13

A Mapmaker's Paradise

What's the good of Mercator's North Poles and
Equators, Tropics, Zones and Meridian Lines?
So the Bellman would cry: and the crew would reply,
"They are merely conventional signs!"
—Lewis Carroll (Charles Dodgson), *The Hunting of the Snark* (1876)

From the sublime beauty of Euler's formulas we now turn to a more mundane matter: the science of map making. It is common knowledge that one cannot press the peels of an orange against a table without tearing them apart: no matter how carefully one tries to do the job, some distortion is inevitable. Surprisingly, it was not until the middle of the eighteenth century that this fact was proved mathematically, and by none other than Euler: his theorem says that it is impossible to map a sphere onto a flat sheet of paper without distortion. Had the earth been a cylinder, or a cone, the mapmaker's task would have been easier: these surfaces are *developable*—they can be flattened without shrinking or stretching. This is because these surfaces, though curved, have essentially the geometry of a plane. But the underlying geometry of a sphere is fundamentally different from that of a plane; consequently, one cannot create a map of the earth that faithfully reproduces *all* its features.

To cope with this problem, cartographers have devised a variety of *map projections*—functions (in the mathematical sense) that assign to every point on the sphere an "image" point on the map. The choice of a particular projection depends on the intended purpose of the map; one map may show the correct distance between two points on the globe (of course up to a scaling factor), another the relative area of countries, and yet another the direction between two points. But preserving any of these features always comes at the expense of other features: every map projection is a compromise between conflicting demands.

The simplest of all is the *cylindrical projection*: imagine the earth—represented by a perfectly spherical globe of radius R—to be wrapped in a cylinder touching it at the equator (fig. 77). Imagine further that rays of light emanate from the center of the globe in all directions. A point P on the globe is then projected onto a point P', the "shadow" or "image" of P on the cylinder. When the cylinder is unwrapped, we get a flat map of the entire earth—or *almost* the entire earth: the North and South Poles, being on the axis of the cylinder, have their images at infinity.

Clearly the cylindrical projection maps all circles of longitude (meridians) onto equally spaced vertical lines, while circles of latitude (or "parallels," as they are known in geography) show as horizontal lines whose spacing increases with latitude. In order to find the relation between a point P and its image P', we must first express the location of P in terms of its *longitude* (measured eastward or westward along the equator from the prime meridian through Greenwich, England) and *latitude* (measured northward or southward from the equator along any meridian). Denoting the longitude and latitude of P by the Greek letters λ (lambda) and ϕ and the coordinates of P' by x and y, we have

$$x = R\lambda,\ y = R\tan\phi. \qquad (1)$$

The most striking feature of the cylindrical projection is the excessive north-south stretching at high latitudes, resulting in a drastic distortion of the shape of continents (fig. 78); this, of course, is a consequence of the presence of $\tan\phi$ in the second of equations (1). The cylindrical projection has often been

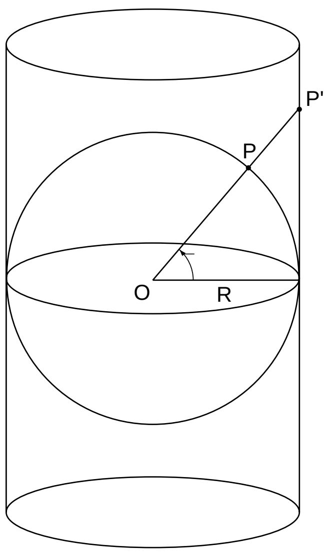

FIG. 77. Cylindrical projection of the globe.

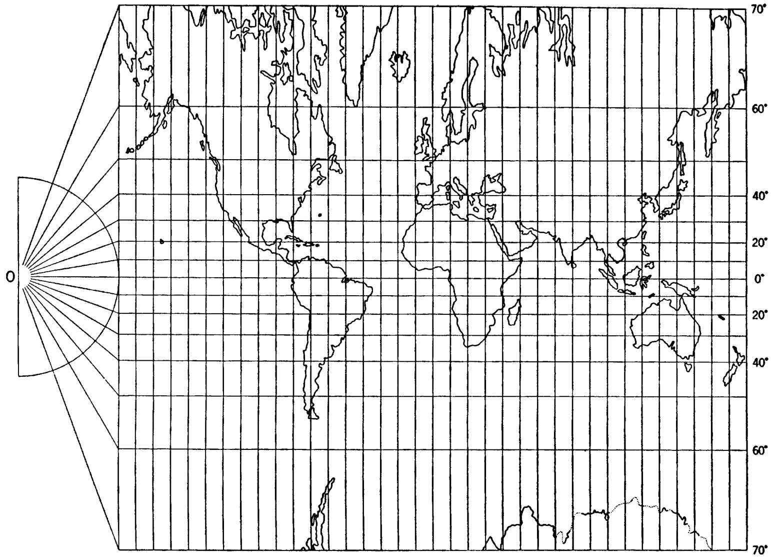

Fig. 78. World map on a cylindrical projection.

confused with Mercator's projection, which it resembles superficially; however, except for the fact that both use a rectangular grid, the two projections are based on entirely different principles, as we will shortly see.

A second projection, known already to Hipparchus in the second century *b.c.*, is the *stereographic projection*. We place the globe on a flat sheet of paper, touching it at the South Pole S (fig. 79). We now connect every point P on the globe by a straight line to the *North* Pole N and extend this line until it meets the plane of the map at the point P'; P' is the image of P under the projection.

The stereographic projection shows all meridians as straight lines radiating from the South Pole S, while circles of latitude show as concentric circles around S. The equator goes over to a circle e, which we may think of as the unit circle. The entire northern hemisphere is then mapped onto the exterior of e and

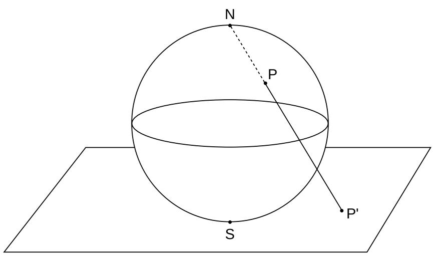

Fig. 79. Stereographic projection from the North Pole.

the southern hemisphere onto its interior. The closer a point is to the North Pole, the farther out will its image be on the map. There is one point on the globe with no image on the map: the North Pole itself. Its image is at infinity.

Let the globe have a unit *diameter*; this will ensure that circle e (the equator on the map) has a unit radius. Consider now a point P with latitude ϕ on the globe. We wish to determine the location of its image P' on the map. Figure 80 shows a cross section of the globe, with E representing a point on the equator; we have $SN = 1$, $\angle ONE = 45°$, $\angle EOP = \phi$, and $\angle ENP = \phi/2$. Therefore, $\angle ONP = (45° + \phi/2)$ and thus P' is located at a distance

$$SP' = \tan(45° + \phi/2) \tag{2}$$

from the South Pole on the map.

Equation (2) leads to an interesting result. Let P and Q be two points with the same longitude but opposite latitudes on the globe. How will their images be related on the map? Replacing ϕ with $-\phi$ in equation (2), we have

$$SQ' = \tan(45° - \phi/2) = \frac{1 - \tan\phi/2}{1 + \tan\phi/2}$$
$$= \frac{1}{\tan(45° + \phi/2)} = \frac{1}{SP'}.$$

Thus $SP' \cdot SQ' = 1$. Two points in the plane fulfilling this condition are said to be *inverse points* with respect to the unit circle; thus the stereographic projection sends two points with equal but opposite latitudes on the globe to two mutually inverse points on the map. This allows us to deduce all properties of the stereographic projection from the theory of inversion. It is known, for example, that the angle of intersection of two curves remains unchanged, or *invariant*, when each of the curves is subjected to inversion. From this one can show that the stereographic projection is direction-preserving, or *conformal*; that is, small regions on the globe preserve their shape on the map (hence the name conformal).[1] Figure 81 shows the northern

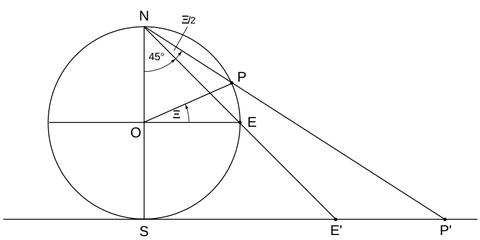

FIG. 80. Geometry of the stereographic projection.

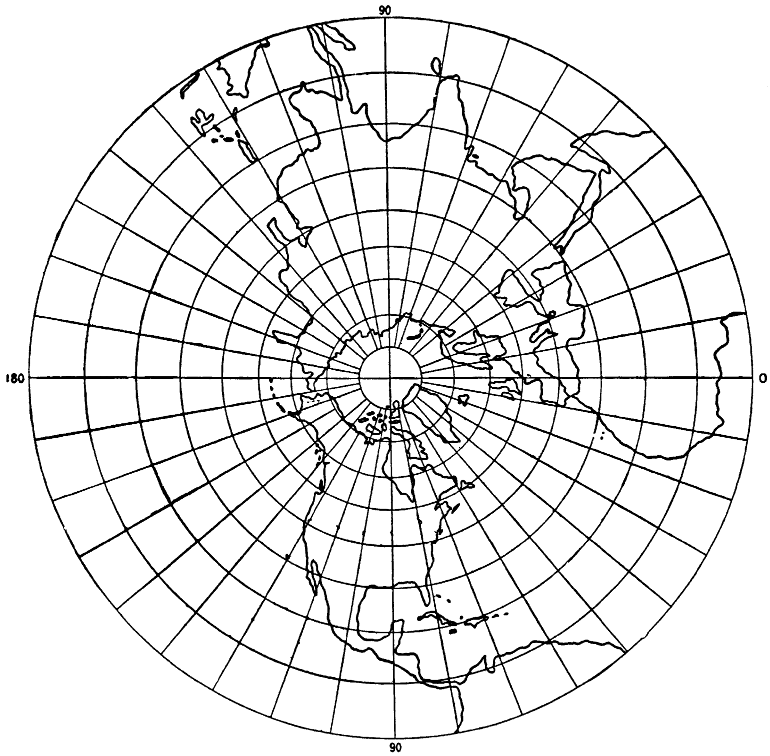

FIG. 81. The northern hemisphere on a stereographic projection.

hemisphere on a stereographic projection in which the globe touches the map at the North (rather than South) Pole; clearly the shape of continents is close to what it is on the globe.

✧ ✧ ✧

In chapter 10 we encountered the azimuthal equidistant map, which shows the true distance and direction from a given fixed point to any other point on the globe. However, distances and directions between any other two points are *not* preserved on this map; hence its usefulness for navigation is greatly limited. One would rather have a map that shows the correct direction, or compass bearing, from *any* point on the globe to *any* other point. But up until the middle of the sixteenth century no such map existed.

Imagine you are the navigator of a boat about to leave port headed in a certain direction. You set your compass at your chosen bearing, say 45°east of north, and then follow that bearing steadfastly, ignoring—for the sake of argument—any land masses that might be in your way. What path will you follow? For many years it had been believed that a path of constant bearing—known as a *rhumb line* or *loxodrome*[2]—is an arc of a great circle (see p. 136). But the Portuguese Pedro Nuñes (or Nonius, 1502–1578) showed that the rhumb line is actually a spiral curve that gets ever closer to either pole, winding around it indefinitely but never reaching it. The Dutch artist Maurits C. Escher (1898–1972) has depicted the rhumb line in one of his works, *Sphere Surface with Fish* (1958), shown in figure 82.

The challenge that faced cartographers in the sixteenth century was to design a map projection *that would show all rhumb lines as straight lines*. Such a map would enable a navigator to join his points of departure and destination by a straight line, measure the angle, or bearing, between this line and the north, and then follow this bearing at sea. On all existing projections, however, a straight-line course laid off on the map did not correspond to a rhumb line at sea. As a result, navigation was an

Fig. 82. Maurits C. Escher's *Sphere Surface with Fish* (1958).

extremely tricky business—and a risky one too, for many lives were lost due to ships failing to reach their destination. It befell a Flemish mapmaker to come to the mariners' rescue.

✧ ✧ ✧

Gerardus Mercator, by general consensus the most famous mapmaker in history, was born Gerhard Kremer in Rupelmonde in Flanders (now Belgium but then part of Holland) on March 5, 1512. Only twenty years earlier Christopher Columbus had made his historic voyage to the New World, and young Kremer's imagination was fired by the new geographical discoveries. He entered the University of Louvain in 1530, and soon after graduating established himself as one of Europe's leading mapmakers and instrument designers. As was customary for learned men at the time, he Latinized his name to Mercator ("merchant," a literal translation of the Dutch word *kramer*), and it is by this name that he has been known ever since.

Mercator's promising career was threatened in 1544 when he was arrested as a heretic for practicing Protestantism in a Catholic country. He barely saved his life and subsequently fled to neighboring Duisburg (now Germany), where he settled in 1552. He remained there for the rest of his life.[3]

Before Mercator, mapmakers decorated their charts with fanciful mythological figures and imaginary lands of their own creation: their maps were more works of art than true representations of the earth. Mercator was the first to base his maps entirely on the most recent data collected by explorers, and in so doing he transformed cartography from an art to a science. He was also one of the first to bound in one volume a collection of separate maps, calling it an "atlas" in honor of the legendary globe-holding mythological figure that decorated the title page; this work was published in three parts, the last appearing in 1595, one year after his death.[4]

It was in 1568 that Mercator set himself the task of inventing a new map projection that would answer the mariner's needs and change global navigation from a haphazard, risky endeavor to an exact science. From the outset he was guided by two principles: the map was to be laid out on a rectangular grid, with all circles of latitude represented by horizontal lines parallel to the equator and equal to it in length, and all meridians showing as vertical lines perpendicular to the equator; and the map would be *conformal*, for only such a map could preserve the true direction between any two points on the globe.

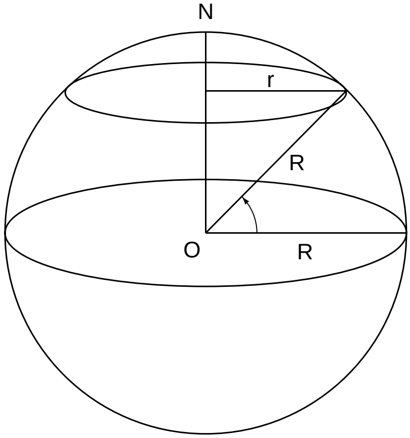

FIG. 83. Circle of latitude ϕ on the globe.

Now on the globe, the cirlces of latitude decrease in size as their latitude increases, until they shrink to a point at either pole. But on Mercator's map these same circles show as horizontal lines of *equal* length. Consequently, each parallel on the map is stretched horizontally (i.e., in an east-west direction) by a factor that depends on the latitude of that parallel. Figure 83 shows a circle of latitude ϕ. Its circumference is $2\pi r = 2\pi R \cos\phi$ on the globe, whereas on the map its length is $2\pi R$; it is thus stretched by a factor $2\pi R/(2\pi R \cos\phi) = \sec\phi$. Note that this stretching factor is *a function of* ϕ: the higher the latitude, the greater the stretching ratio, as table 6 shows.

And now Mercator was ready to produce his trump card: in order for the map to be conformal, the east-west stretching of the parallels must be accompanied by an equal north-south stretching of the spacing *between* the parallels, and this north-south stretching progressively increases as one goes to higher latitudes. In other words, the degrees of latitude, which on the globe are equally spaced along each meridian, must gradually be increased on the map (fig. 84). This is the key principle behind his map.

However, in order to implement this plan, the spacing between successive parallels had first to be determined. Exactly how Mercator did this is not known (and is still being debated

Table 6. Sec ϕ for Some Selected Latitudes

ϕ	0°	15°	30°	45°	60°	75°	80°	85°	87°	89°	90°
$\sec\phi$	1.00	1.04	1.15	1.41	2.00	3.86	5.76	11.47	19.11	57.30	∞

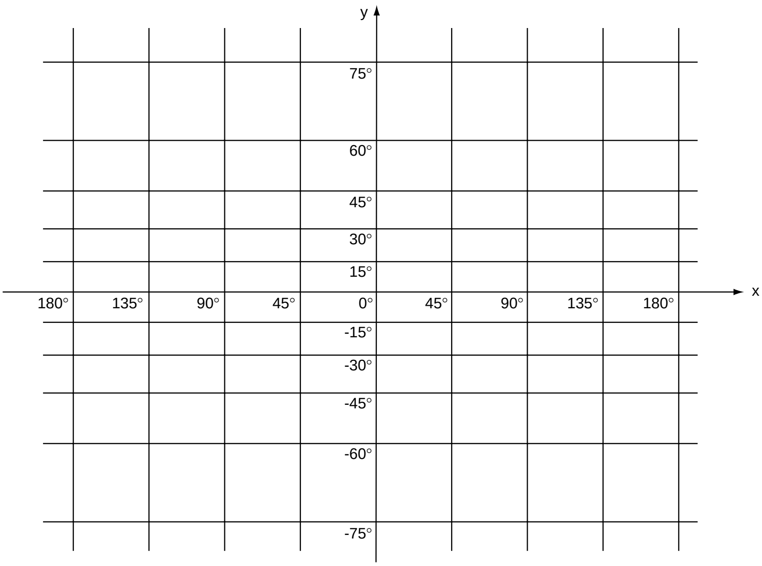

FIG. 84. The Mercator grid.

by historians of cartography);[5] he left no written record of his method except for the following brief explanation which was printed on his map:

> In making this representation of the world, we had to spread on a plane the surface of the sphere in such a way that the positions of places shall correspond on all sides with each other both in true direction and in distance. . . .[6] With this intention we had to employ a new proportion and a new arrangement of the meridians with reference to the parallels. . . . For these reasons we have progressively increased the degrees of latitude towards each pole in proportion to the lengthening of the parallels with reference to the equator.[7]

Even this vague explanation makes it clear that Mercator had a full grasp of the mathematical principles underlying his map. Having created his grid, it now remained for him to put skin to skeleton—to superimpose on this grid the outline of the continents as they were known in his time. He published his world map (or "chart," as mariners prefer to call it) in 1569 under the title *New and Improved Description of the Lands of the World, amended and intended for the Use of Navigators*. It was a huge map, printed on twenty-one sections and measuring 54 by 83 inches. It is one of the most treasured cartographical artifacts of all time: only three copies of the original are known to survive.[8]

✧ ✧ ✧

Mercator died in Duisburg on December 2, 1594, having lived a long life that brought him fame and wealth. Yet his most famous achievement, the map that bears his name, was not immediately embraced by the maritime community, who could not understand its excessive distortion of the shape of continents. The fact that Mercator had not given a full account of how he had "progressively increased" the distance between the parallels only added to the confusion. It was left to Edward Wright (ca. 1560–1615), an English mathematician and instrument maker, to give the first accurate account of the principles underlying Mercator's map. In a work entitled *Certaine Errors in Navigation . . .* , published in London in 1599, he wrote:

> The parts of the meridians at euery poynt of latitude must increase with the same proportion wherewith the Secantes increase. By perpetuall addition of the Secantes answerable to the latitude of each parallel vnto the summe compounded of all former secantes . . . we may make a table which shall truly shew the points of latitude in the meridians of the nautical planisphaere.[9]

In other words, Wright used *numerical integration* to evaluate $\int_0^\phi \sec\phi \, d\phi$. Let us follow his plan, using modern notation.

Figure 85 shows a small spherical rectangle defined by the circles of longitudes λ and $\lambda + \Delta\lambda$ and circles of latitudes ϕ and $\phi + \Delta\phi$, where λ and ϕ are measured in radians. (Because the choice of the "zero meridian" is arbitrary, only the difference in longitude $\Delta\lambda$ is shown in the figure.) The sides of this rectangle have length $(R\cos\phi)\,\Delta\lambda$ and $R\,\Delta\phi$, respectively. Let a point $P(\lambda, \phi)$ on the sphere go over to the point $P'(x, y)$ on the map (where $y = 0$ corresponds to the equator). Then the spherical rectangle will be mapped onto a planar rectangle defined by the lines $x, x + \Delta x, y$, and $y + \Delta y$, where $\Delta x = R\,\Delta\lambda$. Now, the requirement that the map be conformal means that these two rectangles must be *similar* (which in turn means that the direction from $P(\lambda, \phi)$ to a neighboring point $Q(\lambda + \Delta\lambda, \phi + \Delta\phi)$ is the same as between their images on the map). Thus we are led to the equation

$$\frac{\Delta y}{R\,\Delta\lambda} = \frac{R\,\Delta\phi}{R\cos\phi\,\Delta\lambda}$$

or

$$\Delta y = (R\sec\phi)\,\Delta\phi. \qquad (3)$$

In modern terms, equation (3) is a *finite difference* equation. It can be solved numerically by a step-by-step procedure: we put

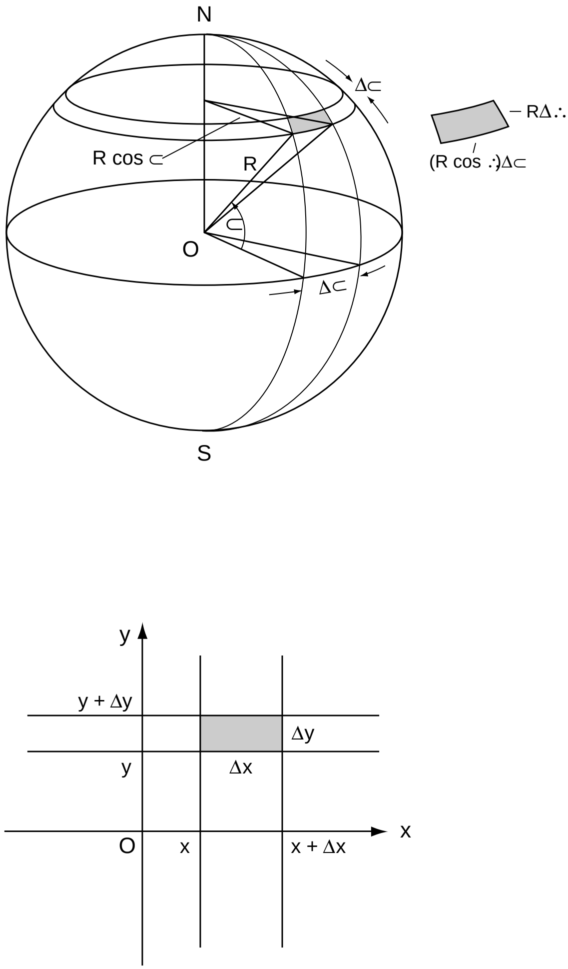

FIG. 85. Spherical rectangle on the globe and its projection on Mercator's map.

$\Delta y_i = y_i - y_{i-1}$, $i = 1, 2, 3, \ldots$, and decide on a fixed increment $\Delta\phi$. Starting with the equator ($y_0 = 0$), we increase ϕ by $\Delta\phi$, find Δy_1 from equation (3), and then find $y_1 = y_0 + \Delta y_1$; we again increase ϕ by $\Delta\phi$ and find $y_2 = y_1 + \Delta y_2$, and so on until the desired range of latitudes has been covered. This *numerical integration* is a tedious, time-consuming procedure—unless one has a programmable calculator or computer, neither of which was available to Wright. Nevertheless he carried the plan through, continually adding the secants at intervals of one minute of arc.[10] He published his results in a table of "meridional parts" for latitudes from 0° to 75°. So at last the method of construction of Mercator's map became known.

Nowadays, of course, we would write equation (3) as a *differential equation*: we let both $\Delta\phi$ and Δy become infinitely small, and in the limit get

$$\frac{dy}{d\phi} = R \sec \phi, \tag{4}$$

whose solution is

$$y = R \int_0^{\phi} \sec t \, dt \tag{5}$$

(we have used t instead of ϕ in the integrand to distinguish the variable of integration from the upper limit of the integral). Today this integral is given as an exercise in a second-semester calculus class (we will say more about it shortly). But Wright's book appeared some seventy years before Newton and Leibniz invented the calculus, so he could not avail himself of its techniques. He had no choice but to resort to numerical integration.

Being a scholar, Wright wrote his book for readers versed in mathematics. But to the ordinary mariner such theoretical explanations meant very little. So Wright devised a simple physical model that he hoped would explain to the uninitiated the principles behind Mercator's map: imagine we wrap the globe in a cylinder touching it along the equator. Let the globe "swell like a bladder" so that each point on its surface comes into contact with the cylinder. When the cylinder is unwrapped, you get a Mercator map.

Unfortunately for posterity—and due to no fault of Wright's—this descriptive model became the source of a persisting myth: that Mercator's map is obtained by projecting rays of light from the center of the globe to the wrapping cylinder (this actually produces the cylindrical projection we discussed earlier in this chapter). Technically, Mercator's "projection" is not a projection at all, at least not in the geometric sense of the word: it can only be obtained by a mathematical procedure that, at its core, involves an infinitesimal process and hence the calculus. Mercator himself never used the cylinder concept, and his projection—except for a superficial similarity—has nothing to do with the cylindrical projection. But once created, myths are slow to die, and even today one finds erroneous statements to that effect in many geography textbooks.

Additional misunderstandings came from the map's excessive distortion of lands at high latitudes: Greenland, for example, appears larger than South America, though in reality it is only one-ninth as big. Moreover, a straight-line segment connecting two points on the map does *not* represent the shortest distance

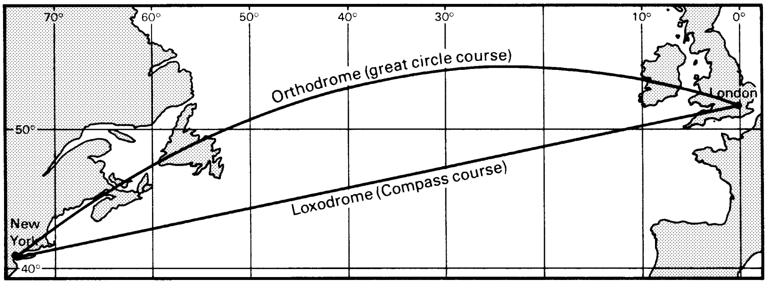

FIG. 86. Rhumb line (loxodrome) and great-circle arc on Mercator's projection.

between them on the globe (unless both points are on the equator or on the same meridian), as figure 86 shows. These "shortcomings" have often been used to criticize Mercator's map. An anonymous contemporary of Wright, evidently offended by this unjust criticism, vented his frustration in these lines:

Let no one dare to attribute the shame
of misuse of projections to Mercator's name;
but smother quite, and let infamy light
upon those who do misuse, publish or recite.

As we have seen, Mercator himself had already realized that no single map could at once preserve distance, shape, *and* direction. Having the navigator's need in mind, he chose to sacrifice distance and shape in order to preserve direction. Yet many people's perception of the world still comes from the large Mercator map that hung from the wall of their high school classroom.

✧ ✧ ✧

Wright's book appeared in 1599, thirty years after Mercator published his new world map. Slowly the maritime community began to appreciate the great value of the map to navigators, and in due time it became the standard map for global navigation, a status which it has kept ever since. When *nasa* began its space exploration in the 1960s, a huge Mercator map, on which the trajectories of satellites were being continuously monitored, dominated the Mission Control room in Houston, Texas. And the first maps of the moons of Jupiter and Saturn, photographed from close range by the Pioneer and Voyager spacecraft, were drawn on his projection.

But let us return to the seventeenth century, where the story now shifts to the mathematical arena. In 1614 John Napier (1550–1617) of Scotland published his invention of logarithms, the single most important aid to computational mathematics since the Hindu-Arabic numeration system was brought to Europe in the Middle Ages.[11] Shortly thereafter Edmund Gunter (1581–1626), an English mathematician and clergyman, published a table of logarithmic tangents (1620). Around 1645 Henry Bond, a mathematics teacher and authority on navigation, compared this table with Wright's meridional table and noticed to his surprise that the two tables matched, provided the entries in Gunter's table were written as $(45^\circ + \phi/2)$. He conjectured that $\int_0^\phi \sec t\, dt$ *is equal to* $\ln \tan\,(45^\circ + \phi/2)$, where "ln" stands for natural logarithms (logarithms to the base $e = 2.718\ldots$), but he could not prove it. Soon his conjecture became one of the outstanding mathematical problems of the 1650s. It was unsuccessfully attempted by John Collins, Nicolaus Mercator (no relation to Gerhard), Edmond Halley (of comet fame), and others—all contemporaries of Isaac Newton and active participants in the developments that led to the invention of the calculus.

Finally in 1668 James Gregory, whom we have already met in connection with the Gregory-Leibniz series, succeeded in proving Bond's conjecture; his proof, however, was so difficult that Halley denounced it as being full of "complications." So it befell Isaac Barrow (1630–1677), who preceded Newton as the Lucasian Professor of Mathematics at Cambridge University, to give an "intelligent" proof of Bond's conjecture (1670). And in so doing he seems to have been the first to use the technique of decomposition into partial fractions, so effective in solving numerous indefinite integrals. The details of his proof are given in Appendix 2.

✧ ✧ ✧

We are now in a position to write down the coordinates (x, y) of a point P' on Mercator's Map in terms of longitude λ and latitude ϕ of the corresponding point P on the globe. The difference equation $\Delta x = R\,\Delta\lambda$ (p. 174) has the obvious solution $x = R\lambda$, and the integral appearing in equation (5) is equal to $\ln \tan\,(45^\circ + \phi/2)$; we thus have

$$x = R\lambda,\ y = R \ln \tan\,(45^\circ + \phi/2).^{12} \qquad (6)$$

Figure 87 shows the world as it appears on Mercator's map; because of the excessive north-south stretching at high latitudes, the map is confined to latitudes from 75° north to 60° south.

But our story does not quite end yet. The reader may have noticed that the expression $\tan\,(45^\circ + \phi/2)$ inside the logarithm

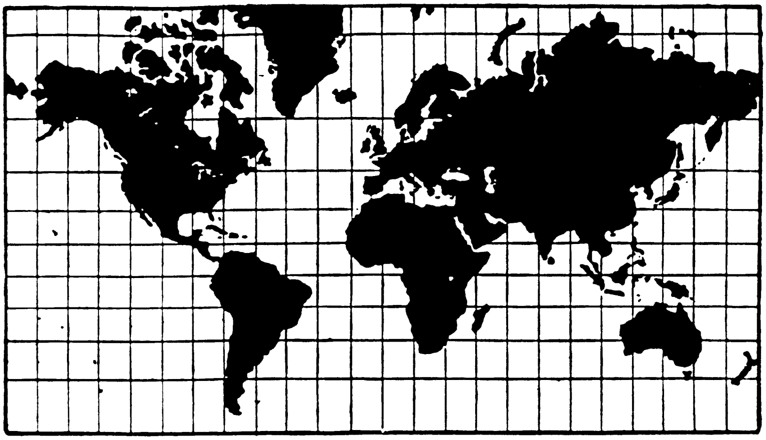

Fig. 87. The world on Mercator's map.

in equation (6) is the same as that appearing in equation (2) in connection with the stereographic projection. This is no coincidence. One of the great achievements of eighteenth-century mathematics was to extend the algebra of ordinary functions such as $\sin x$, e^x, and $\ln x$ to *imaginary* and even *complex* values of the variable x. This development began with Euler and reached its climax in the nineteenth century with the theory of functions of a complex variable. As we shall see in the next chapter, this extension enables us to regard Mercator's projection as a conformal *mapping* (in the mathematical as well as geographical sense) of the stereographic projection by means of the function $w = \ln u$, where both u and w are complex variables.

Notes and Sources

1. A detailed discussion of the stereographic projection and its relation to inversion can be found in my book *To Infinity and Beyond: A Cultural History of the Infinite* (1987: rpt. Princeton, N.J.: Princeton University Press, 1991), pp. 95–98 and 239–245. For other map projections, see Charles H. Deetz and Oscar S. Adams, *Elements of Map Projection* (New York: Greenwood Press, 1969), and John P. Snyder, *Flattening the Earth: Two Thousand Years of Map Projections* (Chicago: University of Chicago Press, 1993).

2. From the Greek *loxos* = slanted, and *dromos* = course, i.e., a slanted line. The term was coined in 1624 by the Dutch scientist Willebrord van Roijen Snell (1581–1626), known for his law of refraction in optics.