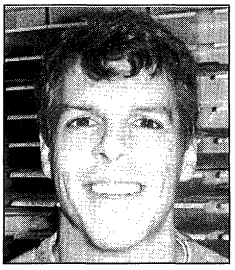
Conformality, the Exponential Function, and World Map Projections

Timothy G. Feeman



Tim Feeman (timothy.feeman@villanova.edu) received his Ph.D. degree in 1984 at the University of Michigan under Allen Shields and has been at Villanova University since 1986. His main research interests are operator theory and the connections between cartography and mathematics. He is active in progressive politics and enjoys running, cycling, and spending time with his family.

From the stereographic projection the ancient Greeks used to chart the heavens to modern circle-packing techniques that can chart the inner reaches of the human brain, conformal maps have played a central role in mathematical analysis and applications for over two millennia. It seems a shame, then, that undergraduate mathematics students rarely learn about them, the more so since a basic knowledge of calculus and some facility with trigonometry are enough to make a significant start. In this paper, we use single-variable calculus to establish the conformality of the stereographic projection and the Mercator world map. Along the way, the complex exponential function arises in a natural way that establishes it as a conformal mapping—without using any complex analysis. Furthermore, the exponential function provides a link between the symmetric structures of the two conformal maps. We conclude by indicating how one can construct conformal maps of surfaces of revolution other than the sphere.

Two conformal maps of the sphere

Throughout, our model for the earth is a sphere of radius 1 parametrized using standard spherical coordinates. That is, every point on the sphere is determined by two angles u and v, called the *latitude* and *longitude*, respectively, as shown in Figure 1. Notice that we must first select two antipodal points to serve as the poles, $u = 0$ and $u = \pi$, and a semi-circle connecting them to serve as the 'prime meridian' $v = 0$.

A map of the sphere, or of any surface for that matter, is said to be *conformal* if the images of any two intersecting paths on the surface intersect at an angle equal to that between the original paths themselves, where the angle between paths is just the angle between their tangent vectors. Thus, conformal maps portray shapes accurately, at least locally. This generally comes at the expense of severe distortions in area.

The *stereographic projection* is a conformal map of the sphere that was known to the ancient Greeks at least by the time of Hipparchus of Rhodes during the second century BC. In 1695, Edmund Halley, motivated by his interest in star charts, published the first proof that this map is conformal. The stereographic projection is constructed by projecting points on the sphere onto a tangent plane from a light source located at the point on the sphere directly opposite the point of tangency. (See Figure 2.) To get a polar perspective, take the plane to be tangent to the sphere at the pole $u = \pi$ and put the light source at the opposite pole $u = 0$. The effect is that the meridian at v projects

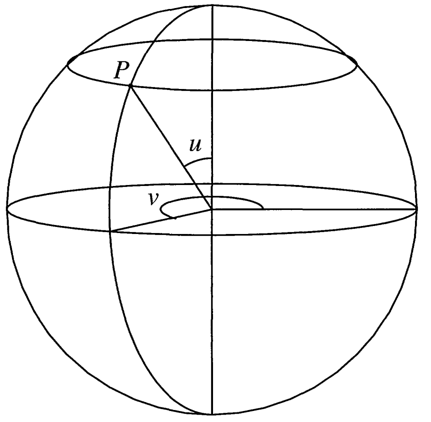


Figure 1. The point P has latitude u and longitude v.

onto the half line with polar coordinate equation $\theta = v$ while the parallel at u projects onto the circle centered at the origin with radius

$$r = r(u) = 2\sin u/(1 - \cos u) = 2\cot(u/2). \tag{1}$$

This can be determined using similar triangles, for example. Notice that the function $r(u)$ is decreasing.

In 1569, Mercator described a technique for creating a conformal world map that shows paths of constant compass bearing as straight lines and, thus, facilitates navigation. As we shall soon see, an exact implementation of Mercator's technique requires the evaluation of a certain integral. Working nearly a century before the introduction of calculus, he was nonetheless able to present a map that approximated the true result. By adjusting his calculations every five degrees of latitude as he worked up from the equator, he arrived at an approximating sum for the integral. His work was refined by

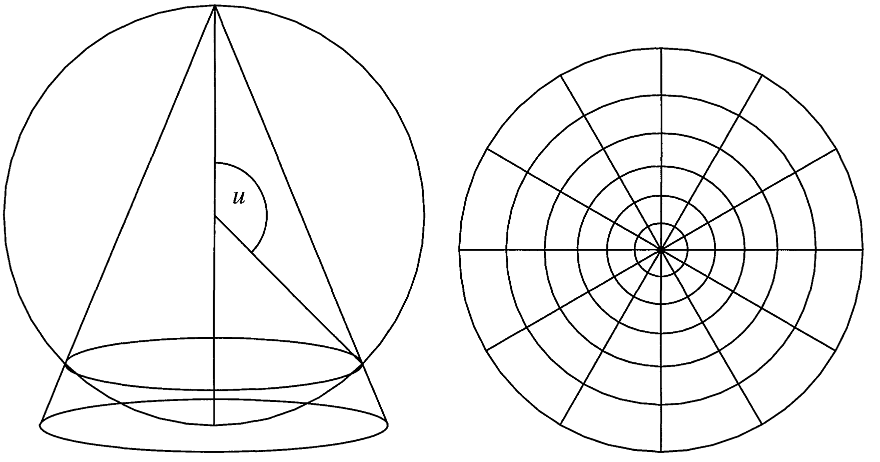


Figure 2. Stereographic projection with graticule for a hemisphere; parallels are shown in increments of $\pi/12$.

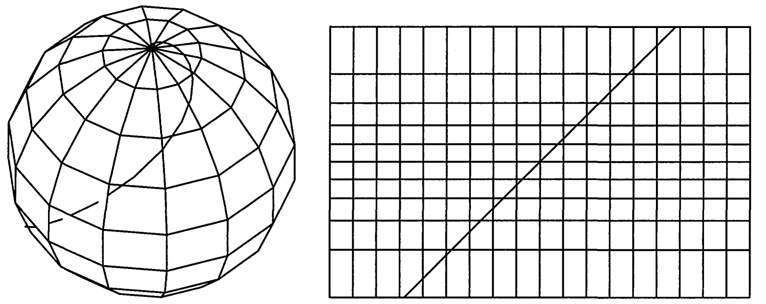


Figure 3. A path of constant bearing on the globe and its image under Mercator's map.

others after the advent of calculus to produce a truly conformal map. Variations of it are used today for satellite tracking and other navigational applications.

The Mercator map portrays all meridians as vertical lines running north-south and all parallels as horizontal line segments running east-west. The difficulty is to determine the necessary spacing between the horizontal lines so that every path of constant compass bearing on the sphere will resolve into a straight line on the map. (See Figure 3.) A navigator using this map can set the ship's bearing simply by drawing a line between the current location and the desired destination. Using Mercator's map in conjunction with a gnomonic projection, which shows shortest routes on the globe as straight lines, a navigator can prepare an approximation of the shortest route that requires relatively few changes in bearing.

To solve the problem of spacing the parallels, let's first assume that the overall horizontal dimension of the map is 2π and that the meridian at longitude v is shown as the vertical line with equation $x = v$, for $0 \leq v \leq 2\pi$. Thus, the map is "true" along the equator of our sphere of radius 1. In this case, as we demonstrate later, conformality is achieved by placing the parallel with latitude u at height

$$y = h(u) := \ln\left[\cot(u/2)\right]. \tag{2}$$

Notice that the function h is decreasing.

Proving conformality

As mentioned above, a map is conformal if the angle between two intersecting paths is always preserved. To check for conformality, then, it would appear to be necessary to check these angles for all possible pairs of paths. However, the two maps being considered here have the important feature that "rectangles" on the sphere, by which we mean regions bounded by two meridians and two parallels, are mapped to either "polar rectangles" or Cartesian rectangles in the plane. Under these circumstances, as Figure 4 indicates, all angles will be preserved, and conformality achieved, if, at each point, the scale factor along the parallel, represented by the horizontal edge in the figure, is equal to the scale factor along the meridian, represented by the vertical edge in the figure.

The scale factor of a map is a ratio of lengths—the length of a path on the map divided by the length of the actual path that it represents. For a map of a curved surface, this measurement is a *local* phenomenon that can vary from point to point and in different directions from the same point. So the *local* scale factor of the map at a given

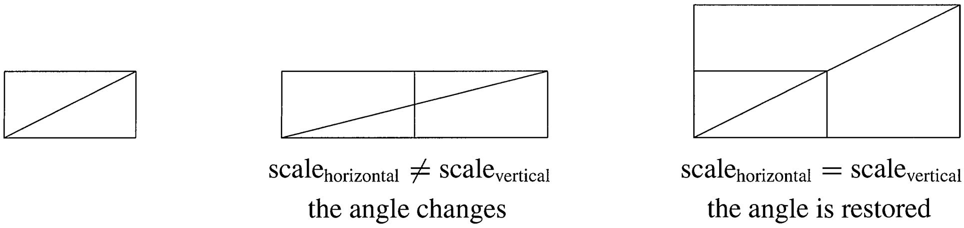


Figure 4. How scale factors affect angles.

point along a given path is the ratio of the arclength elements along the image of the path and along the path itself. For our present purposes, the only paths we need to look at are the parallels and meridians.

To compute the necessary scale factors for the stereographic projection, start at a point $\mathbf{p}$ on the sphere with latitude and longitude given by the pair (u, v) and travel along the parallel to the point $(u, v + dv)$. The distance covered is $\sin(u)\,dv$ since the parallel at u is a circle of radius $\sin(u)$. On the map, the image of this path is an arc of angle dv along a circle with radius $2\cot(u/2)$. The distance covered is thus $2\cot(u/2)\,dv$. The scale factor along this parallel, denoted by M_p, is therefore given by the ratio of these incremental distances. That is,

$$M_p = \frac{2\cot(u/2)\,dv}{\sin(u)\,dv} = \csc^2(u/2).$$

Notice that this is independent of the longitude angle v.

Next, from the point $\mathbf{p} = (u, v)$ on the sphere, travel along the meridian to the point $(u + du, v)$. The distance is du since the meridian has radius 1. On the map, we are moving between circles of radii $r(u)$ and $r(u + du)$, a distance of $-dr$. (Remember that r is decreasing.) The local scale factor along the meridian at the point $\mathbf{p}$, denoted M_m, is the ratio of map distance to distance on the sphere. That is,

$$M_m = -\frac{dr}{du} = -2\frac{d}{du}\left(\cot(u/2)\right) = \csc^2(u/2).$$

Thus, $M_p = M_m$ at every point for the stereographic projection, establishing its conformality.

For the Mercator map, again start at a point $\mathbf{p} = (u, v)$ on the sphere and travel along the parallel to $(u, v + dv)$ or along the meridian to $(u + du, v)$. The corresponding distances are $\sin(u)\,du$ and du, respectively, as before. On the map the corresponding arclengths are dv and $|h(u + du) - h(u)| = -dh$. (Recall that h is decreasing.) The scale factors are, therefore, $M_p = dv/[\sin(u)\,dv] = \csc u$, along the parallel, and $M_m = -dh/du$, along the meridian.

Figure 4 guarantees that conformality will hold provided that $-dh/du = \csc u$. Along with the initial condition $h(\pi/2) = 0$, this yields

$$h(u) = \int_u^{\pi/2} \csc t\,dt = \ln|\csc u + \cot u| = \ln\{\cot(u/2)\},$$

as we stated above.

Both the stereographic projection and Mercator's map are uniquely determined up to translation and global scale change within their respective frameworks.

Taken together, (1) and (2) show that $r(u) = 2e^{h(u)}$. Thus, by transferring the Mercator map image back onto the sphere and, from there, into the plane again using the stereographic projection, we get a mapping from the vertical strip $\{(x, y) : 0 \le x \le 2\pi\}$, occupied by the Mercator map, into the plane that sends the point (x, y) to the point with polar coordinates $(2e^y, x)$. This is the composition of two conformal mappings—the inverse of Mercator's conformal map and the stereographic projection—and, hence, must be conformal as well. Using a half-sized scale model of the stereographic projection shows that mapping (x, y) to the point with polar coordinates (e^y, x) is also conformal. Moreover, if we first reflect the xy-plane across the line $y = x$, then we get a conformal map from the horizontal strip $S = \{(x, y) : 0 \le y \le 2\pi\}$ into the plane that maps the point (x, y) in S to the point with polar coordinates (e^x, y).

Identifying the complex number $z = x + iy$ with the point (x, y) in the plane, we see that this mapping sends a complex number z in the strip S to the complex number whose expression in polar form is $e^x \cdot e^{iy} = e^{x+iy} = e^z$. Thus, the exponential function is a conformal mapping that maps the strip S onto the punctured plane (with the origin removed, since the pole that is depicted as the origin on the stereographic projection is not depicted on Mercator's map). Any rectangle $\{(x, y) : 0 \le y \le 2\pi,\ a \le x \le b\}$ is mapped onto an annulus.

In fact, the exponential function serves as the intermediary not only between these two conformal maps of the sphere but between their symmetry structures as well.

Figure 5 illustrates a classical geometric concept of symmetry in which the two points P and Q are said to be symmetric about the circle indicated. In the picture, P and Q lie on the same radial line emanating from O, the center of the circle. The line PA is perpendicular to OQ and intersects the circle at the same two points as the lines through Q that are tangent to the circle.

The right triangles ΔOAQ and ΔOPA in the figure are similar, so the proportionality of the lengths of corresponding sides yields the equation $|OQ|/|OA| = |OA|/|OP|$, or $|OP| \cdot |OQ| = |OA|^2$. That is, to be symmetric about the circle indicated in the figure, two points P and Q must lie on the same radial line emanating from O and the product of their distances to the center must equal the square of the circle's radius.

This concept arises in classical complex analysis, for instance in the study of fractional linear transformations, which have the property that the image of any circle or straight line (thought of as a circle of infinite radius) is again a circle or straight line. A theorem known as the *symmetry principle* states that, if a fractional linear transformation carries a circle C onto a circle C_1, then it transforms any pair of points symmetric about C into a pair of points symmetric about C_1. (See [**1**, Theorem 3.3.15].)

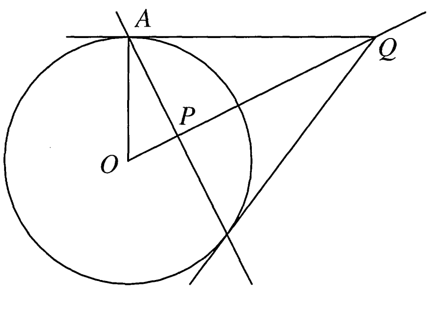


Figure 5. The points P and Q are symmetric about the circle; $|OP| \cdot |OQ| = |OA|^2$.

Returning to our two conformal maps of the sphere, take two points P and Q that lie on the same vertical line on Mercator's map, so P and Q have the same longitude. Let h_1 and h_2 be the y-coordinates of P and Q, respectively. Then certainly P and Q are symmetric about the horizontal line $y = \frac{1}{2}(h_1 + h_2)$ halfway between them. Transferring P and Q back to the sphere using the inverse of Mercator's map and then applying the stereographic projection, we get two points P_1 and Q_1 that lie on the same radial half-line emanating from the origin (so they have the same longitude) and are at distances $r_1 = 2e^{h_1}$ and $r_2 = 2e^{h_2}$, respectively, from the origin. Hence,

$$|OP_1| \cdot |OQ_1| = r_1 \cdot r_2 = 4e^{h_1+h_2} = \left[2e^{\frac{1}{2}(h_1+h_2)}\right]^2. \tag{3}$$

In other words, P_1 and Q_1 are symmetric about the circle of radius $r = 2e^{\frac{1}{2}(h_1+h_2)}$ centered at the origin. This circle on the stereographic projection corresponds to the same parallel as the horizontal line $y = \frac{1}{2}(h_1 + h_2)$ on Mercator's map.

Surfaces of revolution

Though the sphere is of special interest because of its resemblance to the shape of our planet, the methods described above can be applied to more general surfaces of revolution as well. This is so because, for such surfaces, it is possible to define latitude and longitude in a natural way that enables us to define maps that are analogous to our conformal maps of the sphere. The interested reader is encouraged to apply the following general outline to his or her favorite surfaces of revolution.

One way of generating a surface of revolution is to begin with a function f, defined and continuous on some (open or closed, finite or infinite) interval I, such that f is continuously differentiable and satisfies $f(x) > 0$ on the interior of I. A surface is then formed by revolving the curve $y = f(x)$ on the interval I about the x-axis one complete revolution.

Letting v represent the angle of rotation about the x-axis, this surface can be parametrized by $(x, f(x)\cos v, f(x)\sin v)$ for x in I and for $0 \le v \le 2\pi$. Fixing a value of v produces a *meridian* of the surface so that each meridian looks like a copy of the original profile curve $y = f(x)$. Each fixed value of x in I corresponds to a *parallel* of surface. Thus, the parallel at x is a circle of radius $f(x)$ that is perpendicular to every meridian. The arclength elements on the surface are $f(x)\,dv$ and $ds = \sqrt{1 + [f'(x)]^2}\,dx$ along the parallels and meridians, respectively.

A map employing polar coordinates, called an *azimuthal projection* by cartographers, portrays the meridians as "evenly spaced" half lines and the parallel at x as a circle centered at the origin with radius $r = r(x)$, where the non-negative function r is selected by the mapmaker and determines the map's mathematical properties. The scale factors of this map along each parallel and meridian are $M_p = r(x)/f(x)$ and $M_m = \pm dr/ds = \pm dr/[\sqrt{1 + [f'(x)]^2}\,dx]$, where the $\pm$ depends on whether r is increasing or decreasing. The condition for conformality, that $M_p = M_m$ at every point, is a separable differential equation that yields the solution

$$r(x) = r(c)\exp\left\{\pm\int_c^x \frac{\sqrt{1 + (f'(t))^2}}{f(t)}\,dt\right\} \tag{4}$$

for x in the interior of I, where c is a suitably chosen base point in I.

In similar fashion, we can construct a conformal map of the surface that, like Mercator's map of the sphere, is rectangular in appearance (called a *cylindrical projection* by cartographers). If we take the overall horizontal dimension of the map to be 2π and depict the meridians as vertical lines and the parallel corresponding to the real number x as a segment of the horizontal line $y = h(x)$, then we get $M_p = 1/f(x)$ and $M_m = \pm dh/ds = \pm dh/[\sqrt{1+(f'(x))^2}\,dx]$. The condition for conformality, that $M_p = M_m$, is again a separable differential equation, this time with solution

$$h(x) = h(c) + \int_c^x \frac{\sqrt{1+(f'(t))^2}}{f(t)}\,dt \tag{5}$$

for x in the interior of I. Comparing this formula with (4) shows that $r(x) = e^{h(x)}$ once again.

This description applies, for instance, to the surface known as the *catenoid*, generated by the function $f(x) = \alpha\cosh(x/\alpha)$, where $\alpha = f(0)$ is a positive constant. The graph of f is called a *catenary*, from the Latin for "chain". The catenoid is a minimal surface, formed, for instance, by a soap film extended between two circular wire frames in parallel planes. (See Figure 6.) In this case, (4) and (5) yield $r(x) = r(0)\,e^{x/\alpha}$ and $h(x) = h(0) + (x/\alpha)$, respectively. Taking $r(0) = 1$ and $h(0) = 0$ produces the simple relationship $r(x) = e^{h(x)}$.

A second example is the pseudosphere, also known as the bugle surface or tractroid, whose generating curve, called the *tractrix*, starts at the point (0, 1) and moves so that its tangent line always reaches the x-axis after running for distance exactly 1. (See Figure 7.) Analytically, this condition is described by the differential equation $f'(x) = -f(x)/\sqrt{1-[f(x)]^2}$, for $x > 0$, with $f(0) = 1$.

Now, with the radius function $r(x)$ decreasing, $r(0) = 1$ and $h(0) = 0$, (4) and (5) yield $h(x) = \int_0^x \frac{-f'(t)}{(f(t))^2}\,dt = \frac{1}{f(x)} - 1$ and $r(x) = \exp\left\{1 - \frac{1}{f(x)}\right\} = \exp\{-h(x)\}$, for $x > 0$. The azimuthal projection maps the pseudosphere conformally onto the punctured unit disc in the plane.

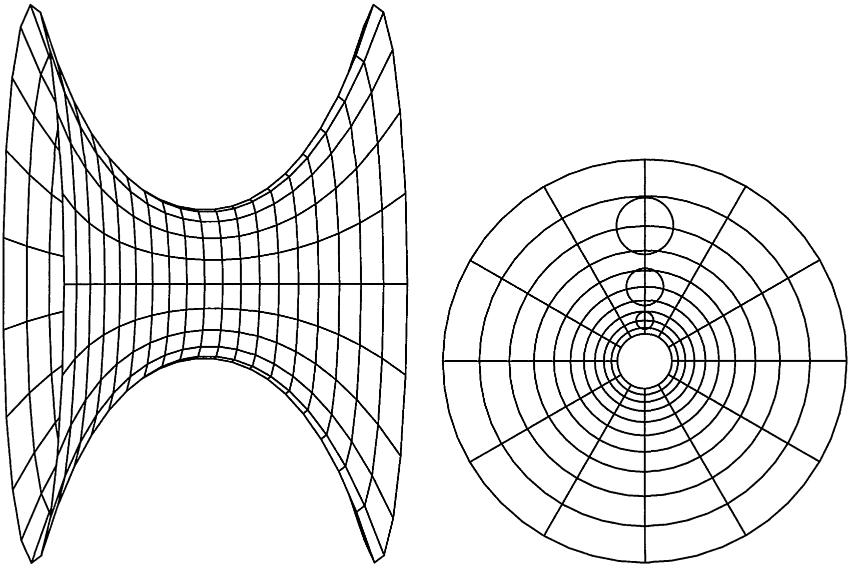


Figure 6. The catenoid and a conformal azimuthal map of the portion with $-1 \le x \le 1$. The small circles show the area distortion of the map.

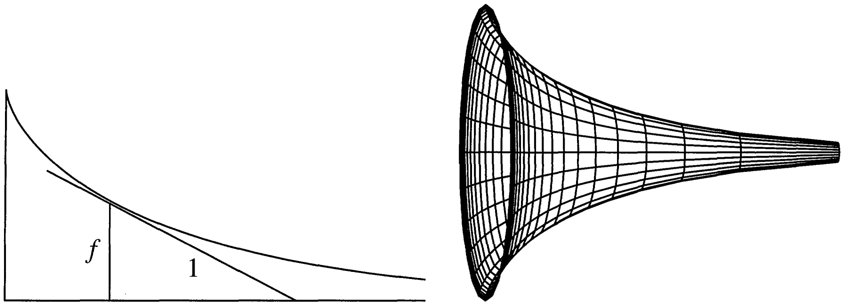


Figure 7. The tractrix, $f' = -f/\sqrt{1-f^2}$; the pseudosphere.

A slightly different approach is needed to handle the torus, which can't be obtained by revolving the graph of a function. Instead, begin with two positive numbers, R_1 and R_2, satisfying $0 < R_1 \leq R_2$. The circle of radius R_1 centered at $(R_2, 0)$ can be parametrized by $(R_2 + R_1\cos(u), R_1\sin(u))$, for $0 \leq u \leq 2\pi$. Revolving this circle about the y-axis generates a torus. Each fixed value of u corresponds to a parallel, a circle of radius $[R_2 + R_1\cos(u)]$. Each meridian is a circle of radius R_1 obtained by intersecting the torus with a plane containing the y-axis.

For a cylindrical projection of the torus, we map the parallel at u to a horizontal line segment of length 2π at height $y = h(u)$, where h is an increasing function satisfying $h(0) = 0$. The images of the meridians are "evenly spaced" vertical lines or line segments. (See Figure 8.) The scale factors are then $M_p = 1/[R_2 + R_1\cos(u)]$ and, since the arclength element on each meridian is $ds = R_1\,du$, $M_m = dh/ds = dh/(R_1\,du)$. For a conformal map, we require that $M_p = M_m$. That is, $h(u) = \int_0^u \frac{R_1}{R_2+R_1\cos(t)}\,dt$. In the case where $R_1 = R_2$, so that the "inner equator" of the torus is just the point at the origin, then we get simply $h(u) = \tan(u/2)$.

Exploiting the role of the exponential function, we will have a conformal azimuthal projection of the torus if we show the meridians as radial half lines emanating from the origin and portray the parallel at u as a circle, centered at the origin, of radius $r(u) = r(0)\cdot\exp\left\{\int_0^u \frac{R_1}{R_2+R_1\cos(t)}\,dt\right\}$.

Remarks

Though we have examined only conformality here, cartographers also have a considerable interest in maps that show areas in their correct proportions. For a map on which the images of the meridians and parallels intersect at right angles, a diagram

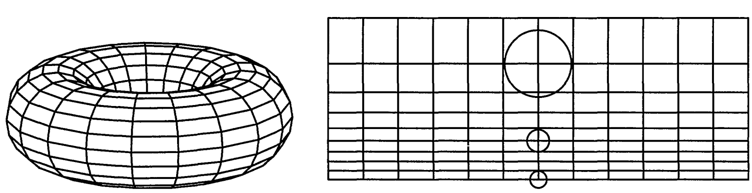


Figure 8. The torus and a conformal cylindrical map of latitudes $0 \leq u \leq 3\pi/4$. The small circles show the area distortion of the map.

similar to Figure 4 shows that the requirement for the map to be area preserving is that $M_p \cdot M_m = 1$. Interested readers are invited to use this to construct area preserving cylindrical and azimuthal maps of their favorite surfaces of revolution.

References

1. Lars V. Ahlfors, *Complex Analysis*, 2nd ed., McGraw-Hill, New York, 1966.
2. Steven G. Krantz, Conformal mappings, *American Scientist*, **87** (1999), 436–445.
3. John P. Snyder, *Flattening the Earth*, University of Chicago Press, 1993.

Pólya Awards, 2000

The recipients of the George Pólya Awards, given for articles of expository excellence published in the *College Mathematics Journal*, are Chip Ross and Jody Sorenson (jointly) and Ezra Brown. Their papers were

"Will the Real Bifurcation Diagram Please Stand Up!" by Chip Ross and Jody Sorenson (January, pp. 2–14).

The bifurcation diagram (also known as the Feigenbaum diagram) for the quadratic map ($x \rightarrow x^2 + c$) is, deservingly, a centerpiece of the introductory theory of dynamical systems. It is interesting to compute, encapsulates a lot of information, and has an intriguing shape. But this classic diagram records only the attracting fixed points. By including the repelling fixed points, one gets a diagram in which the bifurcations are seen to be trifurcations. In this comprehensive and enthusiastic article, Ross and Sorenson present several aspects of this more detailed diagram and its pedagogical relevance. In addition to discussing how they have actively engaged their students in these investigations, they give several examples of how their version of the diagram gives useful information about the quadratic map. A striking example is how the diagram brings out the Cantor set of points x for which $f(x)$ is chaotic. Their work serves as an inspiration in that it shows how an innovative use of computation can shed new light on well-studied examples of mathematical importance.

"Three Fermat Trails to Elliptic Curves" by Ezra Brown (May, pp. 162–172).

The paper quickly seduces readers into following a path, and a story, Ezra Brown clearly wants us to enjoy. And we do. By elegantly intertwining historical nuggets with an intriguing set of illustrative problems, all beautifully and simply explained, Brown makes this introduction to the study of elliptic curves a compelling introduction to the subject.

