

Hence, $\rho = 1 - z$, $u = x/(1 - z)$, $v = y/(1 - z)$ and we have

$$\Pi(x, y, z) = \left(\frac{x}{1 - z}, \frac{y}{1 - z}, 0\right).$$



On the other hand, from Eq. 6.10 and $x^2 + y^2 + z^2 = 1$ we get $\rho = 2/(u^2 + v^2 + 1)$ and hence

$$\mathbf{q} = \left(\frac{2u}{u^2 + v^2 + 1}, \frac{2v}{u^2 + v^2 + 1}, \frac{u^2 + v^2 - 1}{u^2 + v^2 + 1}\right).$$

If we denote the right-hand side by $\boldsymbol{\sigma}_1(u,v)$, then $\boldsymbol{\sigma}_1$ is a parametrization of S^2 with the north pole removed. Parametrizing the xy-plane by $\boldsymbol{\sigma}_2(u,v) = (u,v,0)$, we then have

$$\Pi(\boldsymbol{\sigma}_1(u,v)) = \boldsymbol{\sigma}_2(u,v).$$

According to Corollary 6.3.4, to show that Π is conformal we have to show that the first fundamental forms of $\boldsymbol{\sigma}_1$ and $\boldsymbol{\sigma}_2$ are proportional. The first fundamental form of $\boldsymbol{\sigma}_2$ is $du^2 + dv^2$. As to $\boldsymbol{\sigma}_1$, we get

$$\begin{aligned}(\boldsymbol{\sigma}_1)_u &= \left(\frac{2(v^2-u^2+1)}{(u^2+v^2+1)^2}, \frac{-4uv}{(u^2+v^2+1)^2}, \frac{4u}{(u^2+v^2+1)^2}\right),\\ (\boldsymbol{\sigma}_1)_v &= \left(\frac{-4uv}{(u^2+v^2+1)^2}, \frac{2(u^2-v^2+1)}{(u^2+v^2+1)^2}, \frac{4v}{(u^2+v^2+1)^2}\right).\end{aligned}$$

This gives

$$E_1 = (\boldsymbol{\sigma}_1)_u \cdot (\boldsymbol{\sigma}_1)_u = \frac{4(v^2-u^2+1)^2 + 16u^2v^2 + 16u^2}{(u^2+v^2+1)^4} = \frac{4}{(u^2+v^2+1)^2}.$$

Similarly, $F_1 = 0$, $G_1 = 4/(u^2+v^2+1)^2$. Thus, the first fundamental form of $\boldsymbol{\sigma}_2$ is λ times that of $\boldsymbol{\sigma}_1$, where $\lambda = \frac{1}{4}(u^2+v^2+1)^2$.

It is often useful to think of Π as a map to the complex numbers $\mathbb{C}$ rather than to the xy-plane, by identifying $u+iv \in \mathbb{C}$ with $(u,v,0)$. Moreover, we can parametrize S^2 itself in a partly complex way by identifying $(x,y,z) \in S^2$ with $(x+iy, z)$. Then, S^2 becomes the set of pairs (w,z) where $w \in \mathbb{C}$, $z \in \mathbb{R}$ and $|w|^2 + z^2 = 1$. Stereographic projection then takes the simple form

$$\Pi(w,z) = \frac{w}{1-z},$$

and the surface patch $\boldsymbol{\sigma}_1$ is given by

$$\boldsymbol{\sigma}_1(w) = \left(\frac{2w}{|w|^2+1}, \frac{|w|^2-1}{|w|^2+1}\right).$$

The inconvenience of having to exclude the north pole from the domain of definition of Π can be overcome by introducing a 'point at infinity' ∞ and defining the 'extended complex plane' $\mathbb{C}_\infty = \mathbb{C} \cup \{\infty\}$. If we agree that Π maps the north pole to ∞, it defines a *bijection* $\Pi : S^2 \to \mathbb{C}_\infty$. Further discussion of this map is left to the exercises.

Returning now to the general case, it is natural to ask when there is a conformal map between two surfaces. The surprising answer is that this is *always the case locally*: if $\mathbf{p}_1$ and $\mathbf{p}_2$ are points of two surfaces $\mathcal{S}_1$ and $\mathcal{S}_2$, respectively, there are open subsets $\mathcal{O}_1$ of $\mathcal{S}_1$ containing $\mathbf{p}_1$ and $\mathcal{O}_2$ of $\mathcal{S}_2$ containing $\mathbf{p}_2$ and a conformal diffeomorphism $\mathcal{O}_1 \to \mathcal{O}_2$. This follows from the following theorem:

Theorem 6.3.6

Every surface has an atlas consisting of conformal surface patches.

Indeed, if $\boldsymbol{\sigma}_1$ and $\boldsymbol{\sigma}_2$ are conformal parametrizations of $\mathcal{S}_1$ and $\mathcal{S}_2$, the map $\boldsymbol{\sigma}_1(u,v) \mapsto \boldsymbol{\sigma}_2(u,v)$ will be conformal as it is the composite of the conformal diffeomorphism $\boldsymbol{\sigma}_2$ and the inverse of the conformal diffeomorphism $\boldsymbol{\sigma}_1$.

We shall prove a special case of Theorem 6.3.6 later (see Theorem 12.4.1), but the general case is beyond the scope of this book.

EXERCISES

6.3.1 Show that every local isometry is conformal. Give an example of a conformal map that is not a local isometry.

6.3.2 Show that *Enneper's surface*

$$\boldsymbol{\sigma}(u,v) = \left(u - \frac{u^3}{3} + uv^2, v - \frac{v^3}{3} + vu^2, u^2 - v^2\right)$$

is conformally parametrized.

6.3.3 Recall from Example 6.1.3 that the first fundamental form of the latitude–longitude parametrization $\boldsymbol{\sigma}(\theta,\varphi)$ of S^2 is

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

Find a smooth function ψ such that the reparametrization $\tilde{\boldsymbol{\sigma}}(u,v) = \boldsymbol{\sigma}(\psi(u),v)$ is conformal. Verify that $\tilde{\boldsymbol{\sigma}}$ is, in fact, the Mercator parametrization in Exercise 5.3.2.

6.3.4 Let $\Phi : U \to V$ be a diffeomorphism between open subsets of $\mathbb{R}^2$. Write

$$\Phi(u,v) = (f(u,v), g(u,v)),$$

where f and g are smooth functions on the uv-plane. Show that Φ is conformal if and only if

$$\text{either } (f_u = g_v \text{ and } f_v = -g_u) \text{ or } (f_u = -g_v \text{ and } f_v = g_u). \tag{6.11}$$

Show that, if $J(\Phi)$ is the Jacobian matrix of Φ, then $\det(J(\Phi)) > 0$ in the first case and $\det(J(\Phi)) < 0$ in the second case.

6.3.5 (This exercise requires a basic knowledge of complex analysis.) Recall that the transition map between two surface patches in an atlas for a surface $\mathcal{S}$ is a smooth map between open subsets of $\mathbb{R}^2$. Since

$\mathbb{R}^2$ is the 'same' as the complex numbers $\mathbb{C}$ (via $(u, v) \leftrightarrow u + iv$), we can ask whether such a transition map is *holomorphic*. One says that $\mathcal{S}$ is a *Riemann surface* if $\mathcal{S}$ has an atlas for which all the transition maps are holomorphic. Deduce from Theorem 6.3.6 and the preceding exercise that every orientable surface has an atlas making it a Riemann surface. (You will need to recall from complex analysis that a smooth function Φ as in the preceding exercise is holomorphic if and only if the first pair of equations in (6.11) hold – these are the *Cauchy–Riemann equations*. If the second pair of equations in (6.11) hold, Φ is said to be *anti-holomorphic*.)

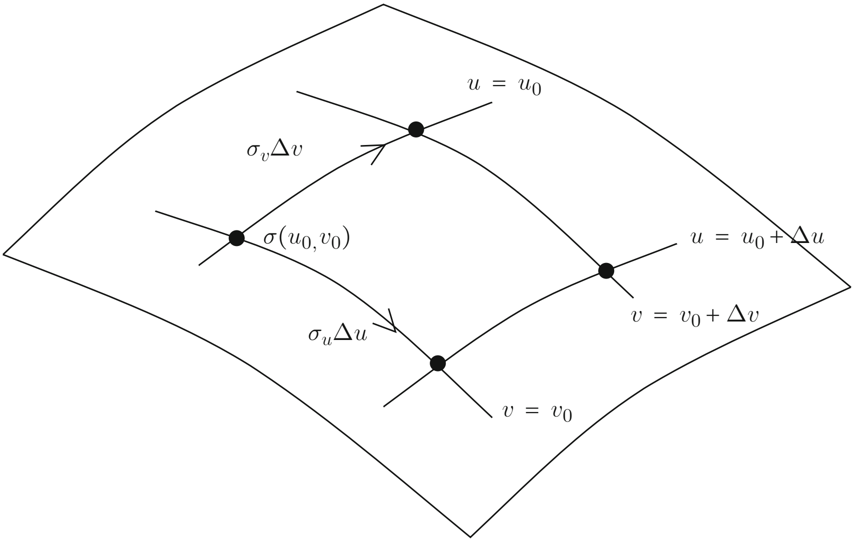
6.3.6 Define a map $\tilde{\Pi}$ similar to Π by projecting from the *south* pole of S^2 onto the xy-plane. Show that this defines a second conformal surface patch $\tilde{\boldsymbol{\sigma}}_1$, which covers the whole of S^2 except the south pole. What is the transition map between these two patches? Why do the two patches $\boldsymbol{\sigma}_1$ and $\tilde{\boldsymbol{\sigma}}_1$ *not* give S^2 the structure of a Riemann surface? How can $\tilde{\boldsymbol{\sigma}}_1$ be modified to produce such a structure?

6.3.7 Show that the stereographic projection map Π takes circles on S^2 to Circles in $\mathbb{C}_\infty$, and that every Circle arises in this way. (A circle on S^2 is the intersection of S^2 with a plane; a Circle in $\mathbb{C}_\infty$ is a line or a circle in $\mathbb{C}$ – see Appendix 2).

6.3.8 Show that, if M is a Möbius transformation or a conjugate-Möbius transformation (see Appendix 2), the bijection $\Pi^{-1} \circ M \circ \Pi : S^2 \to S^2$ is a conformal diffeomorphism of S^2. It can be shown that every conformal diffeomorphism of S^2 is of this type.

6.4 Equiareal maps and a theorem of Archimedes

Suppose that $\boldsymbol{\sigma} : U \to \mathbb{R}^3$ is a surface patch on a surface $\mathcal{S}$. The image of $\boldsymbol{\sigma}$ is covered by the two families of parameter curves obtained by setting $u =$ constant and $v =$ constant, respectively. Fix $(u_0, v_0) \in U$; since the change in $\boldsymbol{\sigma}(u, v)$ corresponding to a small change Δu in u is approximately $\boldsymbol{\sigma}_u \Delta u$ and that corresponding to a small change Δv in v is approximately $\boldsymbol{\sigma}_v \Delta v$, the part of the surface contained by the parameter curves on the surface corresponding to $u = u_0$, $u = u_0 + \Delta u$, $v = v_0$ and $v = v_0 + \Delta v$ is approximately a parallelogram in the plane with sides given by the vectors $\boldsymbol{\sigma}_u \Delta u$ and $\boldsymbol{\sigma}_v \Delta v$ (the derivatives being evaluated at (u_0, v_0)):



Recalling that the area of a parallelogram in the plane with sides **a** and **b** is $\| \mathbf{a} \times \mathbf{b} \|$, we see that the area of the parallelogram on the surface is approximately

$$\| \boldsymbol{\sigma}_u \Delta u \times \boldsymbol{\sigma}_v \Delta v \| = \| \boldsymbol{\sigma}_u \times \boldsymbol{\sigma}_v \| \Delta u \Delta v.$$

This suggests the following definition.

Definition 6.4.1

The *area* $\mathcal{A}_{\boldsymbol{\sigma}}(R)$ of the part $\boldsymbol{\sigma}(R)$ of a surface patch $\boldsymbol{\sigma} : U \to \mathbb{R}^3$ corresponding to a region $R \subseteq U$ is

$$\mathcal{A}_{\boldsymbol{\sigma}}(R) = \int_R \| \boldsymbol{\sigma}_u \times \boldsymbol{\sigma}_v \| \, dudv.$$

Of course, this integral may be infinite – think of the area of a whole plane, for example. However, the integral will be finite if, say, R is contained in a rectangle that is entirely contained, along with its boundary, in U.

The quantity $\| \boldsymbol{\sigma}_u \times \boldsymbol{\sigma}_v \|$ that appears in the definition of area is easily computed in terms of the first fundamental form $Edu^2 + 2Fdudv + Gdv^2$ of $\boldsymbol{\sigma}$:

Proposition 6.4.2

$\| \boldsymbol{\sigma}_u \times \boldsymbol{\sigma}_v \| = (EG - F^2)^{1/2}.$

Proof

We use a result from vector algebra: if $\mathbf{a}, \mathbf{b}, \mathbf{c}$ and $\mathbf{d}$ are vectors in $\mathbb{R}^3$,

$$(\mathbf{a} \times \mathbf{b}) \cdot (\mathbf{c} \times \mathbf{d}) = (\mathbf{a} \cdot \mathbf{c})(\mathbf{b} \cdot \mathbf{d}) - (\mathbf{a} \cdot \mathbf{d})(\mathbf{b} \cdot \mathbf{c}).$$

Applying this to $\| \boldsymbol{\sigma}_u \times \boldsymbol{\sigma}_v \|^2 = (\boldsymbol{\sigma}_u \times \boldsymbol{\sigma}_v) \cdot (\boldsymbol{\sigma}_u \times \boldsymbol{\sigma}_v)$, we get

$$\| \boldsymbol{\sigma}_u \times \boldsymbol{\sigma}_v \|^2 = (\boldsymbol{\sigma}_u \cdot \boldsymbol{\sigma}_u)(\boldsymbol{\sigma}_v \cdot \boldsymbol{\sigma}_v) - (\boldsymbol{\sigma}_u \cdot \boldsymbol{\sigma}_v)^2 = EG - F^2. \qquad \square$$

Note that, for a regular surface, $EG - F^2 > 0$ everywhere, since for a regular surface $\boldsymbol{\sigma}_u \times \boldsymbol{\sigma}_v$ is never zero.

Thus, our definition of area is

$$\mathcal{A}_{\boldsymbol{\sigma}}(R) = \int_R (EG - F^2)^{1/2} du dv. \tag{6.12}$$

We sometimes denote $(EG - F^2)^{1/2} du dv$ by $d\mathcal{A}_{\boldsymbol{\sigma}}$. But we have still to check that this definition is sensible, i.e., it is unchanged if $\boldsymbol{\sigma}$ is reparametrized. This is certainly not obvious, since E, F and G change under reparametrization (see Exercise 6.1.4).

Proposition 6.4.3

The area of a surface patch is unchanged by reparametrization.

Proof

Let $\boldsymbol{\sigma} : U \to \mathbb{R}^3$ be a surface patch and let $\tilde{\boldsymbol{\sigma}} : \tilde{U} \to \mathbb{R}^3$ be a reparametrization of $\boldsymbol{\sigma}$, with reparametrization map $\Phi : \tilde{U} \to U$. Thus, if $\Phi(\tilde{u}, \tilde{v}) = (u, v)$, we have

$$\tilde{\boldsymbol{\sigma}}(\tilde{u}, \tilde{v}) = \boldsymbol{\sigma}(u, v).$$

Let $\tilde{R} \subseteq \tilde{U}$ be a region, and let $R = \Phi(\tilde{R}) \subseteq U$. We have to prove that

$$\int_R \| \boldsymbol{\sigma}_u \times \boldsymbol{\sigma}_v \| du dv = \int_{\tilde{R}} \| \tilde{\boldsymbol{\sigma}}_{\tilde{u}} \times \tilde{\boldsymbol{\sigma}}_{\tilde{v}} \| d\tilde{u} d\tilde{v}.$$

We showed in the proof of Proposition 4.2.7 that

$$\tilde{\boldsymbol{\sigma}}_{\tilde{u}} \times \tilde{\boldsymbol{\sigma}}_{\tilde{v}} = \det(J(\Phi))\, \boldsymbol{\sigma}_u \times \boldsymbol{\sigma}_v,$$

where $J(\Phi)$ is the Jacobian matrix of Φ. Hence,

$$\int_{\tilde{R}} \parallel \tilde{\boldsymbol{\sigma}}_{\tilde{u}} \times \tilde{\boldsymbol{\sigma}}_{\tilde{v}} \parallel \, d\tilde{u}d\tilde{v} = \int_{\tilde{R}} |\det(J(\Phi))| \parallel \boldsymbol{\sigma}_u \times \boldsymbol{\sigma}_v \parallel \, d\tilde{u}d\tilde{v}.$$

By the change of variables formula for double integrals, the right-hand side of this equation is exactly

$$\int_R \parallel \boldsymbol{\sigma}_u \times \boldsymbol{\sigma}_v \parallel \, dudv.$$

□

Now that we have a good definition of area, we can ask which maps between surfaces are area-preserving.

Definition 6.4.4

Let $\mathcal{S}_1$ and $\mathcal{S}_2$ be two surfaces. A local diffeomorphism $f : \mathcal{S}_1 \to \mathcal{S}_2$ is said to be *equiareal* if it takes any region in $\mathcal{S}_1$ to a region of *the same area* in $\mathcal{S}_2$ (we assume that each of the regions is sufficiently small, so that it is contained in the image of some surface patch).

We have the following analogue of Theorem 6.2.2.

Theorem 6.4.5

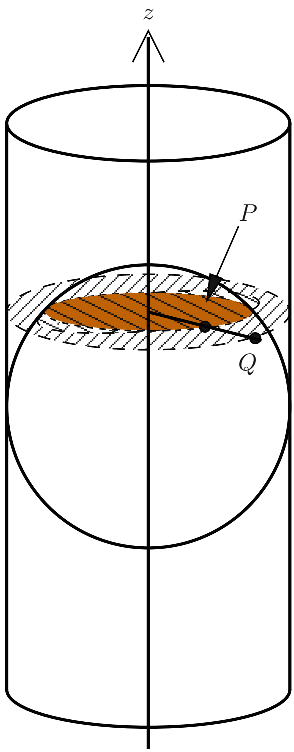
A local diffeomorphism $f : \mathcal{S}_1 \to \mathcal{S}_2$ is equiareal if and only if, for any surface patch $\boldsymbol{\sigma}(u, v)$ on $\mathcal{S}_1$, the first fundamental forms

$$E_1 du^2 + 2F_1 dudv + G_1 dv^2 \quad \text{and} \quad E_2 du^2 + 2F_2 dudv + G_2 dv^2$$

of the patches $\boldsymbol{\sigma}$ on $\mathcal{S}_1$ and $f \circ \boldsymbol{\sigma}$ on $\mathcal{S}_2$ satisfy

$$E_1 G_1 - F_1^2 = E_2 G_2 - F_2^2. \tag{6.13}$$

The proof is very similar to that of Theorem 6.2.2 and we leave it as Exercise 6.4.6. As with isometries and conformal maps, it is obvious that any composite of equaireal diffeomorphism is equiareal, and that the inverse of any equiareal difeomorphism is equaireal.



One of the most famous examples of an equiareal map was found by Archimedes. Legend has it that the discovery was inscribed onto his tombstone by the Roman general Marcellus who led the siege of Syracuse in which Archimedes perished. Naturally, since calculus was not available to him, Archimedes' proof of his theorem was quite different from ours.

Consider the unit sphere $x^2 + y^2 + z^2 = 1$ and the unit cylinder $x^2 + y^2 = 1$. The sphere is contained inside the cylinder, and the two surfaces touch along the circle $x^2 + y^2 = 1$ in the xy-plane. For each point $\mathbf{p} \in S^2$ other than the poles $(0, 0, \pm 1)$, there is a unique straight line parallel to the xy-plane and passing through the point $\mathbf{p}$ and the z-axis. This line intersects the cylinder in two points, one of which, say $\mathbf{q}$, is closest to $\mathbf{p}$. Let f be the map from S^2 (with the two poles removed) to the cylinder that takes $\mathbf{p}$ to $\mathbf{q}$.

To find a formula for f, let (x, y, z) be the Cartesian coordinates of $\mathbf{p}$, and (X, Y, Z) those of $\mathbf{q}$. Since the line through $\mathbf{p}$ and $\mathbf{q}$ is parallel to the xy-plane, we have $Z = z$ and $(X, Y) = \lambda(x, y)$ for some scalar λ. Since (X, Y, Z) is on the cylinder,

$$1 = X^2 + Y^2 = \lambda^2(x^2 + y^2),$$

$$\therefore \quad \lambda = \pm(x^2 + y^2)^{-1/2}.$$

Taking the + sign gives the point $\mathbf{q}$, so we get

$$f(x, y, z) = \left(\frac{x}{(x^2 + y^2)^{1/2}}, \frac{y}{(x^2 + y^2)^{1/2}}, z \right).$$

We shall show in the proof of the next theorem that f is a diffeomorphism.

Theorem 6.4.6 (Archimedes' Theorem)

The map f is an equiareal diffeomorphism.

Proof

We take the atlas for the surface $\mathcal{S}_1$ consisting of the sphere minus the north and south poles with two patches, both given by the formula

$$\boldsymbol{\sigma}_1(\theta, \varphi) = (\cos\theta\cos\varphi, \cos\theta\sin\varphi, \sin\theta),$$

and defined on the open sets

$$\{-\pi/2 < \theta < \pi/2,\ 0 < \varphi < 2\pi\} \text{ and } \{-\pi/2 < \theta < \pi/2,\ -\pi < \varphi < \pi\}.$$

The image of $\boldsymbol{\sigma}_1(\theta, \varphi)$ under the map f is the point

$$\boldsymbol{\sigma}_2(\theta, \varphi) = (\cos\varphi, \sin\varphi, \sin\theta) \tag{6.14}$$

of the cylinder. It is easy to check that this gives an atlas for the surface $\mathcal{S}_2$, consisting of the part of the cylinder between the planes $z = 1$ and $z = -1$, with two patches, both given by Eq. 6.14 and defined on the same two open sets as $\boldsymbol{\sigma}_1$. We have to show that Eq. 6.13 holds.

We computed the coefficients E_1, F_1 and G_1 of the first fundamental form of $\boldsymbol{\sigma}_1$ in Example 6.1.3:

$$E_1 = 1, \quad F_1 = 0, \quad G_1 = \cos^2\theta.$$

For $\boldsymbol{\sigma}_2$, we get $(\boldsymbol{\sigma}_2)_\theta = (0, 0, \cos\theta)$, $(\boldsymbol{\sigma}_2)_\varphi = (-\sin\varphi, \cos\varphi, 0)$, and so

$$E_2 = \cos^2\theta, \quad F_2 = 0, \quad G_2 = 1.$$

It is now clear that Eq. 6.13 holds.

Note that, since f corresponds simply to the identity map $(\theta, \varphi) \mapsto (\theta, \varphi)$ in terms of the parametrizations $\boldsymbol{\sigma}_1$ and $\boldsymbol{\sigma}_2$ of the unit sphere and cylinder, respectively, it follows that f is a diffeomorphism. □

The following classical result provides a beautiful application of Archimedes' theorem. A *spherical triangle* is a triangle on a sphere whose sides are arcs of great circles.

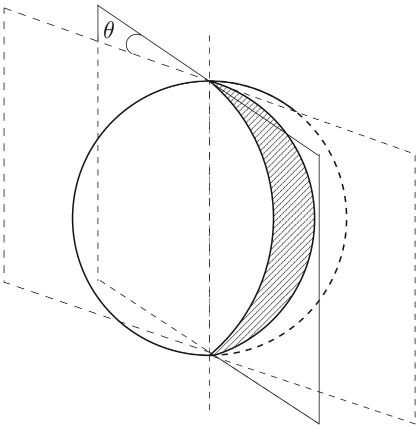
Theorem 6.4.7

The area of a spherical triangle on the unit sphere S^2 with internal angles α, β and γ is

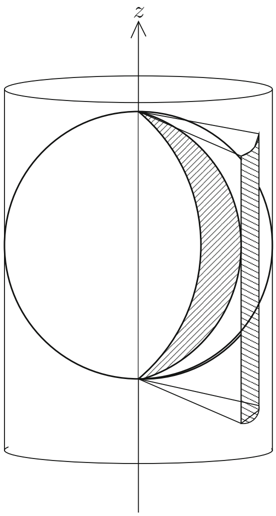
$$\alpha + \beta + \gamma - \pi.$$

Proof

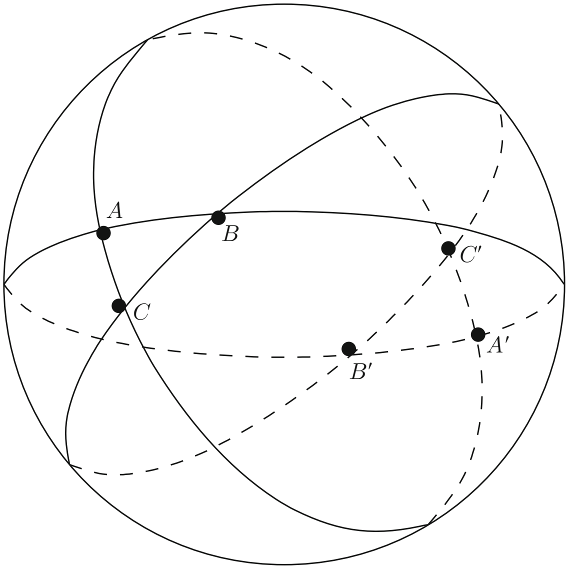
We begin by using Archimedes' Theorem 6.4.6 to compute the area of a 'lune', i.e., the area enclosed between two great circles:



We can assume that the great circles intersect at the poles, since this can be achieved by applying a rotation of S^2, and this does not change areas. If θ is the angle between them, the image of the lune under the map f is a curved rectangle on the cylinder of width θ and height 2 (see next page). If we now apply the isometry which unwraps the cylinder on the plane, this curved rectangle on the cylinder will map to a genuine rectangle on the plane, with width θ and height 2. By Archimedes' theorem, the lune has the same area as the curved rectangle on the cylinder, and since every isometry is an equiareal map (see Exercise 6.4.6), this has the same area as the genuine rectangle in the plane, namely 2θ. Note that this correctly gives the area of the whole sphere to be 4π.



Turning now to the proof of the theorem, let A, B and C be the vertices of the triangle (so that α is the angle at A, etc.). The three great circles, of which the sides of the triangle are arcs, divide S^2 into eight triangles, as shown in the following diagram (in which A' is the antipodal point of A, etc.).



Note that the two triangles with vertices A, B, C and A', B, C together form a lune with angle α, etc. Hence, denoting the triangle with vertices A, B, C by ABC and its area by $\mathcal{A}(ABC)$, etc., we have, by the preceding calculation,

$$\mathcal{A}(ABC) + \mathcal{A}(A'BC) = 2\alpha,$$
$$\mathcal{A}(ABC) + \mathcal{A}(AB'C) = 2\beta,$$
$$\mathcal{A}(ABC) + \mathcal{A}(ABC') = 2\gamma.$$

Adding these equations, we get

$$2\mathcal{A}(ABC) + \{\mathcal{A}(ABC) + \mathcal{A}(A'BC) + \mathcal{A}(AB'C) + \mathcal{A}(ABC')\} = 2\alpha + 2\beta + 2\gamma. \tag{6.15}$$

Now, the triangles ABC, $AB'C$, $AB'C'$ and ABC' together make a hemisphere (namely, the hemisphere containing the vertex A with boundary the great circle passing through B and C), so

$$\mathcal{A}(ABC) + \mathcal{A}(AB'C) + \mathcal{A}(AB'C') + \mathcal{A}(ABC') = 2\pi. \tag{6.16}$$

Finally, since the map that takes each point of S^2 to its antipodal point is an isometry, and hence equiareal, we have

$$\mathcal{A}(A'BC) = \mathcal{A}(AB'C').$$

Inserting this into Eq. 6.16, we see that the term in { } on the left-hand side of Eq. 6.15 is equal to 2π. Rearranging now gives the result. □

In Chapter 13, we shall obtain a far-reaching generalization of this result in which S^2 is replaced by an arbitrary surface, and great circles by arbitrary curves on the surface.

EXERCISES

6.4.1 Determine the area of the part of the paraboloid $z = x^2 + y^2$ with $z \leq 1$ and compare with the area of the hemisphere $x^2 + y^2 + z^2 = 1$, $z \leq 0$.

6.4.2 A sailor circumnavigates Australia by a route consisting of a triangle whose sides are arcs of great circles. Prove that at least one interior angle of the triangle is $\geq \frac{\pi}{3} + \frac{10}{169}$ radians. (Take the Earth to be a sphere of radius 6,500km and assume that the area of Australia is 7.5 million square kilometres.)

6.4.3 A *spherical polygon* on S^2 is the region formed by the intersection of n hemispheres of S^2, where n is an integer ≥ 3. Show that, if $\alpha_1, \ldots, \alpha_n$ are the interior angles of such a polygon, its area is equal to

$$\sum_{i=1}^{n} \alpha_i - (n-2)\pi.$$

6.4.4 Suppose that S^2 is covered by spherical polygons such that the intersection of any two polygons is either empty or a common edge or vertex of each polygon. Suppose that there are F polygons, E edges and V vertices (a common edge or vertex of more than one polygon being counted only once). Show that the sum of the angles of all the polygons is $2\pi V$. By using the preceding exercise, deduce that $V - E + F = 2$. (This result is due to Euler; it will be generalized in Chapter 13.)

6.4.5 Show that:

(i) Every local isometry is an equiareal map.

(ii) A map that is both conformal and equiareal is a local isometry.

Give an example of an equiareal map that is not a local isometry.

6.4.6 Prove Theorem 6.4.5.

6.4.7 Let $\boldsymbol{\sigma}(u, v)$ be a surface patch with standard unit normal $\mathbf{N}$. Show that

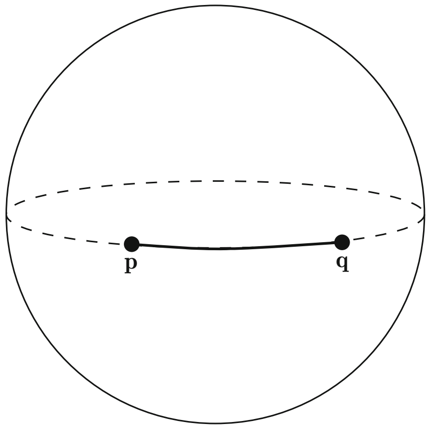
$$\mathbf{N} \times \boldsymbol{\sigma}_u = \frac{E\boldsymbol{\sigma}_v - F\boldsymbol{\sigma}_u}{\sqrt{EG - F^2}}, \quad \mathbf{N} \times \boldsymbol{\sigma}_v = \frac{F\boldsymbol{\sigma}_v - G\boldsymbol{\sigma}_v}{\sqrt{EG - F^2}}.$$

6.5 Spherical geometry

We conclude this chapter with a brief discussion of the simplest example of a geometry different from Euclid's, namely spherical geometry. The study of spherical geometry, like that of plane geometry, began in antiquity. Its importance was astronomical: to locate an object in the sky such as a star, one imagines a fixed large sphere centred on the observer; the straight line connecting the observer to the star intersects the sphere in a point whose position gives the direction in which the observer must look in order to see the star. Thus, the three-dimensional universe is projected onto the surface of a sphere. Of course, spherical geometry is also important because we live on the surface of a sphere, to a reasonably good approximation.

If we are to develop spherical geometry by analogy with Euclidean plane geometry, the first thing to do is to decide what should be the analogue of straight lines. Now straight lines are the shortest curves joining any two of their points (Exercise 1.2.4), so it is natural to ask what the corresponding shortest curves are on the sphere. We are going to show that these are arcs of *great circles*.

For simplicity, we work with the unit sphere S^2. If $\mathbf{p}$ and $\mathbf{q}$ are two distinct points of S^2, there is always at least one great circle passing through $\mathbf{p}$ and $\mathbf{q}$. To see this, note first that if $\mathbf{p}$ and $\mathbf{q}$ are *antipodal* points, i.e., if $\mathbf{p} = -\mathbf{q}$, the intersection of S^2 with any plane containing this diameter is a great circle through $\mathbf{p}$ and $\mathbf{q}$. If $\mathbf{p}$ and $\mathbf{q}$ are not antipodal points, the plane passing through the origin perpendicular to the (non-zero) vector $\mathbf{p} \times \mathbf{q}$ intersects S^2 in a great circle passing through $\mathbf{p}$ and $\mathbf{q}$. The argument shows, in fact, that if $\mathbf{p}$ and $\mathbf{q}$ are not antipodal there is a unique great circle passing through them both; in this case $\mathbf{p}$ and $\mathbf{q}$ divide this great circle into two circular arcs, one shorter than the other. If $\mathbf{p}$ and $\mathbf{q}$ are antipodal, there are infinitely many great circles passing through both points, each of which is divided by $\mathbf{p}$ and $\mathbf{q}$ into two semicircles (see below).



Proposition 6.5.1

Let $\mathbf{p}$ and $\mathbf{q}$ be distinct points of S^2. If $\mathbf{p} \neq -\mathbf{q}$, the short great circle arc joining $\mathbf{p}$ and $\mathbf{q}$ is the unique curve of shortest length joining $\mathbf{p}$ and $\mathbf{q}$. If $\mathbf{p} = -\mathbf{q}$, any great semicircle joining $\mathbf{p}$ and $\mathbf{q}$ is a shortest curve joining these two points.

Proof

By using a rotation of S^2 (which is an isometry of S^2 – see Exercise 6.1.2) we can assume that $\mathbf{p}$ is the north pole $(0, 0, 1)$, and by a further rotation about the z-axis we can assume in addition that $\mathbf{q}$ is a point on the great semicircle $\mathcal{C}$ passing through the north and south poles and the point $(1, 0, 0)$, say $(\cos\alpha, 0, \sin\alpha)$, where $-\frac{\pi}{2} \leq \alpha \leq \frac{\pi}{2}$. Then the distance from $\mathbf{p}$ to $\mathbf{q}$ measured along the short great circle arc joining them is $\pi/2 - \alpha$.

The first fundamental form of the latitude-longitude parametrization $\boldsymbol{\sigma}(\theta, \varphi)$ is $d\theta^2 + \cos^2\theta\, d\varphi^2$ (Example 6.1.3) so the length of a curve $\boldsymbol{\gamma}(t)$ passing through $\mathbf{p}$ when $t = t_0$ and through $\mathbf{q}$ when $t = t_1$, say, is

$$\int_{t_0}^{t_1} \left(\dot\theta^2 + \cos^2\theta\, \dot\varphi^2\right)^{1/2} dt.$$

The integrand is not less than $|\dot\theta|$, so the length of the part of $\boldsymbol{\gamma}$ between $\mathbf{p}$ and $\mathbf{q}$ is not less than

$$\int_{t_0}^{t_1} |\dot\theta|\, dt = \int_{\alpha}^{\pi/2} d\theta = \pi/2 - \alpha,$$

which is the length of the short great circle arc passing through $\mathbf{p}$ and $\mathbf{q}$.

Conversely, if $\boldsymbol{\gamma}$ has exactly this length, we must have

$$\left(\dot\theta^2 + \cos^2\theta\, \dot\varphi^2\right)^{1/2} = |\dot\theta|,$$

and hence

$$\cos\theta\, \dot\varphi = 0$$

for all t between t_0 and t_1. Since $\cos\theta = 0$ only at the north and south poles $(0, 0, \pm 1)$, we must therefore have $\dot\varphi = 0$ at all other points of $\boldsymbol{\gamma}$; this means that φ is a constant, which must be zero since $\boldsymbol{\gamma}$ passes through $\mathbf{p}$, and so $\boldsymbol{\gamma}$ is part of $\mathcal{C}$. □

Thus, great circles are the spherical analogues of straight lines in Euclidean geometry. One immediate difference between spherical and plane geometry is that *there are no parallel lines in spherical geometry*, for any two great circles intersect (the two planes containing the two great circles intersect in a diameter of S^2, the endpoints of which are the points of intersection of the two great circles).

The *spherical distance* $d_{S^2}(\mathbf{p}, \mathbf{q})$ between two points $\mathbf{p}, \mathbf{q} \in S^2$ is the length of the short great circle arc joining $\mathbf{p}$ and $\mathbf{q}$. This is simply the angle between the vectors $\mathbf{p}$ and $\mathbf{q}$ in the range $0 \leq d_{S^2}(\mathbf{p}, \mathbf{q}) \leq \pi$: in symbols,

$$\cos d_{S^2}(\mathbf{p}, \mathbf{q}) = \mathbf{p} \cdot \mathbf{q}.$$

There is a beautiful formula for the spherical distance in terms of the stereographic projection map Π (see Example 6.3.5). Recall that Π defines a bijection from S^2 to the extended complex plane $\mathbb{C}_\infty$; we write $d_{S^2}(\Pi^{-1}(w), \Pi^{-1}(z))$ simply as $d_{S^2}(w, z)$.

Proposition 6.5.2

If $w, z \in \mathbb{C}$, the spherical distance $d_{S^2}(z, w)$ between the points of S^2 corresponding to w and z under stereographic projection is given by

$$\tan \frac{1}{2} d_{S^2}(w, z) = \frac{|w - z|}{|1 + \bar{w}z|}.$$

Proof

From Example 6.3.5, the point of S^2 corresponding to $w \in \mathbb{C}$ is

$$\Pi^{-1}(w) = \left(\frac{w + \bar{w}}{|w|^2 + 1}, \frac{w - \bar{w}}{i(|w|^2 + 1)}, \frac{|w|^2 - 1}{|w|^2 + 1} \right).$$

Hence,

$$\begin{aligned} \cos d_{S^2}(w, z) &= \Pi^{-1}(w) \cdot \Pi^{-1}(z) \\ &= \frac{(w + \bar{w})(z + \bar{z}) - (w - \bar{w})(z - \bar{z}) + (|w|^2 - 1)(|z|^2 - 1)}{(|w|^2 + 1)(|z|^2 + 1)} \\ &= \frac{2(\bar{w}z + w\bar{z}) + (1 - |w|^2)(1 - |z|^2)}{(|w|^2 + 1)(|z|^2 + 1)}. \end{aligned} \tag{6.17}$$

On the other hand, let t denote the right-hand side of the formula in the statement of the proposition. Then,

$$\begin{aligned} \frac{1 - t^2}{1 + t^2} &= \frac{|1 + \bar{w}z|^2 - |w - z|^2}{|1 + \bar{w}z|^2 + |w - z|^2} \\ &= \frac{(1 + \bar{w}z)(1 + w\bar{z}) - (w - z)(\bar{w} - \bar{z})}{(1 + \bar{w}z)(1 + w\bar{z}) + (w - z)(\bar{w} - \bar{z})} \\ &= \frac{2(\bar{w}z + w\bar{z}) + (1 - |w|^2)(1 - |z|^2)}{(|w|^2 + 1)(|z|^2 + 1)}. \end{aligned} \tag{6.18}$$

The proposition follows on comparing Eqs. 6.17 and 6.18 and recalling the identity

$$\cos \theta = \frac{1 - \tan^2 \frac{1}{2}\theta}{1 + \tan^2 \frac{1}{2}\theta}.$$ □

Much of Euclidean geometry deals with the properties of triangles. We shall always consider only spherical triangles with sides of length less than π.

Proposition 6.5.3

Suppose that a spherical triangle has sides of length A, B and C, and let α, β and γ be its internal angles (so that α is the angle opposite the side of length A, etc., and $0 \leq \alpha, \beta, \gamma < \pi$). Then,

$$\text{(i)} \qquad \cos\gamma = \frac{\cos C - \cos A \cos B}{\sin A \sin B},$$

$$\text{(ii)} \qquad \frac{\sin\alpha}{\sin A} = \frac{\sin\beta}{\sin B} = \frac{\sin\gamma}{\sin C}.$$

Two formulas similar to that in (i) can, of course, be obtained by making the cyclic permutations $A \to B \to C \to A$, $\alpha \to \beta \to \gamma \to \alpha$.

Part (i) is called the 'cosine rule' for spherical triangles because it becomes the usual cosine rule when A, B, C are small, in which case the spherical triangle is 'almost' a plane triangle: using the approximations $\cos A = 1 - \frac{1}{2}A^2$ and $\sin A = A$, etc. we get

$$C^2 = A^2 + B^2 - 2AB\cos\gamma.$$

Similarly (ii) reduces to the familiar sine rule for plane triangles when A, B, C are small.

Proof 6.5.3 Let $\mathbf{a}$, $\mathbf{b}$ and $\mathbf{c}$ be the vertices of the triangle, so that α is the angle at $\mathbf{a}$, etc. Since A is the angle (measured in radians) between the unit vectors $\mathbf{b}$ and $\mathbf{c}$, etc., we have

$$\cos A = \mathbf{b}\cdot\mathbf{c}, \quad \cos B = \mathbf{c}\cdot\mathbf{a}, \quad \cos C = \mathbf{a}\cdot\mathbf{b}. \tag{6.19}$$

Next, the side of the triangle of length C is an arc of the great circle that is the intersection of S^2 with the plane Π_C through the origin and perpendicular to the vector $\mathbf{a}\times\mathbf{b}$ (and similarly for the other sides). Let $\Pi_{\mathbf{c}}$ be the plane passing through the vertex $\mathbf{c}$ parallel to the tangent plane of S^2 there. Then $\Pi_{\mathbf{c}}$ intersects the planes Π_A and Π_B in two straight lines that are tangent to the sides of the triangle passing through $\mathbf{c}$. It follows that γ is the angle between these two lines, which in turn is equal to the angle between Π_A and Π_B, i.e., the angle between $\mathbf{b}\times\mathbf{c}$ and $\mathbf{a}\times\mathbf{c}$:

$$\cos\gamma = \frac{(\mathbf{b}\times\mathbf{c})\cdot(\mathbf{a}\times\mathbf{c})}{\parallel\mathbf{b}\times\mathbf{c}\parallel\,\parallel\mathbf{a}\times\mathbf{c}\parallel}. \tag{6.20}$$

Of course, there are similar formulas for $\cos\alpha$ and $\cos\beta$.

Now

$$\| \mathbf{b} \times \mathbf{c} \| = \sin A, \quad \| \mathbf{a} \times \mathbf{c} \| = \sin B.$$

On the other hand, the triple product identity (see the proof of Proposition 6.4.2) gives

$$(\mathbf{b} \times \mathbf{c}) \cdot (\mathbf{a} \times \mathbf{c}) = (\mathbf{a} \cdot \mathbf{b})(\mathbf{c} \cdot \mathbf{c}) - (\mathbf{b} \cdot \mathbf{c})(\mathbf{a} \cdot \mathbf{c}) = \cos C - \cos A \cos B,$$

using Eq. 6.19. Inserting these formulas in Eq. 6.20 gives formula (i).

For (ii), we have

$$\sin\alpha = \frac{\| (\mathbf{a} \times \mathbf{c}) \times (\mathbf{a} \times \mathbf{b}) \|}{\sin B \sin C} = \frac{\| ((\mathbf{a} \times \mathbf{c}) \cdot \mathbf{b})\mathbf{a} - ((\mathbf{a} \times \mathbf{c}) \cdot \mathbf{a})\mathbf{b} \|}{\sin B \sin C} = \frac{|(\mathbf{a} \times \mathbf{c}) \cdot \mathbf{b}|}{\sin B \sin C}.$$

Hence,

$$\frac{\sin\alpha}{\sin A} = \frac{|(\mathbf{a} \times \mathbf{c}) \cdot \mathbf{b}|}{\sin A \sin B \sin C}. \tag{6.21}$$

Now, the scalar triple product $(\mathbf{a} \times \mathbf{c}) \cdot \mathbf{b}$ is unchanged, up to a sign, by any permutation of the vectors $\mathbf{a}$, $\mathbf{b}$ and $\mathbf{c}$. It follows that the left-hand side of Eq. 6.21 is unchanged under any permutation of the vertices of the triangle. This gives formula (ii). □

As a special case, we have the spherical analogue of Pythagoras' theorem:

Corollary 6.5.4

Suppose that a spherical triangle has sides of length A, B and C and that the angle opposite the side of length C is a right angle. Then,

$$\cos C = \cos A \cos B.$$

The formal analogy between Eqs. 6.19 and 6.20 suggests that we should consider the spherical triangle with vertices

$$\mathbf{a}^* = \frac{\mathbf{b} \times \mathbf{c}}{\| \mathbf{b} \times \mathbf{c} \|}, \quad \mathbf{b}^* = \frac{\mathbf{c} \times \mathbf{a}}{\| \mathbf{c} \times \mathbf{a} \|}, \quad \mathbf{c}^* = \frac{\mathbf{a} \times \mathbf{b}}{\| \mathbf{a} \times \mathbf{b} \|}.$$

Note that the cyclic order $\mathbf{a} \to \mathbf{b} \to \mathbf{c} \to \mathbf{a}$ of the vertices is preserved in these formulas; if the cyclic order was reversed the sign of all three vectors would change. The triangles with vertices $\mathbf{a}^*, \mathbf{b}^*, \mathbf{c}^*$ and $-\mathbf{a}^*, -\mathbf{b}^*, -\mathbf{c}^*$ are called the *dual triangles* of the triangle with vertices $\mathbf{a}, \mathbf{b}, \mathbf{c}$.

Note that each of the two dual triangles is obtained from the other by applying the antipodal map $\mathbf{v} \mapsto -\mathbf{v}$ of S^2; since this is an isometry of $\mathbb{R}^3$ (see Appendix 1), it is also an isometry of S^2 (Exercise 6.1.2) so the two dual triangles have the same angles and sides of the same length. Geometrically, $\pm\mathbf{a}^*$ are the endpoints of the diameter of S^2 perpendicular to the plane that intersects S^2 in the great circle passing through $\mathbf{b}$ and $\mathbf{c}$: they are called the *poles* of this great circle (thus, the north and south poles of S^2 are the poles of the equator).

Note also that $\pm\mathbf{a}$ are the poles of the great circle through $\mathbf{b}^*$ and $\mathbf{c}^*$, since $\mathbf{a}$ is perpendicular to $\mathbf{b}^*$ and $\mathbf{c}^*$. It follows that the dual triangles of the triangle with vertices $\mathbf{a}^*, \mathbf{b}^*, \mathbf{c}^*$ are the original triangle with vertices $\mathbf{a}, \mathbf{b}, \mathbf{c}$ and its image under the antipodal map. This can also be verified algebraically:

$$\begin{aligned} \mathbf{b}^* \times \mathbf{c}^* = \frac{(\mathbf{c}\times\mathbf{a})\times(\mathbf{a}\times\mathbf{b})}{\parallel \mathbf{c}\times\mathbf{a} \parallel\parallel \mathbf{a}\times\mathbf{b} \parallel} &= \frac{((\mathbf{c}\times\mathbf{a})\cdot\mathbf{b})\mathbf{a}}{\parallel \mathbf{c}\times\mathbf{a} \parallel\parallel \mathbf{a}\times\mathbf{b} \parallel}, \\ \therefore \quad \frac{\mathbf{b}^*\times\mathbf{c}^*}{\parallel \mathbf{b}^*\times\mathbf{c}^* \parallel} &= \pm\mathbf{a}, \end{aligned}$$

the sign being that of $(\mathbf{c}\times\mathbf{a})\cdot\mathbf{b} = \mathbf{a}\cdot(\mathbf{b}\times\mathbf{c})$. Thus, the dual triangle of the triangle with vertices $\mathbf{a}^*, \mathbf{b}^*, \mathbf{c}^*$ is the original triangle if $\mathbf{a}\cdot(\mathbf{b}\times\mathbf{c}) > 0$ and is its image under the antipodal map if $\mathbf{a}\cdot(\mathbf{b}\times\mathbf{c}) < 0$.

Proposition 6.5.5

Let α, β, γ and A, B, C be the angles and the lengths of the sides of a spherical triangle, so that α is the angle opposite the side of length A, etc. Let $\alpha^*, \beta^*, \gamma^*, A^*, B^*, C^*$ be the corresponding quantities for either of the dual triangles. Then,

$$\begin{aligned} \alpha^* = \pi - A, \quad \beta^* = \pi - B, \quad \gamma^* = \pi - C, \\ A^* = \pi - \alpha, \quad B^* = \pi - \beta, \quad C^* = \pi - \gamma. \end{aligned}$$

Proof

Denoting the vertices of the triangle by $\mathbf{a}, \mathbf{b}, \mathbf{c}$ as above, Eq. 6.19 gives

$$\cos A^* = \mathbf{b}^*\cdot\mathbf{c}^* = \frac{(\mathbf{c}\times\mathbf{a})\cdot(\mathbf{a}\times\mathbf{b})}{\parallel \mathbf{c}\times\mathbf{a} \parallel\parallel \mathbf{a}\times\mathbf{b} \parallel} = -\cos\alpha,$$

so, since both α and A^* are between 0 and π,

$$A^* = \pi - \alpha. \tag{6.22}$$

The formula $\alpha^* = \pi - A$ is obtained by applying Eq. 6.22 to the dual triangles.

□

Corollary 6.5.6

With the notation in Proposition 6.5.3, we have

$$\cos A = \frac{\cos\alpha + \cos\beta\cos\gamma}{\sin\beta\sin\gamma},$$

together with two similar formulas obtained by making the permutations $A \to B \to C \to A$, $\alpha \to \beta \to \gamma \to \alpha$.

Proof

Just apply part (i) of Proposition 6.5.3 to the dual triangle and use Proposition 6.5.5. □

This formula is important because it shows that *the sides of a spherical triangle are determined by its angles*, unlike the situation in plane geometry in which there are 'similar' triangles with the same angles but possibly different sizes. The 'reason' for this is that in spherical geometry there is an absolute standard of length, namely the radius of the sphere.

Much of Euclidean geometry is concerned with the question of when two geometrical figures (such as triangles) are *congruent*, which means that one figure can be 'moved' so that it coincides with the other. The types of 'motions' that are allowed are those that do not change the size or shape of the triangles, namely the *isometries* of the plane (see Appendix 1). Hence, we need to determine the isometries of the sphere.

We know that any isometry of $\mathbb{R}^3$ that preserves S^2 will give an isometry of S^2 (see Exercise 6.1.2). The following proposition shows that we get all the isometries of S^2 this way (cf. Theorem A.1.5 and its proof).

Proposition 6.5.7

Every isometry of S^2 is a composite of reflections in planes passing through the origin. In fact, at most three reflections are required.

Proof

The first thing to observe is that isometries of S^2 must take great circles to great circles, since these are the curves of shortest length and isometries preserve length.

Let F be any isometry of S^2, and let $\boldsymbol{e}_1 = (1,0,0)$, $\boldsymbol{e}_2 = (0,1,0)$ and $\boldsymbol{e}_3 = (0,0,1)$. If $F(\boldsymbol{e}_1) = \boldsymbol{e}_1$ let G_1 be the identity map. Otherwise, let G_1 be the reflection in the plane perpendicular to the line joining $\boldsymbol{e}_1$ to $F(\boldsymbol{e}_1)$ and passing through its mid-point; note that since $\| \boldsymbol{e}_1 \| = \| F(\boldsymbol{e}_1) \|$, this plane passes through the origin so G_1 is an isometry of S^2. Then $G_1 \circ F$ fixes $\boldsymbol{e}_1$. If $\boldsymbol{e}_2 = G_1(F(\boldsymbol{e}_2))$ let G_2 be the identity map. Otherwise, let G_2 be the reflection in the perpendicular bisector of the line joining $\boldsymbol{e}_2$ and $G_1(F(\boldsymbol{e}_2))$. Since $\| \boldsymbol{e}_2 \| = \| G_1(F(\boldsymbol{e}_2)) \|$ (because F and G_1 are isometries), this plane passes through the origin so G_2 is an isometry of S^2, and since

$$\| \boldsymbol{e}_1 - G_1(F(\boldsymbol{e}_2)) \| = \| G_1(F(\boldsymbol{e}_1)) - G_1(F(\boldsymbol{e}_2)) \| = \| \boldsymbol{e}_1 - \boldsymbol{e}_2 \|,$$

$\boldsymbol{e}_1$ is fixed by G_2. Hence, $G_2 \circ G_1 \circ F$ fixes $\boldsymbol{e}_1$ and $\boldsymbol{e}_2$. Now the north and south poles $\pm\boldsymbol{e}_3$ are the only two points whose spherical distance from $\boldsymbol{e}_1$ and $\boldsymbol{e}_2$ is equal to $\pi/2$, so $G_2 \circ G_1 \circ F$ must either fix $\boldsymbol{e}_3$ or take it to $-\boldsymbol{e}_3$. In the first case let G_3 be the identity, in the second let G_3 be reflection in the xy-plane. Then, $H = G_3 \circ G_2 \circ G_1 \circ F$ is an isometry of S^2 that fixes $\boldsymbol{e}_1$, $\boldsymbol{e}_2$ and $\boldsymbol{e}_3$.

Since H fixes $\boldsymbol{e}_1$ and $\boldsymbol{e}_2$ it must fix each point of the equator, since the equator is the unique great circle passing through these two points and any point on the equator is uniquely determined by its spherical distances from them. Similarly, H must fix each point of the great circle passing through $\boldsymbol{e}_1$ and $\boldsymbol{e}_3$. If $\mathbf{a}$ is any point of S^2 other than the poles $\pm\boldsymbol{e}_3$, the unique great circle $\mathcal{C}$ passing through $\mathbf{a}$ and the poles intersects the equator at a point $\mathbf{b}$, say. Since H fixes $\mathbf{b}$ and the poles, it fixes every point of $\mathcal{C}$ by the previous argument. In particular, H fixes $\mathbf{a}$. Since $\mathbf{a}$ was an arbitrary point of the sphere, H must be the identity map.

Hence, $F = G_1 \circ G_2 \circ G_3$ is a product of ≤ 3 reflections. □

One of the most striking differences between Euclidean and spherical geometry is contained in the following result, which is strongly suggested by Corollary 6.5.6.

Proposition 6.5.8

In spherical geometry, similar triangles are congruent.

This means that if two spherical triangles have vertices $\mathbf{a}, \mathbf{b}, \mathbf{c}$ and $\mathbf{a}', \mathbf{b}', \mathbf{c}'$, and if the angle of the first triangle at $\mathbf{a}$ is equal to that of the second triangle at $\mathbf{a}'$, and similarly for the other two angles, there is an isometry of S^2 that takes $\mathbf{a}$ to $\mathbf{a}'$, $\mathbf{b}$ to $\mathbf{b}'$ and $\mathbf{c}$ to $\mathbf{c}'$. We leave the proof to Exercise 6.5.2.

EXERCISES

6.5.1 Find the angles and the lengths of the sides of an equilateral spherical triangle whose area is one quarter of the area of the sphere.

6.5.2 Show that similar spherical triangles are congruent.

6.5.3 The *spherical circle* of centre $\mathbf{p} \in S^2$ and radius R is the set of points of S^2 that are a spherical distance R from $\mathbf{p}$. Show that, if $0 \leq R \leq \pi/2$:

(i) A spherical circle of radius R is a circle of radius $\sin R$.

(ii) The area inside a spherical circle of radius R is $2\pi(1 - \cos R)$.

What if $R > \pi/2$?

6.5.4 This exercise describes the transformations of $\mathbb{C}_\infty$ corresponding to the isometries of S^2 under the stereographic projection map $\Pi : S^2 \to \mathbb{C}_\infty$ (Example 6.3.5). If F is any isometry of S^2, let $F_\infty = \Pi \circ F \circ \Pi^{-1}$ be the corresponding bijection $C_\infty \to \mathbb{C}_\infty$.

(i) A Möbius transformation

$$M(w) = \frac{aw + b}{cw + d},$$

where $a, b, c, d \in \mathbb{C}$ and $ad - bc \neq 0$, is said to be *unitary* if $d = \bar{a}$ and $c = -\bar{b}$ (see Appendix 2). Show that the composite of two unitary Möbius transformations is unitary and that the inverse of a unitary Möbius transformation is unitary.

(ii) Show that if F is the reflection in the plane passing through the origin and perpendicular to the unit vector (a, b) (where $a \in \mathbb{C}$, $b \in \mathbb{R}$ – see Example 5.3.4), then

$$F_\infty(w) = \frac{-a\bar{w} + b}{b\bar{w} + \bar{a}}.$$

(iii) Deduce that if F is any isometry of S^2 there is a unitary Möbius transformation M such that either $F_\infty = M$ or $F_\infty = M \circ J$ where $J(w) = -\bar{w}$.

(iv) Show conversely that if M is any unitary Möbius transformation, the bijections $\mathbb{C}_\infty \to \mathbb{C}_\infty$ given by M and $M \circ J$ are both of the form F_∞ for some isometry F of S^2.