correspondence between points P_1 of S_1 and points P_2 of S_2 such that

$$\text{distance between } P_1 \text{ and } P_1' \text{ in } S_1 = \text{distance between } P_2 \text{ and } P_2' \text{ in } S_2$$

where the distances are measured *within* the respective surfaces. A more precise statement of the theorema egregium then is: if S_1, S_2 are isometric, then S_1, S_2 have the same Gaussian curvature at corresponding points. The converse statement is not true: there are surfaces S_1, S_2 that are not isometric even though there is a one-to-one (and continuous) correspondence between them for which Gaussian curvature is the same at corresponding points. An example is given in Strubecker (1964, Vol. 3, p. 121), involving surfaces of nonconstant Gaussian curvature.

For surfaces of constant Gaussian curvature there is better agreement between isometry and curvature, as we shall see in the next section. From now on, unless otherwise qualified, "curvature" will mean Gaussian curvature.

17.4 Surfaces of Constant Curvature

The simplest surface of constant positive curvature is the sphere of radius r, which has curvature $1/r^2$ at all points. Other surfaces of curvature $1/r^2$ may be obtained by bending portions of the sphere; however, all such surfaces have either edges or points where they are not smooth, as was proved by Hilbert (1901). The plane, as we have observed, has zero curvature, and so have all surfaces obtained by bending the plane or portions of it.

It remains to investigate whether there are surfaces of constant *negative* curvature. In ordinary space, such a surface has principal curvatures of opposite sign at each point, giving it the appearance of a saddle (Figure 17.7). A number of surfaces of constant negative curvature were given by Minding (1839). The most famous of them is the *pseudosphere*, the surface of revolution obtained by rotating a tractrix about the x-axis (Figure 17.8). This surface was investigated as early as 1693 by Huygens, who found its surface area, which is finite, and the volume and center of mass of the solid it encloses, which is also finite [Huygens (1693a)].

The pseudosphere is in some ways the negative-curvature counterpart of the cylinder, and hence one may wonder whether there is a surface of constant negative curvature that is more like a plane. Hilbert (1901) proved that there is no smooth unbounded surface of constant negative curvature in ordinary space, so this rules out planelike surfaces and also accounts

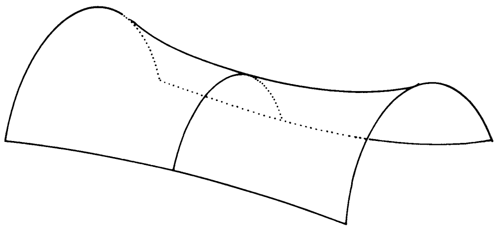

Figure 17.7: A saddle

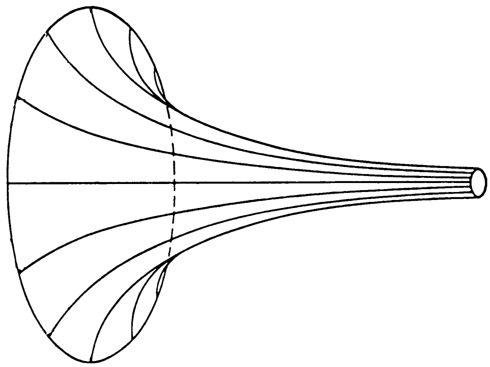

Figure 17.8: The pseudosphere

for the "edge" on the pseudosphere. One can, however, obtain a "plane" of negative curvature by introducing a nonstandard notion of length into the Euclidean plane. This discovery of Beltrami (1868a) is discussed in the next chapter, along with other implications of negative curvature for noneuclidean geometry.

These geometric implications can also be glimpsed if we return to the question of whether surfaces S_1, S_2 of equal curvature are isometric. Even with constant curvature this is still not true, since a plane is not isometric to a cylinder. What *is* true, though, is that any sufficiently small portion of the plane can be mapped isometrically into any part of the cylinder. Minding (1839) showed that the analogous result is true for any two surfaces S_1, S_2 of the same constant curvature. Taking $S_1 = S_2$, this result can be interpreted as saying *rigid motion* is possible within S_1; a body within S_1 can be moved, without any shrinking or stretching, to any part of S_1 large enough to contain it. The latter restriction is necessary, for example, for the pseudosphere since it becomes indefinitely narrow as $x \to \infty$.

The possibility of rigid motion was fundamental to Euclid's geometry of the plane, and with the discovery of curved surfaces that support rigid motion, Euclid's geometry could be seen as a special case—the zero curvature case—of something broader. The broader notion of geometry on a surface begins to take shape once one has an appropriate notion of "straight line." This is developed in the next section.

EXERCISES

The construction of the tractrix as the involute of the catenary in Section 17.2 gives a remarkable insight into the two principal curvatures of the pseudosphere.

17.4.1 Interpreting PQ in Figure 17.5 as the radius of curvature of the tractrix, and hence as the curvature of a section of the pseudosphere, suggest an interpretation of QR as a radius of curvature.

17.4.2 Assuming that PQ and QR are in fact principal radii of curvature, deduce from Exercise 17.2.4 that

$$\text{Gaussian curvature of the pseudosphere at any point} = -1.$$

17.5 Geodesics

A "straight line," or *geodesic* as it is called, can be defined equivalently by a shortest-distance property or a zero curvature property. The shortest-distance definition was historically first, even though it is mathematically deeper and subject to the inconvenience that a geodesic is *not* necessarily the shortest way between two points. On a sphere, for example, there are two geodesics between two nearby points P_1, P_2: the short portion and the long portion of the great circle through P_1, P_2. We can cover both by saying that a geodesic gives the shortest distance between any two of its points that are sufficiently close together. In talking about shortest distance, even between nearby points P_i, P_j, one still has the calculus of variations problem of finding which curve from P_i to P_j has minimum length. Nevertheless, this is how geodesics were first defined, by Jakob and Johann Bernoulli, and Euler (1728a) found a differential equation for geodesics from this approach.

A more elementary approach is to define the *geodesic curvature* κ_g at P of a curve C on a surface S as the ordinary curvature of the orthogonal projection of C in the tangent plane to S at P. As one might expect, geodesic curvature can also be defined intrinsically, and κ_g was introduced in this way by Gauss (1825). A geodesic is then a curve of zero geodesic curvature. This is the definition of Bonnet (1848).

The latter definition immediately shows that great circles on the sphere are geodesics, since their projections onto tangent planes are straight lines. Other examples are the horizontal lines, vertical circles, and helices on the cylinder (Figure 17.9). These all come from straight lines on the plane that is rolled up to form the cylinder. Geodesics on the pseudosphere, and other surfaces of negative curvature, are not all so simple to describe. However, the next chapter shows that they become simple when one maps the surface of constant negative curvature suitably onto a plane.

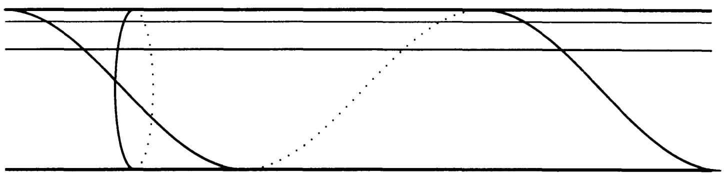

Figure 17.9: Geodesics on the cylinder

EXERCISES

17.5.1 Are the circles on the pseudosphere, in planes perpendicular to its axis, geodesics? Give a qualitative argument to support your answer.

It may be easier to answer this question if one first considers the *cone*, a surface also obtained by bending the plane. To avoid worrying about the apex, where the cone is not smooth, we omit this point.

17.5.2 Show that the circles on the cone, in planes perpendicular to its axis, are *not* geodesics.

17.5.3 Show that there are nonsmooth geodesics on the cone, that is, curves of geodesic curvature zero except at certain points where they have no tangent.

17.6 The Gauss–Bonnet Theorem

In Section 17.2 we observed that

$$\int_{\mathscr{C}} \kappa\, ds = 2\pi$$

for a simple closed curve $\mathscr{C}$ in the plane. This result has a profound generalization to curved surfaces known as the *Gauss–Bonnet* theorem. On a curved surface κ must be replaced by the geodesic curvature κ_g, and the theorem states

$$\int_{\mathscr{C}} \kappa_g \, ds = 2\pi - \iint_{\mathscr{R}} \kappa_1 \kappa_2 \, dA,$$

where A denotes area and $\mathscr{R}$ is the region enclosed by $\mathscr{C}$ [Bonnet (1848)]. Gauss himself published only a special case, or rather the limit of a special case, in which $\mathscr{C}$ is a geodesic triangle. In this case, of course, $\kappa_g = 0$ along the sides of $\mathscr{C}$, and κ_g becomes infinite at the corners. By rounding off the corners by small arcs ds, one sees (Figure 17.10) that

$$\int_{\mathscr{C}*} \kappa_g \, ds \cong \alpha' + \beta' + \gamma',$$

where α', β', γ' are the external angles of the triangle and $\mathscr{C}^*$ is the rounded approximation to the triangle $\mathscr{C}$.

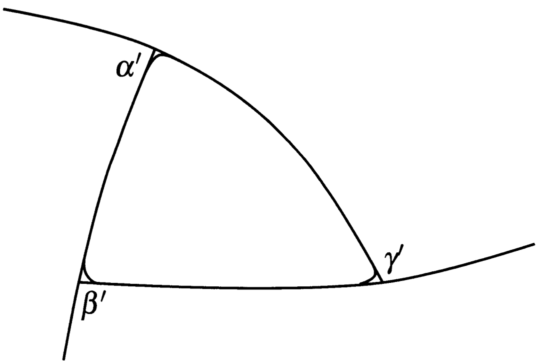

Figure 17.10: Rounding off a geodesic triangle

Then by letting the size of the roundoffs tend to zero one gets

$$\begin{aligned}\int_{\mathscr{C}*} \kappa_g \, ds &= \alpha' + \beta' + \gamma' \\ &= 3\pi - (\alpha + \beta + \gamma),\end{aligned}$$

where α, β, γ are the internal angles of the triangle. Introducing the quantity

$$(\alpha + \beta + \gamma) - \pi$$

called the *angular excess* of the triangle (because an ordinary triangle has angle sum π), we have

$$\int_{\mathscr{C}} \kappa_g \, ds = 2\pi - \text{angular excess},$$

and the result of Gauss (1827) was that

$$\text{angular excess} = \iint_{\mathscr{R}} \kappa_1 \kappa_2 \, dA.$$

We see that the *integral* of the Gaussian curvature has a more elementary geometric meaning than the curvature $\kappa_1 \kappa_2$. It appears, in fact, that Gauss thought about angular excess first, then the curvature integral, and only last about the curvature itself. The decomposition into principal curvatures probably came later, when he reworked his geometric ideas into analytic form, reversing the order of discovery in the process. Dombrowski (1979) made a plausible reconstruction of the original approach, using clues from the unpublished work of Gauss.

The role of angular excess can be seen more plainly in the case of constant curvature $\kappa_1 \kappa_2 = c$. In this case

$$\text{angular excess} = c \times \text{area of } \mathscr{R},$$

so the angular excess gives a measure of area, a result Gauss claimed, in a letter (1846a), to have known in 1794. In fact, the special case of this result for the sphere was known to Thomas Harriot in 1603 [see Lohne (1979)]. Harriot's elegant proof goes as follows (Figure 17.11).

Prolonging the sides of triangle ABC yields a partition of the sphere into four pairs of congruent, diametrically opposite triangles (Figure 17.11a). We denote the area of ABC and its diametric opposite $A'B'C'$ by $\Delta_{\alpha\beta\gamma}$. The other three pairs represent areas $\Delta_\alpha, \Delta_\beta, \Delta_\gamma$, which complement $\Delta_{\alpha\beta\gamma}$ in "slices" of the sphere of angles α, β, γ, respectively (Figure 17.11b).

Since the area of a slice is $2r^2$ times the angle, where r is the radius of the sphere, we have

$$\Delta_{\alpha\beta\gamma} + \Delta_\alpha = 2r^2\alpha,$$
$$\Delta_{\alpha\beta\gamma} + \Delta_\beta = 2r^2\beta,$$
$$\Delta_{\alpha\beta\gamma} + \Delta_\gamma = 2r^2\gamma,$$

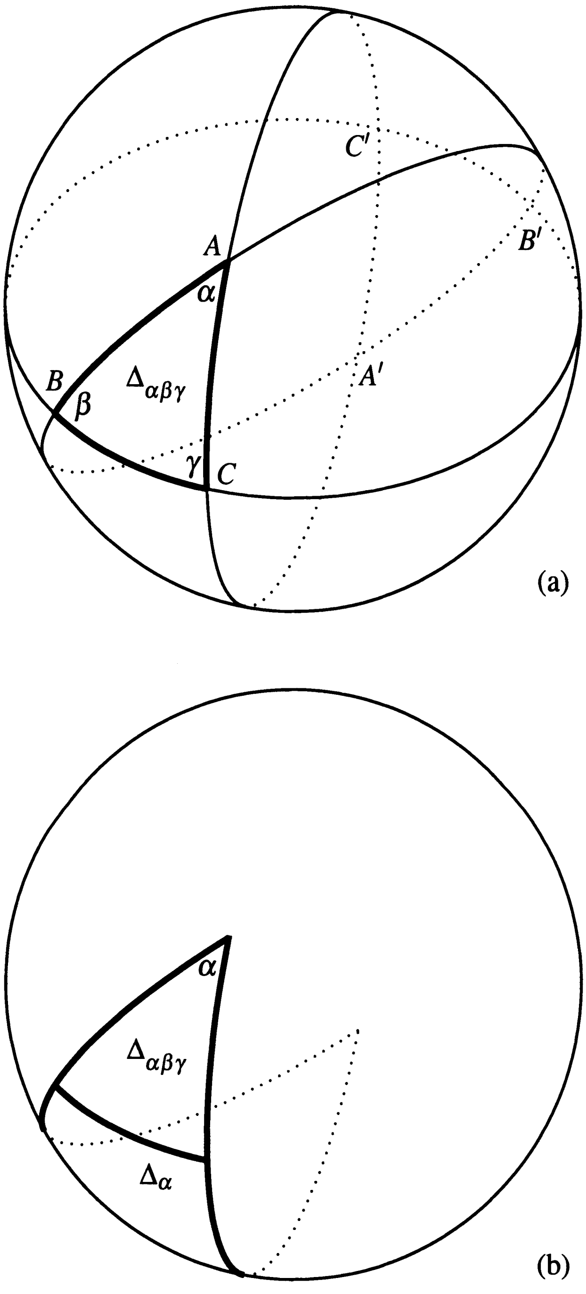

Figure 17.11: Area of a spherical triangle

whence, by addition

$$3\Delta_{\alpha\beta\gamma} + (\Delta_\alpha + \Delta_\beta + \Delta_\gamma) = 2r^2(\alpha + \beta + \gamma). \tag{1}$$

On the other hand,

$$2(\Delta_{\alpha\beta\gamma} + \Delta_\alpha + \Delta_\beta + \Delta_\gamma) = \text{area of sphere} = 4\pi r^2,$$

and substituting this in (1) gives

$$\Delta_{\alpha\beta\gamma} = r^2(\alpha + \beta + \gamma - \pi)$$

as required, since $1/r^2$ = curvature of the sphere.

Gauss was interested in the counterpart of this result for negative curvature, in which case the angle sum of a triangle is less than π and one has angular *defect* rather than angular excess. His investigations in this case led him not only to Gaussian curvature but also to noneuclidean geometry.

Exercises

It is surprising at first that area on the sphere should be measured by angles rather than lengths. However, there are general reasons (apart from the Gauss–Bonnet theorem) why area should be measured by angular excess, and this idea only fails where angular excess is zero—that is, in the Euclidean plane.

17.7.1 Consider a triangle Δ split into two triangles Δ_1 and Δ_2 by a line through a vertex. Show that

$$\text{excess}(\Delta) = \text{excess}(\Delta_1) + \text{excess}(\Delta_2).$$

17.7.2 Deduce from Exercise 17.7.1 that if any polygon Π is split into triangles Δ_i, then

$$\text{excess}(\Pi) = \text{excess}(\Delta_1) + \text{excess}(\Delta_2) + \cdots.$$

Thus the angular excess function has the same *additive* property as an area function. It can be shown that any additive function, provided it is continuous, is a constant multiple of area [see Bonola (1912), p. 46].

17.7 Biographical Notes: Harriot and Gauss

The discoveries of Thomas Harriot described in this chapter and the last seem to entitle him to a secure place in the history of mathematics, perhaps alongside others who made a few penetrating contributions, such as

Desargues and Pascal. Unfortunately, Harriot's place is not yet clear. It was clouded by exaggerated claims made by seventeenth- and eighteenth-century admirers, and until recently the disorder and inaccessibility of his papers have made any claims difficult to verify. In addition, Harriot was a very secretive man, and little is known about his life. He lived in the world of Sir Walter Raleigh, Christopher Marlowe, and Guy Fawkes—a lurid and fascinating world, but a very dangerous one—and probably believed that secrecy was necessary for his survival. As a result, our present understanding of Harriot [see the biography by Shirley (1983)] is based on a meager set of facts about him and a good deal of extrapolation from knowledge of his less discreet contemporaries.

All that we know of Harriot's early life comes from a record of his entry into Oxford University in December 1577, stating that his age was then 17 and his father "plebeian." The only other information about his family comes from his will of 1621, which mentions a sister and a cousin. It seems probable that he never had children and never married. At Oxford, Harriot gained the standard bachelor's degree in classics, but he must have picked up some Euclid and astronomy, which were offered to master's candidates. He would also have heard Richard Hakluyt, author of the famous *Voyages*, who was then just beginning to lecture on the geography of the New World opened up by sixteenth-century navigators.

It was probably Hakluyt who inspired Harriot to travel to London in the early 1580s and seek out Sir Walter Raleigh. Raleigh was then about 30 and the most powerful member of Queen Elizabeth's inner circle, with grand dreams of wealth through exploration. Harriot must have impressed Raleigh with his grasp of the mathematical problems of navigation, for around 1583 he joined Raleigh's household as an instructor, with considerable freedom to conduct his own research. Harriot held classes in navigation as part of Raleigh's preparations for the voyage to Virginia in 1585, led by Sir Richard Grenville, which was the first attempt at British settlement in the New World. Although the attempt was unsuccessful, it was the biggest adventure of Harriot's life. He studied Indian languages and customs and wrote a book on the settlement entitled *A Brief and True Report of the New Found Land of Virginia* (1588), the only one of Harriot's works published in his lifetime.

With Raleigh as patron, Harriot was financially secure, and he remained so for the rest of his life. However, he was also at the mercy of Raleigh's political fortunes. By 1592, the 40-year-old Raleigh was finding his role as the favorite of the nearly 60-year-old queen increasingly irksome, and he secretly married one of the queen's servants, Elizabeth Throckmorton. He may have married her as early as 1588, but at any rate the secret was out when Lady Raleigh gave birth to a son in 1592, and Raleigh was imprisoned in the Tower of London. Harriot did not suffer for his direct association with Raleigh, but through him he was linked with Christopher Marlowe, at the latter's sensational trial for atheism in 1593.

Marlowe, the dramatist, had a secret life in espionage and other unsavory activities, and any number of accusations could have been made against him, though which ones were true it is now impossible to say. Unfortunately for Harriot, the second of Marlowe's offenses against religion was said to be that "He affirmeth that Moyses was but a Jugler, and that one Heriots being Sir W. Raleighs man can do more than he." As it happened, the proceedings were terminated by the murder of Marlowe in a tavern brawl, and Harriot was not called to testify, but he was left publicly under suspicion.

Harriot did not desert Raleigh, but he was prudent enough to seek another patron, and he found one in Henry Percy, the Ninth Earl of Northumberland. Henry, known as the "Wizard Earl," was a friend of Raleigh and, like him, interested in science and philosophy. In 1593 he gave Harriot a grant, later to become an annual pension of £80. This sum was twice the salary of the best-paid teachers of the time and it enabled Harriot to maintain a house and servants on the earl's property on the Thames near London. The house, known as Sion House, remained Harriot's home and laboratory for the rest of his life.

But once again Harriot was unlucky in his choice of friends. The earl's cousin, Thomas Percy, was the man who rented the cellars under the Houses of Parliament in the famous plot to blow up King James I with gunpowder on November 5, 1605. Harriot was dragged into the investigation and imprisoned for a short time on suspicion that he had secretly cast a horoscope of the king. James I was terrified of black magic and indiscriminately viewed all mathematicians as astrologers and magicians. In the end, though, no evidence against Harriot was found, and it was the earl who suffered more, being imprisoned in the Tower from 1605 to 1621.

Meanwhile, Raleigh had fared even worse. After several spells in the Tower, he was released in 1616 to lead an expedition in search of the mythical city of gold, El Dorado. When the expedition returned in a shambles, Raleigh was rearrested and executed on an old treason charge from 1603.