The Use of Scale Factors in Map Analysis
An Elementary Approach

ELAINE F. BOSOWSKI AND TIMOTHY G. FEEMAN
Villanova University / Pennsylvania / United States

Abstract Throughout human history there has been a close connection between mathematics and cartography to the benefit of both disciplines. Yet most undergraduate students in both areas are unaware of it. In this paper, itself a cooperative effort between a cartographer and a mathematician, we hope to show that mathematical ideas available to many first-year college students can be used to analyze properties of some commonly used map projections. Specifically, our principal mathematical tool is the Calculus of functions of a single variable. After giving a proof, using only high-school algebra and trigonometry, that there can be no fixed-scale flat map of the earth, we demonstrate how the Calculus can be used to compute the scale factors along meridians and parallels for selected cylindrical and azimuthal projections. Conditions that these scale factors should satisfy for a projection to be either conformal or equal-area are then discussed, and canonical examples of each type are exhibited. Finally, we use scale factors to analyze how each projection distorts areas and angles, and discuss how the classic tool known as Tissot's indicatrix applies to this setting.

Introduction

ONE consequence of changes in cartographic education (as noted particularly in *Cartographica,* 1996) might well be collaborative efforts between geographers and colleagues from other disciplines. Indeed, to the list of sources cited by Keller (1996, 49) as being important areas from which materials for cartography courses might be drawn, "…psychology, cognition, computer sciences, business, language, and art", we would emphatically add *mathematics*. This paper itself is the result of cooperation between a geographer/cartographer and a mathematician interested in blending their approaches to teaching and scientific inquiry.

From diagrams of bus, train, and subway routes and street maps of the towns and cities where we live, to weather charts, shopping-mall directories, graphics accompanying newspaper and magazine articles about global events, and, some would argue (Wood 1992), even traffic signs, we encounter and use all sorts of maps every day of our lives. Maps are tools, products of human effort and creativity, that can be designed for a host of uses. But whatever its purpose, every map employs some type of code to formulate and express its message (MacEachren 1995). For many maps this code begins with a mathematical transformation of the map's subject. This is most readily apparent, perhaps, in maps that portray relatively large portions of the earth's surface, such as those shown in a typical atlas (e.g., *Goode's World Atlas,* 1995), but it is true of many other maps as well.

Despite its presence, the mathematics behind various maps is not often discussed in significant detail in any undergraduate mathematics or geography course.[1] Of course, as both the earth and the flat piece of paper onto which it is to be mapped are multi-dimensional surfaces, a background in multivariable Calculus would seem on first thought to be necessary for understanding the mathematical properties of maps. In fact, this is not the case. As we discuss here, an analysis of a map's scale factors is an important key to understanding certain properties of projections, such as preservation (or distortion) of areas or angles. Because calculating scale factors involves only the comparison of *linear* distances, the basic elements of one-variable Calculus are adequate to the task. Our focus will be on certain specific projections that are commonly used to make atlas maps or are otherwise of historical interest. For the sake of simplicity, we assume throughout that the surface of the earth is a perfect sphere. Though of course this is not actually the case, it is nearly enough so for our purposes.

Elaine F. Bosowski, Dept. of Geography, Villanova University, Villanova, PA USA (See dedication, p. 9)
Timothy G.Feeman, Dept. of Mathematical Sciences, Villanova University, Villanova, PA 19085 USA tfeeman@email.vill.edu
Revised manuscript received April 1998.

1 The possible exception would be as an interesting side topic in a first course in differential geometry taken by advanced mathematics and physics students (McCleary 1994).

There Are No Ideal Maps

Before we can use scale factors to analyze the properties of a given projection, we must examine what we mean by the scale of a map. Many of the maps we encounter every day make no claim of being rendered 'to scale,' but others, such as atlas maps, most road maps, and United States Geological Survey topographic maps, do include some explicit information about the scale of the map. For example, in the legend of an atlas map we might find the representative fraction 1 : 16,000,000, meaning that any path of length 1 unit on the map represents a path of length 16 million units on the surface of the earth. So, as a first attempt at a definition, let us say that a map has a scale factor of M if the ratio of the length of a path on the map to the length of the path that it represents on the earth is M. That is,

$$\text{scalefactor} = M = \frac{\text{distance on map}}{\text{distance on earth}}$$

It turns out, however, that it is dishonest for a map to claim to have a single scale factor, for the reality is that *there can be no flat map of even a portion of the earth's surface that has a fixed scale*. Said differently, every map of even a portion of the earth involves at least some distortion, some compromise in the representation of the earth's features. Standard intuitive arguments (try to flatten an orange peel without stretching or tearing it; try to apply a piece of paper to a globe without wrinkling it) might persuade us of this, but a proof requires only a bit of plane geometry and trigonometry.

Indeed, assume the earth to be spherical and consider the circle C of all points on the earth's surface that lie at a distance of r units away from the north pole, where the distance is measured along the surface of the earth. That is, the distance from any point to the north pole is the length of the appropriate arc of the meridian passing through this point. Within the plane defined by the circle C, however, the radius of C is easily seen to be less than r so that its circumference, which we'll call L, is less than the quantity $2\pi r$. (Figure 1.) If a flat map of the earth had a fixed scale factor of M, then the image of the circle C on the earth would be a circle of radius Mr on the map. The circumference of this image circle would then be $2M\pi r$. But, using the fixed scale, the circumference of the image of C ought to be just M times the circumference of C itself, or $M \cdot L$ which is less than $2M\pi r$. The inequality of these two circumferences for the image of C tells us that such a map cannot exist.

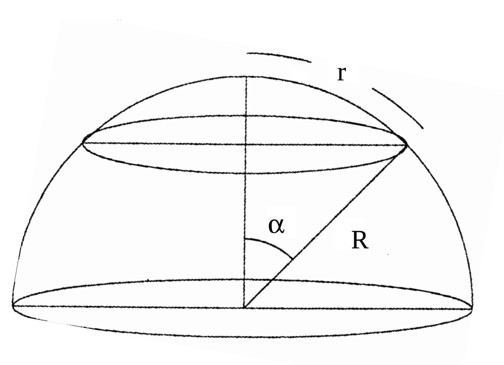

Figure 1 Circumference of the parallel is less than $2\pi r$ on the sphere.

GAUSSIAN CURVATURE

To examine Figure 1 in more detail, observe that the circumference of C is given by $2\pi R \sin \alpha$, where R units is the radius of the earth and α is the central angle formed by the north pole, the centre of the earth, and any point on the circle C. Basic geometric considerations show that $\alpha = r/R$ (in radians) so that the circumference of C is equal to $2\pi R \sin(r/R)$. We can use the Taylor series for the sine function to write the circumference of the circle C on the globe, which we'll denote now by $L(r)$, as

$$L(r) = 2\pi R sin(r/R) = 2\pi R\left(\frac{r}{R} - \frac{r^3}{R^3 \cdot 3!} + \frac{r^5}{R^5 \cdot 5!} - \ldots\right)$$
$$= 2\pi r - \frac{\pi r^3}{3R^2} + \frac{\pi r^5}{60R^4} - \ldots$$

Thus,

$$\lim_{r \to 0} \left(\frac{3}{\pi}\right)\frac{2\pi r - L(r)}{r^3} = \frac{1}{R^2} \qquad (1)$$

This is the numerical quantity which is called the *Gaussian curvature* of the sphere of radius R. A similar procedure can be used in principle to compute Gaussian curvature for other surfaces as well. Choose a base point on the surface and measure the perimeter, $L(r)$, of the 'circle' on the surface consisting of all points lying at a distance r from the base point, where the distance is measured along the surface. Then compute the limit above. The value of the limit is the Gaussian curvature of the surface at the base point. Because the sphere is completely symmetrical, the Gaussian curvature is constant over the entire surface. (See Osserman 1995 for a splendid exposition on the relation between Gaussian curvature and geodesy.)

Computing Scale Factors

We have just seen that the scale factor of a flat map of the earth must vary over the region covered by the map. That is, the scale factor of a projection is a local phenomenon which must change from point to point and can be different in different directions from the same point. Thus, we have no choice but to formulate a notion of 'local' representative fraction, or local scale factor. To do this, consider that the scale factor in any given direction from a given point on a map is *almost constant* if we restrict ourselves to only small distances on the earth it represents. Indeed, this is the reason why many maps, especially those showing fairly small regions, can *claim* to have a constant scale factor and, more important, the reason why they get away with it (Monmonier 1991). This point of view allows us to approximate the local scale of the map thus:

$$M_{local} \approx \frac{\Delta dist_{map}}{\Delta dist_{earth}} \tag{2}$$

where M_{local} is the local scale factor, and the symbol Δdist stands for a small increment of distance on either the map or the earth accordingly. The exact value of M_{local} is obtained by taking the limit of the quotient on the right hand side as the increment $\Delta dist_{earth}$ tends to 0. Students of Calculus will recognize this as the process of taking a derivative. To see how this works, consider two specific classes of map projections. As a general convention, in order to simplify the computations we will work not from the earth directly, but rather from a spherical globe whose radius is taken to be 1 unit.

Maps that depict the entire sphere on a rectangle are called *cylindrical projections*. Typically, such maps show the parallels of latitude as horizontal lines and the meridians as vertical lines equally spaced along the equator. Varying the spacing between the parallels changes the appearance as well as the mathematical properties of the map. In this paper, when dealing with cylindrical projections we will measure latitudes in radians rather than the degrees we see in atlas maps, starting from $-\pi/2$ at the South Pole up to 0 at the equator and $\pi/2$ at the North Pole, because the Calculus is more straighforward that way. The equator will be represented by a horizontal line which we take as the x-axis in the plane of the map. The parallel at latitude u will be shown on the map as a segment of the horizontal line with equation $y = h(u)$, where h is some specific function. (Note that $y = h(0) = 0$ at the equator.) The easiest way to show the meridians as evenly spaced is to map the meridian at longitude v (also in radians) to the vertical line with equation $x = v$. Thus, the overall horizontal dimension of the map will be 2π units, matching the circumference of the equator on our unit globe. Figure 2 shows an example of a cylindrical projection.

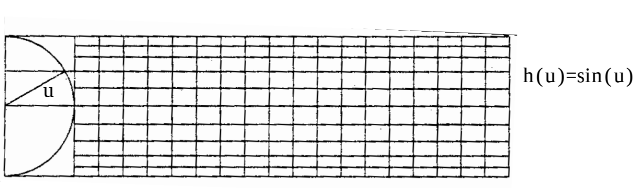

Figure 2 Graticule for Lambert's equal-area cylindrical projection

With the above conventions for a cylindrical projection, the horizontal distance on the map between the vertical lines representing the longitudes v and $(v + t)$ is simply t. At latitude u on the globe, the circumference of the parallel is $2\pi \cos u$, so the portion of the parallel that lies between longitudes v and $(v + t)$ is a circular arc of length $(t/2\pi)(2\pi)\cos u = t \cos u$. Thus, the scale factor along this parallel, denoted by M_p, is equal to

$$M_p = \lim_{t \to 0} \frac{\Delta dist_{map}}{\Delta dist_{globe}} = \lim_{t \to 0} \frac{t}{t \cos u} = \sec u . \tag{3}$$

The second type of map we consider here is an *azimuthal projection* in which the globe, or some portion of it, is projected onto a plane tangent to the globe at a selected point. The *stereographic, gnomonic,* and *orthographic projections* are primary examples of azimuthal maps. These are projections in the truest sense, with the projecting light sources located at the antipode to the tangent point, at the centre of the globe, and at infinity, respectively. The resulting gnomonic and orthographic projections provide maps of at most one hemisphere. See Figures 3, 4, and 5 for illustrations of these three classical azimuthal projections. The *azimuthal equidistant projection*, in which all distances to the central point on the map are shown to a common scale, is also included in this class, although it is not obtained by projection from a light source.

Our work here, which is to compute scale factors, will be most straightforward if we adopt the following conventions for handling azimuthal projections. We will assume, first, that the projection is centred on the South Pole, and, second, that latitude angles are measured upward from 0 at the South Pole to $\pi/2$ at the equator and π at the North Pole. The image of the South Pole will be taken as the origin in the plane of the map. The parallels will be shown on the map as concentric circles centred at the origin, with the circle corresponding to latitude u having a radius of $r(u)$. Our convention of measuring latitude angles upward from the South Pole ensures that the function $r(u)$ increases with u. The meridian at longitude v will be portrayed as a radial line segment emanating from the origin and making an angle of v with the positive x-axis.

In this context, the parallel at latitude u on the globe has radius sin u. Its image is a circle of radius $r = r(u)$ on the map. Therefore, regardless of the longitude, a change of t in the longitude angle will correspond to a distance of $(t/2\pi)(2\pi r(u)) = t\,r(u)$ along the image of the parallel, compared to a distance of $(t/2\pi)(2\pi \sin u) = t \sin u$ along the parallel itself on the globe. Thus, the scale factor along this parallel will be given by

$$M_p = \lim_{t \to 0} \frac{t\,r(u)}{t \sin u} = r(u) \cdot \csc u \tag{4}$$

That no limits were really needed to compute these particular scale factors was due to the fact that the meridians on these maps are evenly spaced along each parallel, just as they are on the globe. Thus, each incremental segment of the parallel is stretched by the same factor. In other words, the spacing of the meridians along a parallel determines the scale factor along that parallel. Similarly, the scale factor along a meridian depends on the spacing of the parallels.

For a cylindrical projection, recall that we represent the equator as lying along the x-axis and the parallel at latitude u as the horizontal line at height $y = h(u)$. The arc of the meridian on the globe lying between latitudes u and $(u + t)$ has length t, while its image on the map has length $h(u + t) - h(u)$. Thus, the scale factor, which we will denote by M_m, along the meridian at a point at latitude u is given by

$$M_m = \lim_{t \to 0} \frac{h(u+t) - h(u)}{t} = h'(u), \tag{5}$$

the derivative of the height function for the parallels.

Similarly, for an azimuthal projection centred on the South Pole, recall that $r(u)$ is the radius of the circle which is the image of the parallel at latitude u. The arc of the meridian on the globe between latitudes u and $(u + t)$ has length t, while the image of this segment on the map has length $r(u + t) - r(u)$. Thus,

$$M_m = \lim_{t \to 0} \frac{r(u+t) - r(u)}{t} = r'(u). \tag{6}$$

For example, Lambert's equal-area cylindrical projection (Figure 2), introduced in the 1770s by Johann Heinrich Lambert, is a rectangular map that places the parallel at latitude u at the height $h(u) = \sin u$. Thus, by Equation 5, the scale factor M_m along any meridian at a point at latitude u is equal to $M_m = h'(u) = \cos u$. By Equation 3, all cylindrical projections have $M_p = \sec u$ at latitude u.

For the stereographic projection (Figure 3), the vertical distance between the North Pole, where the light source is located, and the (horizontal) plane of the parallel at latitude u (measured up from the South Pole) is $1 + \cos u$. Thus, by similar triangles, the image of this latitude circle is a circle on the paper with radius $r = 2 \sin u/(1 + \cos u) = 2 \tan(u/2)$. From Equation 4, the scale factor along the parallel at latitude u is $M_p = 2/(1 + \cos u) = \sec^2(u/2)$. By Equation 6, we get $M_m = dr/du = 2/(1 + \cos u) = \sec^2(u/2)$ for the scale factor along any meridian at a point at latitude u.

For a gnomonic map (Figure 4) of the southern hemisphere centred at the South Pole, similar considerations of plane geometry and trigonometry show that the parallel at latitude u is represented on the map by a circle of radius $r(u) = \tan u$. As the tangent function is undefined for the angle $\pi/2$, we see that the equator is mapped to infinity. The scale factors for this projection are, therefore, $M_p = \tan u \cdot \csc u = \sec u$ and $M_m = \sec^2 u$.

An orthographic projection (Figure 5) of the southern hemisphere onto a plane tangent to the globe at the South Pole will project the parallel at latitude u onto a circle of radius $r = \sin u$ centred at the Pole. The scale factor, M_p, of the map along this parallel is therefore equal to 1; that is, the parallel on the globe of radius 1 has the same length as its image on the map. Because the light source for the orthographic projection is located at infinity, this map corresponds somewhat to an astronaut's view of the earth from space, though the astronaut is probably still too close. This property of 'looking right' is reflected in the fact that concentric circles around the centre of the map are true to scale. By Equation 6, we have $M_m = \cos u$ for this projection.

Finally, an azimuthal equidistant projection centred on the South Pole (Figure 6) is designed to show all distances to the Pole correctly. Thus, the parallel at angle u is shown as a circle of radius $r(u) = u$, this being the distance along the surface of the globe from the South Pole to any point on that parallel. The scale factor of the map along any meridian is $M_m = r'(u) = 1$, and, along this parallel, we have $M_p = u \cdot \csc u$. Note that the limit of M_p as u approaches 0 is equal to 1; that is, the scale of the map is true at the South Pole.

In some sense, knowing the scale factors of a given projection tells us everything there is to know about its

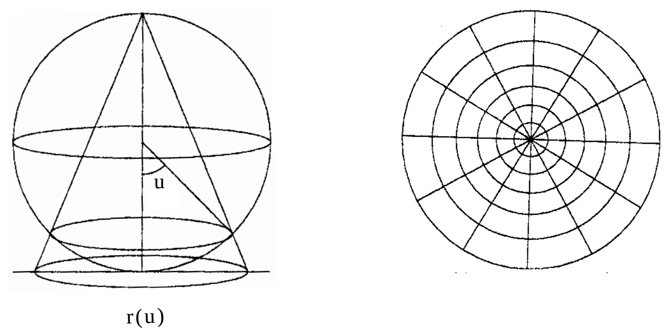

Figure 3 *Stereographic projection with graticule for a hemisphere; parallels are shown in 15° increments from the Pole.*

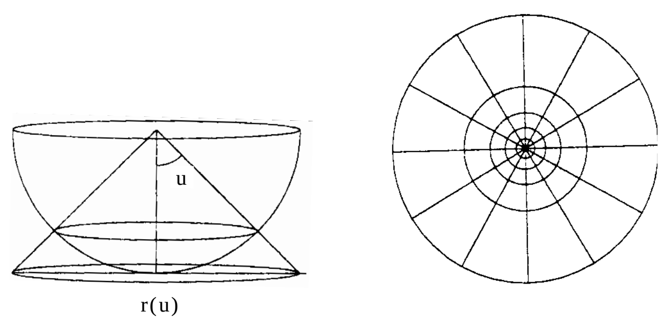

Figure 4 *Gnomonic projection with partial graticule; parallels are shown in 15° increments from the Pole.*

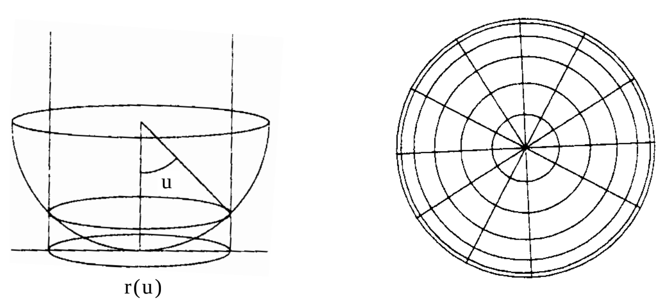

Figure 5 *Orthographic projection with graticule for a hemisphere*

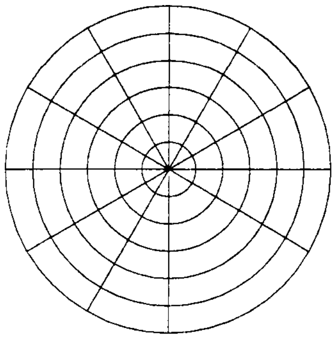

Figure 6 *Graticule for one hemisphere for an azimuthal equidistant projection; parallels are shown in 15° increments from the Pole.*

mathematical properties. Technically, though, this would require computing the scale not only along the meridians and parallels but along any path on the surface of the globe, a daunting task. However, with only one extra condition — *that the projection show the parallels and meridians intersecting at right angles on the map as they do on the globe* — the scale factors along the meridians and parallels are in fact enough to tell us about certain map properties of great importance to map users, including conformality and preservation of areas.

Conformal and Equal-area Maps

CONFORMAL MAPS

A map projection of the globe is *conformal* if the images of any two intersecting paths on the globe intersect at an angle equal to that between the original paths themselves. In particular, because the meridians and parallels on a sphere intersect each other at right angles, it follows that, for a conformal map, the images of the meridians must intersect the images of the parallels at right angles, though this condition alone is not sufficient for conformality.

To see how angles are affected by a map's scale factors, consider a rectangle whose diagonal makes an angle of α with the vertical side, as in Figure 7. If we multiply the base and height of the rectangle by the same factor M, then the angle α is preserved because the triangles in the first and third pictures are similar to each other.

In other words, to preserve the angle, the vertical scale factor must match the horizontal scale factor. If, for a given projection, the images of the parallels and meridians intersect at right angles as they do on the sphere and, if the scale factor M_p along the parallel is equal to the scale factor M_m along the meridian at every point, then every small rectangle on the sphere, represented by the left-most rectangle in Figure 7, will have as its image a rectangle that preserves the angle of the diagonal, represented by the right-most rectangle in Figure 7. Because every angle is the difference between the diagonals of some pair of small rectangles on the globe, it follows that all angles will be preserved. The projection is conformal. Conversely, a conformal projection must show the parallels at right angles to the meridians and, as the diagram indicates, must satisfy the condition $M_p = M_m$ at every point.

Among the maps whose scale factors we have already computed, the stereographic projection is now seen to be a conformal map. (Recall that $M_p = M_m = \sec^2(u/2)$ for this projection.) Indeed, within the class of azimuthal maps, the stereographic is the only conformal projection. To see this, suppose we have an azimuthal map, centred on the South Pole, for which the parallel at latitude u is shown on the map as a circle of radius $r = r(u)$. As we saw above, $M_p = r \csc u$ and $M_m = dr/du$ in this case. The condition for conformality, that M_p and M_m be equal, thus translates into the separable differential equation $dr/du = r \csc u$, or $dr/r = \csc u \, du$. To solve this, we integrate both sides to get $\ln(r) = -\ln(|\csc u + \cot u|) + C$. Exponentiating both sides then yields

$$r = \frac{k}{|\csc u + \cot u|} = \frac{k \sin u}{1 + \cos u} = k \tan\left(\frac{u}{2}\right)$$

for some positive constant k. (The absolute value signs are not needed since $0 \le u \le \pi$ in this context.) In other words, this map is a rescaling by the factor $k/2$ of the stereographic projection.

In 1569, the Flemish geographer and mathematician Gerhard Kremer, better known by his latinized name Mercator, presented a conformal rectangular map. Mercator's aim was to provide a map that would show paths of constant compass bearing on the earth, known as *loxodromes* or *rhumb lines*, as straight lines, and thereby greatly facilitate navigation by means of a compass.

When following a path along the surface of the earth, one's compass bearing at any given point on the

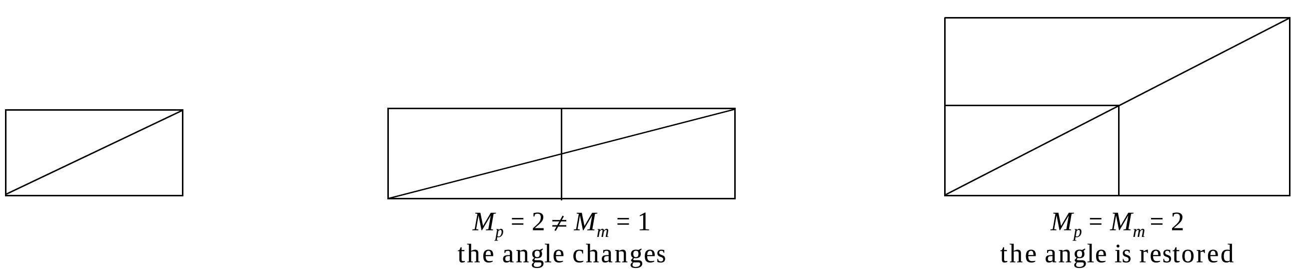

Figure 7 *Angles are affected by scale changes.*

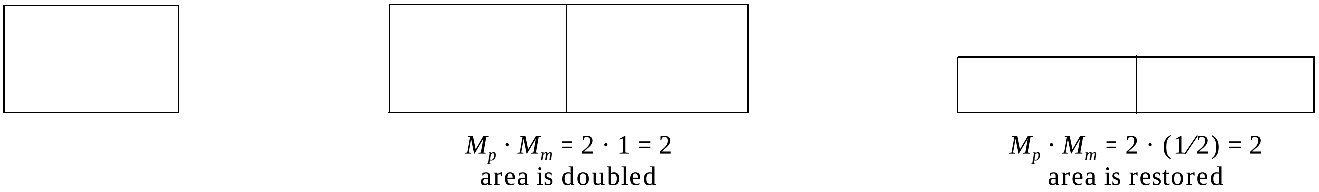

Figure 8 Areas are affected by scale changes.

path is represented by the angle between the direction of the path and the meridian through that particular point. A path of constant compass bearing is therefore one which makes the same angle with every meridian it crosses. On a conformal map, this angle is preserved so that the angle between the image of a loxodrome and the image of every meridian it crosses will correspond to the compass bearing of the loxodrome. On a rectangular map, the meridians are all shown as vertical lines. Therefore a path on the map that makes the same angle with the image of every meridian must be a straight line. In other words, on a conformal rectangular map, loxodromes will be shown as straight lines and the angle between the image of the loxodrome and the vertical will be the compass bearing.

We have already seen that the scale factor along the parallel at latitude u for a rectangular map is given by $M_p = \sec u$. From the condition of conformality, $M_m = M_p$, it follows that we must also have $M_m = \sec u$. To determine the height $h(u)$ at which to place the horizontal line corresponding to the parallel at latitude u, recall that $M_m = h'(u)$. So the function $h(u)$ must must satisfy $h'(u) = \sec u$. Together with the condition that $h(0) = 0$, this implies that

$$h(u) = \int_0^u \sec t \, dt = \ln |\sec u + \tan u|.$$

It is remarkable that Mercator was able to present his map nearly one hundred years before the invention of Calculus, and, indeed, the details of how he constructed it are subject to some speculation. Most likely, he would have seen that, in order to properly place the parallels, he had to proceed in incremental steps, adjusting the (approximate) vertical scale as he went and adding the incremental height displacements that he computed, which amounts to using an approximating sum for the integral.

Incidentally, the stereographic projection and the Mercator map provide a nice illustration of the ‘inverse operation’ relationship between derivatives and integrals. With the stereographic projection, the placement function $r(u)$ is determined by the way the projection is defined and we compute the derivative $r'(u)$ to determine the scale factor along a meridian. That this derivative is equal to the scale factor along the parallel tells us that the map is conformal. For Mercator's map, the problem is reversed. We *start* by wanting to create a conformal map. In other words, we start off with a set of scale factors determined by the conformality condition. We then integrate the scale factor function to determine the placement function $h(u)$ for the parallels.

EQUAL-AREA MAPS

A map projection that shows every region of the globe as having its true area (or its true proportional area, really) is said to be an *equal-area* or *equivalent* projection. To determine the effect of scale factors on areas, consider that if we take a rectangle having adjacent sides of lengths a and b and rescale it, multiplying by M_p in the horizontal direction and M_m in the vertical direction, then the new rectangle will have adjacent sides of lengths $M_p\, a$ and $M_m\, b$. The area will have been changed by a factor of $M_p \cdot M_m$.

This works in much the same way for a flat map of the sphere, provided that the map shows the parallels and meridians intersecting at right angles just as they do on the globe. This requirement, which applies to all the cylindrical and azimuthal projections we have considered here, ensures that the image of any incremental rectangle on the globe is an incremental rectangle on the map. If, at a given point on the sphere, the map projection has scale factors of M_p along the parallel through the point and M_m along the meridian through the point, then the map projection will effectively multiply areas by a factor of $M_p \cdot M_m$, at least locally at that point. Thus, in the special case in which the images of the parallels and meridians intersect at right angles, the condition for a map projection to show all areas to their true proportional sizes is that the product $M_p \cdot M_m$ be a constant.

For example, Lambert's equal area cylindrical projection, for which $M_p = \sec u$ and $M_m = d(\sin u)/du = \cos u$ so that $M_p \cdot M_m = 1$, preserves areas, as one might have guessed from its name. In fact, as $M_p = \sec u$ for all cylindrical projections, it follows that Lambert's is really the only equivalent projection of this type.

Among the azimuthal maps considered so far, none preserves areas. But we can construct one that does. As before, the map is assumed to be centred on the South Pole and we measure latitude angles upward from it. As we have seen, if $r = r(u)$ denotes the radius of the image of the parallel at latitude u, then the scale factor along this parallel is given by $M_p = r \csc u$, while along any meridian we have $M_m = dr/du$ at a point at latitude u. The condition for the map to be truly area preserving, then, is that $M_p \cdot M_m = 1$. That is, $r \csc(u)\ (dr/du) = 1$. This separable differential equation can be rewritten as $r\, dr = \sin u\, du$. Integrating both sides yields the equation $r^2/2 = -\cos u + C$. Substituting the value $r(0) = 0$ gives $r^2/2 = 1 - \cos u$. We know that $r \geq 0$, so $r = \sqrt{2} \cdot \sqrt{1 - \cos u} = 2\sin(u/2)$. As any map that shows areas in their correct proportions must be truly area-preserving, relative to a globe of some radius, it follows that the projection just constructed is the only azimuthal equal-area projection. In fact, this projection, first presented by Johann Lambert in 1772 and known as *Lambert's azimuthal equal-area projection*, is widely used for atlas maps today (see, for example, *Goode's World Atlas* 1995).

Distortion of Areas and Angles

For a cylindrical or azimuthal projection of a sphere to be both conformal and area preserving, both of the conditions $M_m = M_p$ and $M_m \cdot M_p = 1$ would have to be satisfied at every point. Clearly, this implies that $M_m = M_p = 1$ at every point; that is, the scale of the projection would have to be true at every point. But our analysis in Section 2 of the way in which the earth's surface is curved led us to conclude that no flat map of even a small portion of the earth's surface can have a single fixed scale factor. Thus, every flat map of a sphere must involve some distortions of either angles or areas or both. In order to best understand the strengths and weaknesses of any particular map projection, and to best be able to compare projections in order to select one over another as more appropriate for a certain use, we must tackle the essentially mathematical task of measuring and analyzing distortion.

A standard tool for quantifying distortions in angles and areas is *Tissot's indicatrix*, developed in the nineteenth century by the French mathematician, Tissot (Snyder 1993). The starting point for this technique is Tissot's observation that, for any map projection, there is at each point on the sphere a pair of perpendicular directions whose images in the projection are also perpendicular. Tissot called these the principal directions at the given point, and measured distortion using the scale factors of the projection along these directions. There is, for example, some discussion of Tissot's indicatrix in Robinson, et al (1995), but there is no attempt to apply it to any real map projections. In fact, for the cylindrical and azimuthal projections under consideration in this paper, the principal directions are just the parallels and the meridians. Thus, we have already computed the scale factors needed to implement Tissot's programme.

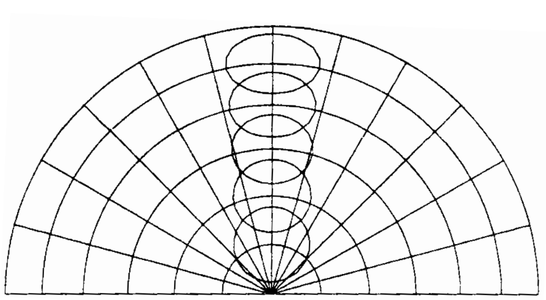

Figure 9 Tissot's indicatrix for Lambert's equal-area azimuthal projection. All ellipses have the same area, but are more elongated away from the Pole.

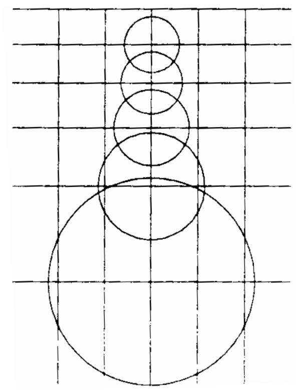

Figure 10 Tissot's indicatrix for Mercator's projection. All ellipses are circles, but the areas increase away from the equator.

To visualize Tissot's indicatrix, we draw at each point on the map an ellipse whose principal axes have lengths $2 \cdot M_p$ and $2 \cdot M_m$. As the area of such an ellipse is just $\pi M_p M_m$, the equal-area condition $M_p M_m = 1$ implies that all the ellipses have the same area on an equal-area map. More generally, the variation in the areas of the ellipses over the range of any given map is an indicator of the area distortions inherent in the map. Similarly, for a conformal map, the condition $M_p = M_m$ implies that all of the ellipses are circles. In general, a more elongated ellipse indicates a greater distortion of angles at that point on a given map. Some examples are illustrated in Figures 9 and 10. In the sections below, we discuss a relatively simple scheme for precisely quantifying these area and angle distortions.

AREA DISTORTION

For a projection in which the images of the parallels and meridians intersect at right angles, we have seen that the image of a small rectangle of size $a \times b$ on the sphere will be a small rectangle on the map of size $(M_p\, a) \times (M_m\, b)$. Thus, the area has been multiplied by the factor $M_p \cdot M_m$ in going from the sphere to the map and we can, therefore, take this factor as a reasonable measurement of the distortion of areas. That is, we define

$$\text{area distortion factor} = M_p \cdot M_m \qquad (7)$$

This quantity will vary from point to point on the map as the scale factors themselves vary. Tables 1 and 2 give some relevant area-distortion factors for the projections we have discussed. In Tissot's system of ellipses, the factor $M_p M_m$ represents the ratio of the area of the ellipse to the area of a circle of radius 1. The more $M_p\, M_m$ varies from the value 1, the more the map distorts areas.

For the Mercator map, for instance, we see that regions at latitudes around $\pi/4$ (45°) will appear on the map to be twice as large as they actually are relative to

Projection	Area distortion factors			
	$M_p \cdot M_m$	at $u = \pi/6$	at $u = \pi/4$	at $u = \pi/2$
Stereographic	$\sec^4 (u/2)$	1.15	1.37	4
gnomonic	$\sec^3 u$	1.54	2.83	∞
orthographic	$\cos u$	.866	.707	0
equidistant	$u \csc u$	1.047	1.11	1.57
Lambert's equal-area	1	1	1	1

Table 1 Area distortion for azimuthal projections

Projection	Area distortion factors			
	$M_p \cdot M_m$	at $u = 0$	at $u = \pm\pi/4$	at $u = \pm\pi/3$
Mercator	$\sec^2 u$	1	2	4
Lambert's equal-area	1	1	1	1

Table 2 Area distortion for cylindrical projections

regions of the same area near the equator, but only one-half of their size relative to regions of the same area at latitude $\pi/3$ (60°). Also, if keeping area distortions under control is a more important consideration than conformality, these measurements suggest that the azimuthal equidistant projection might be a better choice than the stereographic projection or the Mercator map.

ANGLE DISTORTION

To determine a quantitative method for measuring how much a given projection distorts angles, consider that — for a conformal projection, which preserves all angles — the equation $M_p = M_m$ holds at every point. This suggests that, for projections that show the parallels and meridians intersecting at right angles, a reasonable way of measuring the angle distortion is to use the ratio M_p/M_m. For such projections, conformality holds if this ratio is equal to 1 at every point. For Tissot's ellipses, the number M_p/M_m is the ratio of the lengths of the principal axes of each ellipse. The more elongated the ellipse, the more this ratio varies from the value 1.

In fact, for any angle between a meridian and some other great circle arc on the globe, the ratio M_p/M_m measures the ratio between the *tangents* of the angle and of its image angle on the map. This is illustrated in Figure 11.

In the first part of the figure, the angle θ represents an angle on the globe between a meridian and some other great-circle arc. The ratio of the horizontal side of the triangle to the vertical side is $\tan\theta = b/a$. In the second part of the figure, the angle $\hat{\theta}$ is the projected image on the map of the original angle θ. The ratio of the side lengths of this triangle is

Projection	Angle distortion factors			
	M_p/M_m	at $u = \pi/6$	at $u = \pi/4$	at $u = \pi/2$
stereographic	1	1	1	1
gnomonic	$\cos u$	.866	.707	0
orthographic	$\sec u$	1.115	1.414	∞
equidistant	$u \csc u$	1.047	1.11	1.57
Lambert's equal-area	$\sec^2 (u/2)$	1.072	1.172	2

Table 3 Angle distortion for azimuthal projections

Projection	Angle distortion factors			
	M_p/M_m	at $u = 0$	at $u = \pm\pi/4$	at $u = \pm\pi/3$
Mercator	1	1	1	1
Lambert's equal-area	$\sec^2 u$	1	2	4

Table 4 Angle distortion for cylindrical projections

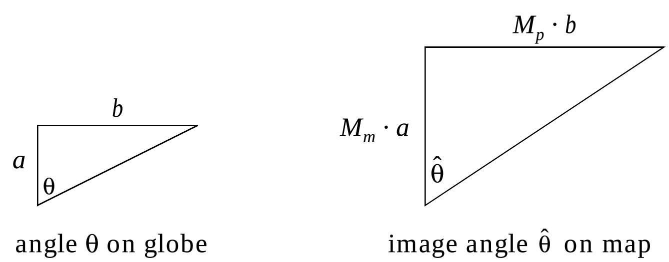

Figure 11 $\tan\theta = b/a$; $\tan\hat{\theta} = (M_p b)/(M_m a)$; $M_p/M_m = \tan\hat{\theta}/\tan\theta$

$$\tan\hat{\theta} = \frac{M_p b}{M_m a} = \frac{M_p}{M_m}\tan\theta; \qquad (8)$$

that is, $\tan\hat{\theta}/\tan\theta = M_p/M_m$. The map has to show parallels and meridians intersecting at right angles for the second figure in the diagram to be a right triangle.

Tables 3 and 4 list some calculations of angle distortions that might be useful to us in selecting a projection. For instance, the function $\sec^2 (u/2)$ grows considerably more slowly than the function $\sec^2 u$, making Lambert's azimuthal equal-area projection significantly more useful than his equal-area cylindrical projection. The azimuthal equidistant projection has an angle-distortion factor that grows even more slowly, but lacks the desirable feature of being area-preserving.

Tissot also used the principal scale factors to measure, at each point on the projection, the *maximum error* between an angle and its image angle. To see how to do this, return to Equation 8 above. Solving for the image angle, $\hat{\theta}$, we get $\hat{\theta} = \arctan((M_p/M_m)\tan\theta)$. If $M_p = M_m$, this implies that $\hat{\theta} = \theta$, as it should for a conformal map. If $M_p \neq M_m$, then we find the maximum error between $\hat{\theta}$ and θ by taking the derivative, with respect to

Projection	Maximum angle error, 2ω		
	at $u = \pi/6$	at $u = \pi/4$	at $u = \pi/2$
gnomonic	-8.24°	-19.76°	-180°
orthographic	8.24°	19.76°	180°
equidistant	2.62°	6.0°	25.66°
Lambert's azimuthal	3.98°	9.06°	38.94°
Lambert's cylindrical	16.42°	38.94°	—

Table 5 Maximum angle error for some non-conformal projections

θ, of the error function $\hat{\theta} - \theta = \arctan((M_p/M_m)\tan\theta) - \theta$, setting the result equal to 0, and solving for θ. This shows that the maximum error occurs when $\theta = (1/2)\arccos((M_p - M_m)/(M_p + M_m))$. This is for angles measured from a meridian. Because an arbitrary angle is the difference between two such angles, we have to double the error computed this way to get the overall maximum angle error, denoted 2ω by Tissot. For the projections we have considered, the scale factors depend on the latitude, so the maximum angle error at any given point will depend only on the latitude of the point. Table 5 gives a comparison of the maximum angle errors encountered at selected latitudes for various projections. As with the angle-distortion factor computed above, we see that, among the non-conformal projections considered, the azimuthal equidistant projection performs relatively well, as does Lambert's equal area azimuthal projection.

OTHER PROJECTIONS AND VARIATIONS

It has been our objective here merely to give an introduction to a few fundamental mathematical aspects of cartography that we believe require only a knowledge of basic trigonometry and geometry and one-variable Calculus to be understood and appreciated. Thus, we have focussed on the properties of conformality and area preservation, and have considered only polar perspective azimuthal projections and cylindrical projections having the equator as the standard line.

The condition we imposed, that the images of the parallels be at right angles to those of the meridians, applies also to conic projections having a polar perspective. When the cone is slit open and laid flat, the angle at the vertex depends on where the cone intersected the sphere before being opened. This additional ingredient in the analysis dissuaded us from discussing these projections in more detail here.

Our analysis above can also be applied to cylindrical projections that have two standard lines. The Gall-Peters equal-area map, to cite a well known example, essentially modifies Lambert's equal-area cylindrical projection to have standard lines at latitudes $\pi/4$ (45°) north and south. Working from a globe of radius 1 unit, the appropriate height function for the parallels is $h(u) = \sqrt{2}\cdot\sin u$. Thus, $M_p = (1/\sqrt{2})\cdot\sec u$, $M_m = h'(u) = \sqrt{2}\cdot\cos u$, $M_p \cdot M_m = 1$, and $M_p/M_m = (1/2)\sec^2 u$.

Quite a few atlas maps are constructed by modifying one of the projections considered here to have a different perspective. For instance, Lambert's azimuthal equal-area projection might be reconfigured to have its centre point on the equator, in order to make an equal-area map of Africa. In such cases, the traditional parallels and meridians generally do not intersect at right angles on the map. Instead, the projection is constructed using substitute parallels and meridians relative to the new perspective. For an azimuthal projection, for instance, the traditional meridians would be replaced in the calculations by the great circles passing through the new central point, while circles concentric to the new centre would substitute for the usual parallels. Scale factors would be computed relative to this surrogate framework. The area- and angle-distortion factors computed here would now measure distortions as one moved away from the map's centre or standard lines.

Atlases also use a variety of projections that are not cylindrical, azimuthal or conic. Analyzing the mathematical properties of such maps — while possible, of course — generally falls outside the framework we have described here and, in any case, would take us well beyond our purpose of highlighting certain basic applications of mathematics to cartography for teachers and students of geography and Calculus.

Acknowledgements

This research was partially supported by the grant NSF-DUE-95-52464

This work is dedicated by Timothy Feeman to the memory of Elaine Bosowski, an inspiring colleague and treasured friend. She was stricken by cancer in the prime of her life. The joy and love she brought to everyone she knew and everything she did remains vibrant and alive.

References

BUGAYEVSKIY, LEV M., and JOHN P. SNYDER. 1995. *Map Projections: A Reference Manual*. London: Taylor and Francis.

Cartographica. 1996. 33/3.

COTTER, CHARLES H. 1966. *The Astronomical and Mathematical Foundations of Geography*. New York: American Elsevier.

DEETZ, CHARLES H., and OSCAR S. ADAMS. 1969. *Elements of Map Projection*. New York: Greenwood Press.

ESPENSHADE, EDWARD B., Jr., et al., eds. 1995. *Goode's World Atlas*, 19th edition. New York: Rand McNally.

KELLER, C. PETER. 1996. "Towards an Introductory Cartographic Curriculum for the 21st Century." *Cartographica* 33/3, 45–53.

MacEACHREN, ALAN M. 1995. *How Maps Work: Representation, Visualization, and Design*. New York: Guilford Press.

McCLEARY, JOHN. 1994. *Geometry From a Differentiable Viewpoint*. Cambridge: Cambridge University Press.

MONMONIER, MARK. 1995. *Drawing the Line: Tales of Maps and*

Cartocontroversy. New York: Henry Holt.

MONMONIER, MARK. 1991. *How To Lie With Maps*. Chicago: University of Chicago Press.

OSSERMAN, ROBERT. 1995. *Poetry of the Universe*. New York: Anchor Books.

PEARSON, FREDERICK, II. 1990. *Map Projections: Theory and Applications*. Boca Raton: CRC Press.

ROBINSON, ARTHUR H., et al. 1995. *Elements of Cartography*. New York: John Wiley.

SNYDER, John P. 1993. *Flattening the Earth: Two Thousand Years of Map Projections*. Chicago: University of Chicago Press.

WOOD, DENIS. 1992. *The Power of Maps*. New York: Guilford Press.

L'utilisation de facteurs d'échelle dans l'analyse cartographique : une approche élémentaire Résumé Tout au long de l'histoire humaine, il y a eu une collaboration et une interaction constantes entre les mathématiques et la cartographie ce, pour le bénéfice des deux disciplines. Toutefois, la plupart des étudiants de premier cycle dans chacune des deux disciplines ne sont pas au courant de cette réalité. Dans cet article, résultat d'une coopération entre un cartographe et un mathématicien, nous espérons montrer que les idées mathématiques accessibles à plusieurs étudiants qui commencent leur université peuvent être utilisées pour analyser les propriétés de certaines projections cartographiques utilisées couramment. De façon spécifique, notre principal outil mathématique est le calcul différentiel de fonctions à variable unique. D'abord, en utilisant la trigonométrie et l'algèbre du niveau secondaire, nous donnons la preuve qu'il ne peut y avoir de carte de la terre à échelle fixe Ensuite, nous démontrons comment on peut utiliser le calcul différentiel pour calculer les facteurs d'échelle le long des méridiens et des parallèles dans des projections azimutales et cylindriques. Nous discutons des conditions que ces facteurs d'échelle devraient rencontrer pour qu'une projection soit conforme ou équivalente, puis nous donnons des exemples pour chaque type. Enfin, nous utilisons des facteurs d'échelle pour analyser comment chaque projection affecte les surfaces et les angles et nous discutons de l'application de l'indicatrice de Tissot à ce cas.