A Tale of Three Circles

CHARLES I. DELMAN
GREGORY GALPERIN
Eastern Illinois University
Charleston, IL 61920
cfcid@eiu.edu, cfgg@eiu.edu

Everyone knows that the sum of the angles of a triangle formed by three lines in the plane is 180°, but is this still true for *curvilinear* triangles formed by the arcs of three circles in the plane? We invite the reader to experiment enough to see that the angle sum indeed depends on the triangle, and that no general pattern is obvious. We give a complete analysis of the situation, showing along the way, we hope, what insights can be gained by approaching the problem from several points of view and at several levels of abstraction.

We begin with an elementary solution using only the most basic concepts of Euclidean geometry. While it is direct and very short, this solution is not complete, since it works only in a special case. The key to another special case turns out to be a model of hyperbolic geometry, leading us to suspect that the various manifestations of the problem lie on a continuum of models of geometries with varying curvature. This larger geometric framework reveals many beautiful unifying themes and provides a single method of proof that completely solves the original problem. Finally, we describe a very simple formulation of the solution, whose proof relies on transformations of the plane, a fitting ending we think, since a transformation may be regarded as a change in one's point of view. The background developed earlier informs our understanding of this new perspective, and allows us to give a purely geometric description of the transformations needed.

For the reader who is unfamiliar with the classical noneuclidean geometries, in which the notions of *line* and *distance* are given new interpretations, we provide an overview that is almost entirely self-contained. Such a reader will be introduced to such things as *angle excess*, *stereographic projection*, and even a sphere of imaginary radius. For the reader who is familiar with the three classical geometries, we offer some new ways of looking at them, which we are confident will reveal some surprises.

Three problems

Consider three circles in the plane intersecting transversely (that is, with no circles tangent to each other) at a common point, P, as in FIGURE 1. What is the sum of the measures of the angles of triangle of circular arcs ABC? Answer: 180°! (The picture gives a hint, but we will spell out the solution shortly.)

Next consider FIGURE 2, showing three circles whose centers are collinear. Now two (obviously congruent) curvilinear triangles are formed. What can be said about the sum of the angle measures in this case? Answer: This time, the sum is less than 180°!

Finally, consider three circles that intersect in the pattern of a generic Venn diagram, as in FIGURE 3. The boundary of the common intersection of their interiors is a convex curvilinear triangle; the sum of its angles is greater than 180°, because the straight-sided triangle with the same vertices lies inside it. What about the other six curvilinear

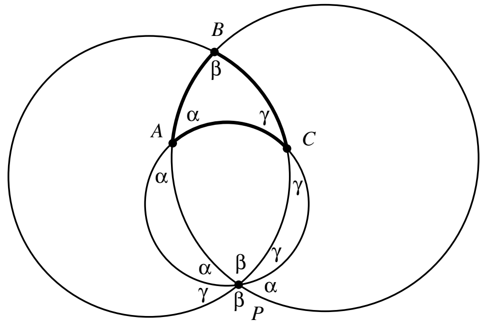


Figure 1 Three circles through a common point

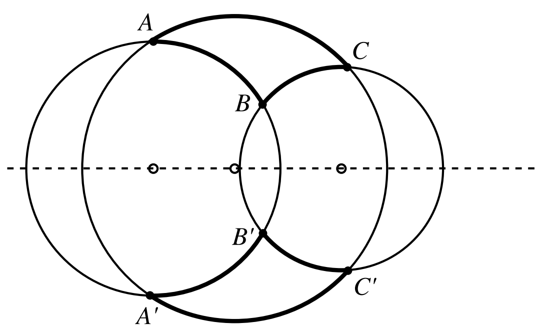


Figure 2 Three circles with collinear centers

triangles formed by these circles? Answer: As the reader might guess, although it is far from obvious from the diagram, the sum in this case is also always greater than 180°!

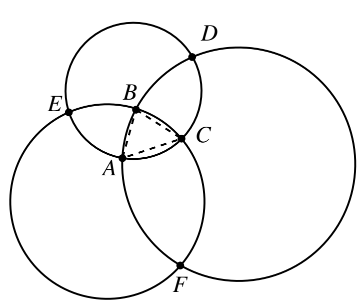


Figure 3 Three circles in Venn diagram position

The solution to the first problem admits an elementary proof. Consider triangle ABC in FIGURE 1. Note that by symmetry, the angles labeled with the letter α have the same measure, as do those labeled by β and by γ. We then see that $2(\alpha + \beta + \gamma) = 360°$, hence $\alpha + \beta + \gamma = 180°$.

The solution to the second problem also admits a simple proof, although a more advanced geometric idea is needed. Namely, the half-plane lying above the line through the three centers may be considered as the upper half-plane model of hyperbolic geometry. In this different sort of geometry, semicircles with centers on the boundary line take the place of lines in Euclidean geometry, and angles between these (hyperbolic) lines are computed using the Euclidean angles between the semicircles. A well-known result in hyperbolic geometry states that the sum of the angle measures of any hyperbolic triangle is always less than 180°.

For the third problem, we will also show that the circles are *lines* in a model of geometry, this time spherical geometry, in which the angle sum of any triangle is more than 180°. More generally, we consider any configuration of three circles that intersect in pairs and give a simple criterion for deciding into which geometry they fall.

The general theorem

Three circles, provided they intersect in the way we have described, determine three lines by their points of pairwise intersection. It may surprise you to learn that these three lines are either concurrent or parallel. The position of the point common to these three lines, as described in the following lemma, is the key to determining which geometry will answer the question about the angle sum.

LEMMA. Let c_1, c_2, and c_3 be three circles in the plane, with any two of them intersecting in two distinct points. Let l_{12}, l_{13}, and l_{23} be the lines determined by these pairwise intersections. Then the lines $\{l_{ij}\}$ are either concurrent or parallel (in which case we consider them to be *concurrent at infinity*). Furthermore, there are exactly three possibilities for the location of the point of concurrency, P:

1. P lies on all three circles (FIGURE 4a);
2. P lies outside all three circles (FIGURE 4b), possibly at infinity; or,
3. P lies inside all three circles (FIGURE 4c).

Finally, if P lies outside all three circles, but not at infinity, the six tangents from P to the three circles all have the same length. The circle centered at P with this common length as radius is perpendicular to all three of the original circles. If P lies at infinity, the line through the centers of all three circles plays this role.

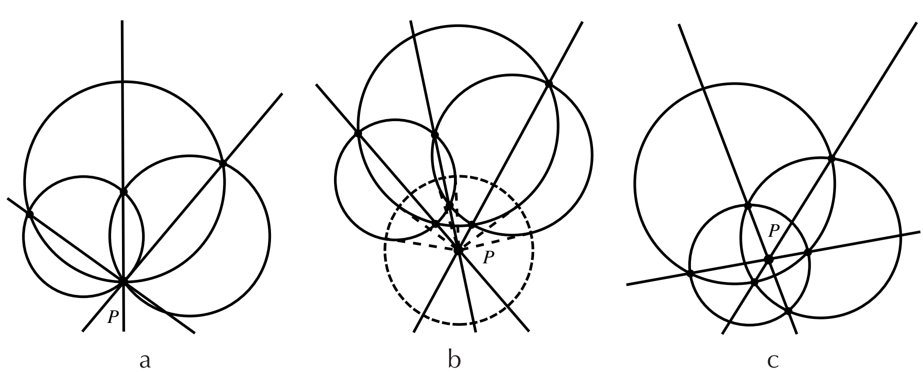


Figure 4 The three possibilities for the location of point P

Proof. Suppose a line through P intersects a circle c in two points A and B (which need not be distinct). The *power of P with respect to c* is defined as the signed product $(PA)(PB)$, where PA denotes the directed distance from P to A. We invite the reader to prove, using similar triangles, that the power does not depend on the line chosen. (A proof may be found in Coxeter and Greitzer [**2**, Theorem 2.11].)

The power of P with respect to c is positive if P is outside c, negative if P is inside c, and 0 if P lies on c. Viewing PA and PB as vectors, we can equivalently define the power as the scalar product $PA \cdot PB$; since the vectors are parallel, the cosine of the angle between them is ± 1 according to whether they point in the same or opposite directions.

If two circles intersect at points A and B, we deduce (by considering line $\overleftrightarrow{PA}$, say) that a point P has the same power with respect to both circles if and only if it lies on line $\overleftrightarrow{AB}$. Suppose first that l_{12} and l_{23} are not parallel and let P be their point of intersection. Then P has the same power with respect to all three circles; hence, P lies on l_{13} and the lines are concurrent. If, on the other hand, two of the lines are parallel, an argument by contradiction shows that all three must be.

If the lines intersect, then, as we have just observed, the power of P with respect to all three circles is the same. Denoting this common power by $\mathcal{P}$, we have:

Case 1. If $\mathcal{P} = 0$, then P lies on all three circles.

Case 2. If $\mathcal{P} > 0$, then P lies outside all three circles.

Case 3. If $\mathcal{P} < 0$, then P lies inside all three circles.

Finally, if P lies outside, then the length of a tangent from P to any of the circles is $\sqrt{\mathcal{P}}$. The circle with this radius and center P is orthogonal to all three circles. ∎

As an interesting digression, we note that for a general pair of circles, which need not intersect, a point has the same power with respect to both if and only if it lies on a particular line, called their *radical axis*, orthogonal to the segment joining their centers. In the course of computing the equation of this line in Cartesian coordinates, Coxeter and Greitzer [**2**, Section 2.2] also prove the beautiful fact that, if the equation of a circle is put in the standard form $F(x, y) = (x - a)^2 + (y - b)^2 - r^2 = 0$, then the power of any point (x, y) with respect to that circle is just $F(x, y)$! This result is not unexpected, since the power is constant on circles concentric with the given one.

We are now prepared to state the theorem. *Triangle* means a curvilinear triangle that is not subdivided by any arcs of the three circles. In case one, only the triangle that does not include P as a vertex is considered.

THEOREM. Let c_1, c_2, and c_3 be three circles in the plane, with each pair intersecting in two distinct points. Let l_{12}, l_{13}, and l_{23} be the lines determined by these pairwise intersections, with common point P. Then the sum of the angles of any triangle formed by the three circles is determined according to the three cases of the lemma:

1. if P lies on all three circles, the sum is equal to 180°;
2. if P lies outside all three circles, the sum is less than 180°; and,
3. if P lies inside all three circles, the sum is greater than 180°.

The proof of the theorem requires some knowledge of the classical noneuclidean geometries, to which we now turn our attention. The reader who is already familiar with these is invited to skip ahead to the section in which the theorem is proved as follows: For any of the possible configurations, we give a conformal map that takes our three circles to the lines of one of the three classical geometries; since the map is conformal, the angle sum is preserved, and thus found to be equal to, less than, or greater than 180° accordingly.

A quick tour of three geometries

A geometry is an abstract mathematical system in which the undefined notions of *point* and *line* are assumed to behave in accordance with certain axioms. (*Euclidean* *and non-Euclidean Geometries*, by Marvin Greenberg [**4**], presents a lively account of the historical development of the classical geometries. Edwin Moise's text, *Elementary Geometry from an Advanced Standpoint* [**7**], is another very comprehensive reference.) In practice, we visualize a geometry by working with a model, which is described more concretely. Objects in the model represent points and lines in such a way that the axioms of the system are fulfilled.

Although there is much to say about the classical geometries, we continue our tale of three circles and focus on the sum of angles in a triangle.

The Cartesian plane, consisting of pairs of real numbers, is a model for Euclidean geometry, provided we adopt the usual concept of distance. Each pair of numbers represents a point, while lines are the solution sets of linear equations. Using algebra to study lines, points, and circles is called the *analytic method*. That the angle measures in a triangle sum to two right angles can be derived either from Euclid's axioms or an analytic approach.

The Cartesian plane and its physical approximations, such as tabletops, in which parallel lines remain equidistant and meet a common transversal line at the same angle on the same side, have historically been described as *flat* (as opposed to a sphere or other *curved* surface); thus, Euclidean geometry is said to be *flat*, or have *curvature zero*.

Any surface on which a distance function, or *metric*, has been defined is a model of some geometry. (These notions may be extended to higher dimensions as well.) The lines on this surface, called *geodesics*, are the curves that locally realize the shortest distances between their points.

Spherical geometry A sphere sitting in $\mathbb{R}^3$, where we know how to measure distances, inherits a metric that measures distances along the surface. Here, the lines are the *great circles*—those circles cut by planes through the sphere's center. The word *local* in the definition of geodesic is important; you must choose the short way around. The geodesic nature of great circles can be understood intuitively by noting that the distance along an arc is proportional to the central angle it subtends, and that central angles obey a sort of triangle inequality: where three planes meet, the sum of any two of the face angles is greater than the third. Thus, a path made up of very small (think *infinitesimal*) arcs will be shortest if all the arcs lie in the same plane through the center. A rigorous proof, as suggested by the preceding discussion, requires integration and other concepts from calculus. (*Geometry from a Differential Viewpoint* by John McCleary [**6**] is an accessible introduction to the application of calculus to geometry.)

Of particular importance to us is the sum of the angles of a spherical (geodesic) triangle. Our calculations will be nicer if we measure angles in radians, which we do from now on. Some experimentation, which we invite the reader to do, suggests that the angle sum of a spherical triangle always exceeds π and decreases with the triangle's area, approaching, but not reaching, π as this area approaches zero (and approaching, but not reaching, 5π as the triangle fills up the whole sphere). This observation suggests that we focus on the amount by which the angle sum of a spherical triangle exceeds π, which we call its *angle excess*. We now prove that a triangle's excess is always positive by precisely examining its relationship with the triangle's area.

Several facts will lead us to the correct relationship. First, notice that the angle excess is additive: if a triangle is subdivided into two smaller triangles, the excesses of the component triangles add up to the excess of the whole, as the reader can calculate. Second, congruent triangles clearly have the same angle excess. These are the essential properties of area: the areas of the parts of a subdivided region add up to the area of the whole, and congruent regions have equal area! Moreover, a sphere, like the plane, is *homogeneous*: any triangle can be rigidly moved, by rotations and reflections of the sphere, to any other position without changing distances or angles (or area). Together, these facts suggest that, for a spherical triangle, angle excess is a constant multiple of area.

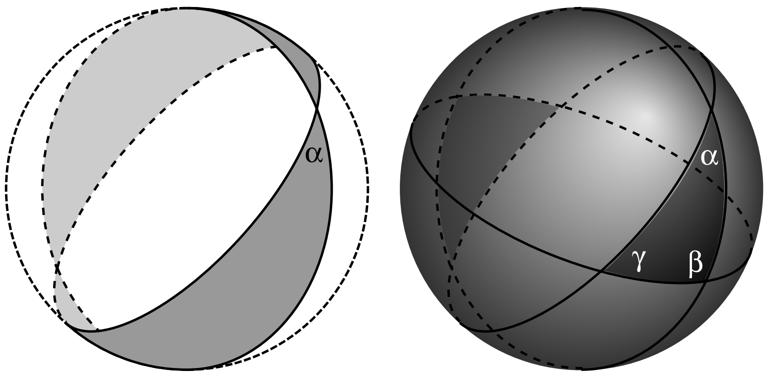


Figure 5 The sector swept out by a spherical angle

To prove that angle excess is proportional to area, observe that the angle between two great circles is proportional to the area of the sector they bound. (Note the essential role of homogeneity here!) On a sphere of radius R, the area of the sector swept out by an angle α is $(\alpha/\pi)(4\pi R^2) = 4R^2\alpha$. (See FIGURE 5a. For simplicity, we have used the same symbol to represent both the angle and its measure.) Let Δ represent the area of a triangle with angles α, β, and γ. The sectors swept out by α, β, and γ cover the sphere redundantly; the triangle and its antipodal image are each covered three times, while the remainder of the sphere is covered exactly once (FIGURE 5b). Thus, with a little algebra, we obtain the formula

$$\alpha + \beta + \gamma - \pi = \frac{\Delta}{R^2}.$$

In particular, on a sphere of unit radius, the angle excess is exactly equal to the area.

Hyperbolic geometry A sphere is said to have *constant positive curvature*, and it is easy to imagine that a small sphere has large curvature, while a large sphere has small curvature. What would it mean for a surface to have constant *negative* curvature? This question will lead us to a model where the sum of the angles in a triangle is always less than two right angles.

To see how we might describe a surface of negative curvature, we start with some formal manipulations on the equation $x^2 + y^2 + z^2 = R^2$. As the constant R^2 increases toward $+\infty$, the curvature of the sphere described by the equation diminishes toward zero. We may view the Euclidean plane as a sphere of infinite radius. What happens when the constant term passes infinity and reappears on the negative end of the number line?

We will call a surface satisfying an equation of the form $x^2 + y^2 + z^2 = -R^2 = (iR)^2$, which has been aptly described as a "sphere of imaginary radius," a *pseudo-sphere*. Of course, the equation will have no solutions unless we expand the domain of our variables from the real to the complex numbers, and its "radius" makes no sense unless we expand our notion of distance to include imaginary numbers.

Distance and angles in Cartesian space are measured via the *dot product*, familiar from multivariable calculus. This product is an example of a symmetric, bilinear form, a type of operation that plays a large role in many branches of mathematics. In addition, the dot product is positive definite: for any vector $\mathbf{v}$, $\mathbf{v}\cdot\mathbf{v}\geq 0$, and $\mathbf{v}\cdot\mathbf{v}=0$ if and only if $\mathbf{v}=0$. The length of a vector $\mathbf{v}$, denoted $|\mathbf{v}|$, is defined as $\sqrt{\mathbf{v}\cdot\mathbf{v}}$ and the angle θ between two vectors is determined by the formula $\cos(\theta)=\mathbf{v}\cdot\mathbf{w}/(|\mathbf{v}||\mathbf{w}|)$.

If we allow the coordinates of our vectors to be any complex numbers, the dot product remains a symmetric, bilinear form, although it is no longer positive definite. A form with these properties is called a *pseudo-metric*. It is this form that we choose to measure distances and angles in complex Cartesian space, $\mathbb{C}^n$, allowing these quantities to have complex values.

Returning to the equation $x^2+y^2+z^2=-R^2$, we consider, to obtain a two-dimensional surface, only those solutions for which x and y are real. The remaining variable, z, is then forced to be purely imaginary, so we let $z=it$. To picture the pseudosphere, we map the solution set of our equation to a surface in $\mathbb{R}^3$, using the map $(x,y,it)\mapsto(x,y,t)$. This image is a two-sheeted hyperboloid, the solution set of the equation $x^2+y^2-t^2=-R^2$. The sheets are analogous to the two hemispheres of an ordinary sphere, the origin $(0,0,0)$ may be thought of as the center, and the points $(0,0,\pm R)$ as the North and South poles.

In order to measure distances and angles, we must remember that our space is an image of a subspace of $\mathbb{C}^3$, where the actual distances and angles are measured using the dot product. Since a vector (x,y,t) actually represents the vector (x,y,it), its actual length is $\sqrt{x^2+y^2+(it)^2}=\sqrt{x^2+y^2-t^2}$. Similarly, the angle between two vectors (x_1,y_1,t_1) and (x_2,y_2,t_2) must be based on the form $x_1x_2+y_1y_2-t_1t_2$. Some readers may recognize this as a Lorentz metric, the three-dimensional version of the pseudo-metric of relativistic space-time. (For a readable and physically motivated, but advanced, introduction to pseudo-metrics, see *Semi-Riemannian Geometry*, by Barret O'Neill [**8**].) The apparent distances and angles in our picture are distorted from their actual values; for example, the true length of every radial vector from the origin to the surface is the imaginary number iR!

To measure distances and angles on the pseudosphere itself, we apply the pseudo-metric to its *tangent vectors*. For any point P in $\mathbb{R}^3$, the *tangent space at P* is the copy of $\mathbb{R}^3$ consisting of all vectors emanating from P. The tangent line at P to a curve through P on the surface is a one-dimensional subspace of this tangent space. The tangent plane to a surface at P is the two-dimensional space composed of all these lines.

It is useful (and intuitive) to write coordinates and other quantities related to tangent vectors in terms of *differential* expressions. If f is a differentiable, real-valued function on $\mathbb{R}^3$, students know how to compute ∇f, and use it to find directional derivatives by taking dot products. We prefer another vocabulary: the *differential* of f is the function that takes a tangent vector $\mathbf{v}_P$, at a point P, to the real number $\nabla f(P)\cdot\mathbf{v}_P$. This gives the best linear approximation to the change in the value of f that results from starting at P and travelling with a displacement $\mathbf{v}_P$.

In particular, the Cartesian coordinates, x, y, and t in our model may be viewed as functions of $P\in\mathbb{R}^3$, and dx, dy, and dt, are the differentials of these functions. If we suppress the vector argument of these differentials, we can use dx, dy, and dt as the first, second, and third coordinates of a general tangent vector with respect to a parallel coordinate system based at P. (In FIGURE 6a, the linear approximations dx, dy, and dt happen to be exact, since orthogonal projection onto an axis is a linear function.) Thus we may conveniently write the (Lorentz) length of a tangent vector as the differential expression $\sqrt{dx^2+dy^2-dt^2}$. This gives us the metric we need in order to talk about the lines of this geometry, which are the geodesics of this metric.

A beautiful fact is that this differential expression is real and positive, when applied to vectors tangent to the pseudosphere. Since the length of a path is computed by integrating the lengths of its tangent vectors, paths between points on the pseudosphere have real, positive length, and it makes sense to talk about geodesics as paths of locally minimal length.

To see that $dx^2 + dy^2 - dt^2 > 0$, it helps to use cylindrical coordinates. As an orthonormal basis for the tangent space to $\mathbb{R}^3$ at the point with cylindrical coordinates (r, θ, t), take a radial unit vector, a circumferential unit vector in the counter-clockwise direction, and a vertical unit vector; with respect to this basis, the coordinates of a tangent vector are $(dr, r d\theta, dt)$. (The factor of r in the circumferential coordinate results from the fact that changing the central angle by a small amount changes the distance travelled by r times that amount. See FIGURE 6b.) The expression $dx^2 + dy^2$, the squared length of the vector's horizontal component, is equal to $dr^2 + r^2\, d\theta^2$. For a vector tangent to the hyperboloid, $dt^2 < dr^2$, since the radial slope of the hyperboloid is less than that of the cone to which it is asymptotic. (Algebraically, it follows from the equation $r^2 - t^2 = -R^2$ that $r^2 < t^2$, and by differentiating we calculate that $dt^2 = (r^2/t^2)\, dr^2$.) Thus $dx^2 + dy^2 - dt^2 = r^2\, d\theta^2 + (dr^2 - dt^2) > 0$.

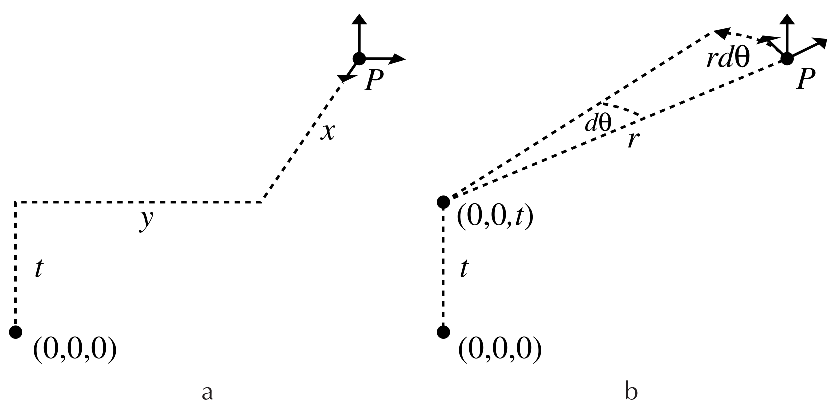


Figure 6 Comparison of two bases for the tangent space at P

The geodesics turn out to be the intersections of the pseudosphere with planes through the origin, analogous to those on an ordinary sphere, although it is harder to see why this is so, and we won't prove it here. The pseudosphere is also homogeneous, and the angle sum of a triangle satisfies a completely analogous formula:

$$\alpha + \beta + \gamma - \pi = \frac{\Delta}{-R^2}.$$

It follows that this sum is always less than π. Of course, there is a different pseudosphere for each value of R. By analogy with the sphere, one might guess that a small value of R yields a pseudosphere of large negative curvature, while a large value of R gives a pseudosphere that is so little curved as to be nearly flat. However great or small the curvature, the sum of the angles in a triangle is still less than two right angles. (A book by B. A. Dubrovin, A. T. Fomenko, and S. P. Novikov [**3**] gives a thorough discussion of both the sphere and pseudosphere.)

Each sheet of the pseudosphere is a model of a geometric object called a *hyperbolic plane*. It would be nice to be able to see this plane looking more like a plane, without having to work with an object as complicated as the Lorentz metric. There is a planar map of the pseudosphere that shows angles accurately, obtained by a method called *stereographic projection*. A similar map is also available for the ordinary sphere. Any map that preserves angle measure is called *conformal*. Since our tale of three circles involves angle measure, conformal maps will be a powerful tool.

Stereographic projection

The sphere To obtain a conformal mapping of the ordinary sphere onto the plane, project from any point of the sphere onto the plane tangent to its antipodal point. Any such map, or its inverse map from the plane to the sphere, is called a *stereographic projection*. Every circle on the sphere (not just the great circles) projects stereographically to a circle or line (which may be thought of as a circle through ∞) in the plane, and it maps to a line if and only if it passes through the point of projection (which maps to ∞). Conversely, every line or circle in the plane is the image of a circle on the sphere. (See FIGURE 7, and note there that the center of a circle on the plane is not, in general, the projection of the center of the corresponding circle on the sphere, but rather the apex of the cone tangent to it.) These valuable properties may be proven by elementary arguments. (The arguments are outlined and nicely illustrated in Hilbert and Cohn-Vossen's classic book, *Geometry and the Imagination* [**5**, pp. 248–251].)

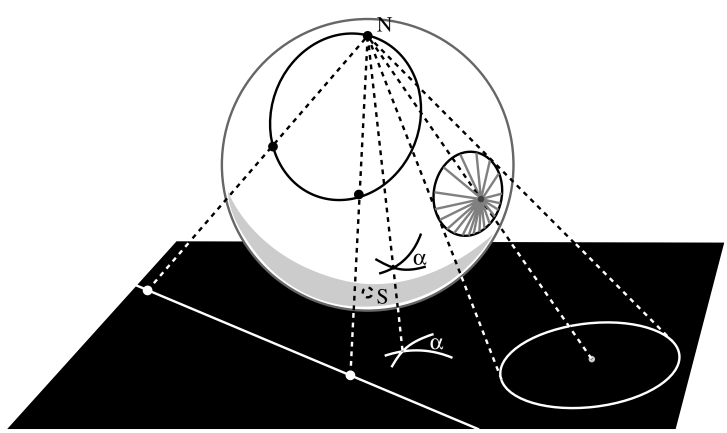


Figure 7 Stereographic projection preserves angles and takes circles to circles or lines

To see that stereographic projection from the sphere to the plane is conformal, consider FIGURE 8a, which shows the cross-section of the sphere cut off by a plane through the point of projection, N, its antipodal point, S, and another point P on the sphere; $\pi(P)$ is the projected image of P, and A is the point where the cross-section of the plane tangent to the sphere at P intersects the tangent plane at S.

An angle with vertex at P in the plane tangent to the sphere is cut by two planes intersecting along line $\overleftrightarrow{NP}$. Since $\angle\pi(P)PA \cong \angle P\pi(P)A$, these planes cut the same angle at $\pi(P)$ in the horizontal plane through S, by symmetry. To see this, imagine an angle formed by two stiff planes of paper; if you snip at the same angle to the spine in either direction, the angle you make is the same. (The projected image of the angle is its reflection in the plane through A that is perpendicular to $\overleftrightarrow{NP\pi(P)}$.)

FIGURE 8b illustrates the local behavior of stereographic projection. If Q is a nearby point on the tangent plane through P, then $\triangle NQP$ is approximately similar to $\triangle N\pi(P)\pi(Q)$. Therefore, stereographic projection is linearly approximated at each point P by a dilation (uniform scaling). The dilation factor varies with the latitude of P, increasing without bound as P approaches N, with a minimum value of 1 at S.

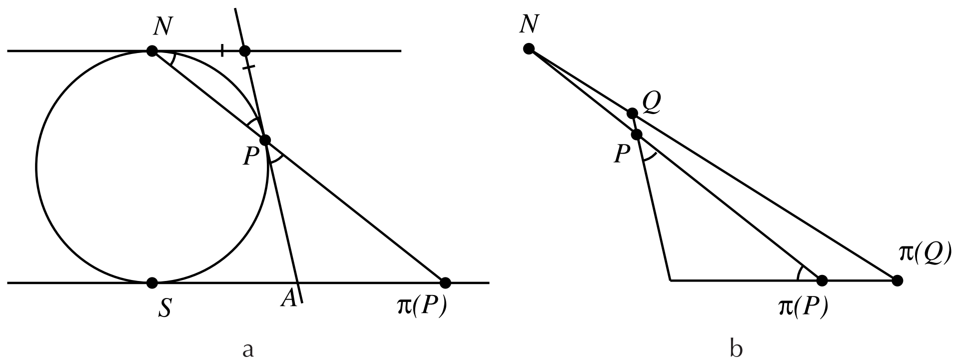


Figure 8 Stereographic projection is conformal and locally approximated by a dilation

The pseudosphere To obtain a conformal map of the pseudosphere onto the plane, project each point of the hyperboloid model onto the horizontal plane through the South pole, $(0, 0, -R)$ (that is, the plane $t = -R$), via the line connecting it to the North pole, $(0, 0, R)$, as in FIGURE 9. This projection maps the southern hemisphere onto the interior of the disk of radius $2R$ centered at the origin, while the northern hemisphere, except for the North pole, goes onto the disk's exterior. The disk's interior is known as a *Poincaré Disk* model of a hyperbolic plane, and its boundary is called the *circle at infinity*. It can be shown that each geodesic maps to a circle or line orthogonal to the circle at infinity (with the points of intersection removed); the geodesics through the poles go to lines. (See B. A. Dubrovin, A. T. Fomenko, and S. P. Novikov [**3**] for a proof.)

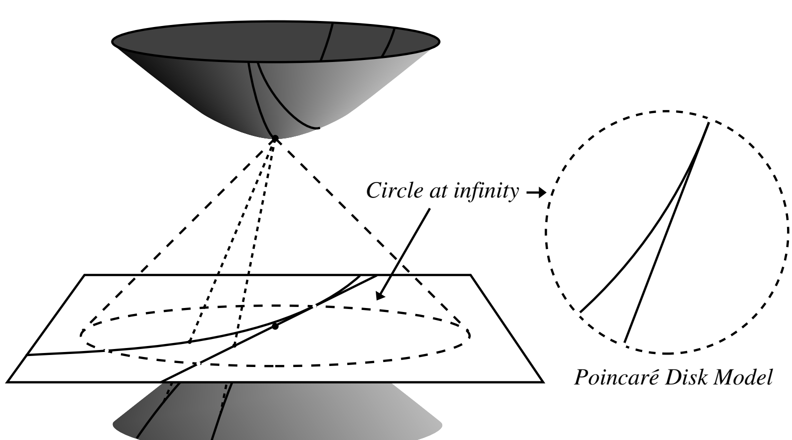


Figure 9 Projection of the hyperboloid model of the pseudosphere onto a plane

The conformality of the projection of the hyperboloid model cannot be demonstrated by elementary geometric arguments because the angles on the hyperboloid are measured using a different bilinear form than the one used to measure angles in the plane. So to prove that angle measure is preserved, we must resort to calculation to compare the angles between tangent vectors to curves, before and after projection. The reader who wishes to just believe us and avoid the technicalities involved may skip the calculation below without any loss of continuity. For those who wish to venture in, the proof provides a nice application of calculus techniques to geometry.

Proof that projection to the Poincaré disk model is conformal We restrict our attention to the southern hemisphere; the proof for the northern hemisphere is similar. Let π denote the projection map, extended to the region $t < R$, and viewed as a map onto $\mathbb{R}^2$ by ignoring the last coordinate $(-R)$ in the image. Suppose curves α and β intersect at $P = (x, y, t)$. The derivative of π at P, $D\pi(P)$, carries vectors tangent to α and β at P to vectors tangent to their projected images at $\pi(P)$ (by the chain rule).

Let $\langle \mathbf{v}_P, \mathbf{w}_P \rangle_L$ denote the Lorentz product of tangent vectors $\mathbf{v}_P$ and $\mathbf{w}_P$ at P, that is,

$$\langle (x_1, y_1, t_1), (x_2, y_2, t_2) \rangle_L = x_1 x_2 + y_1 y_2 - t_1 t_2.$$

We sketch a proof that, if $\mathbf{v}_P$ and $\mathbf{w}_P$ are tangent to the hyperboloid, then $D\pi(P)(\mathbf{v}_P) \cdot D\pi(P)(\mathbf{w}_P) = [4R^2/(R-t)^2]\langle \mathbf{v}_P, \mathbf{w}_P \rangle_L$. In other words, the dot product of the images is just scaled by a constant factor from the Lorentz product of the preimages. It follows that the angle, θ, between $\mathbf{v}_P$ and $\mathbf{w}_P$, which is determined by

$$\cos\theta = \frac{\langle \mathbf{v}_P, \mathbf{w}_P \rangle_L}{\langle \mathbf{v}_P, \mathbf{v}_P \rangle_L^{\frac{1}{2}} \langle \mathbf{w}_P, \mathbf{w}_P \rangle_L^{\frac{1}{2}}}$$

is equal to the angle, ψ, between $D\pi(P)(\mathbf{v}_P)$ and $D\pi(P)(\mathbf{w}_P)$, which is determined by

$$\cos\psi = \frac{D\pi(P)(\mathbf{v}_P) \cdot D\pi(P)(\mathbf{w}_P))}{[D\pi(P)(\mathbf{v}_P) \cdot D\pi(P)(\mathbf{v}_P)]^{\frac{1}{2}} [D\pi(P)(\mathbf{w}_P) \cdot D\pi(P)(\mathbf{w}_P)]^{\frac{1}{2}}},$$

since the scaling factor cancels out. In contrast to the sphere, the scaling factor *decreases* as P moves away from S (t decreases), with a *maximum* value of 1 at S.

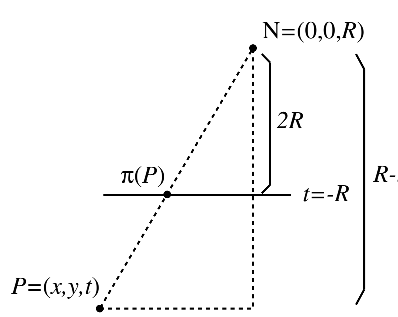


Figure 10 Effect of the projection π on the radial coordinate

Referring to similar triangles, as in FIGURE 10, we see that the image under π of the point P with cylindrical coordinates (r, θ, t) is the point whose polar coordinates are $(\frac{2Rr}{R-t}, \theta)$. The derivative at P of the conversion from Cartesian to cylindrical coordinates takes $(dr, r d\theta, dt)$ to $(dr, d\theta, dt)$. If $\phi : \mathbb{R}^3 \to \mathbb{R}^2$ is the map $(r, \theta, t) \mapsto (\frac{2Rr}{R-t}, \theta)$, its derivative (or Jacobian, if you prefer) at (r, θ, t) is

$$D\phi(r, \theta, t) = \begin{pmatrix} \frac{2R}{R-t} & 0 & \frac{2Rr}{(R-t)^2} \\ 0 & 1 & 0 \end{pmatrix}.$$

Applying $D\phi(r, \theta, t)$ to $(dr, d\theta, dt)$, followed by the derivative at $\phi(r, \theta, t) = (\frac{2Rr}{R-t}, \theta)$ of the conversion from polar back to Cartesian coordinates, we calculate (thanks to the ever-valuable chain rule) that $D\pi(P)(dr, rd\theta, dt) = \frac{2R}{R-t}(dr + \frac{rdt}{R-t}, rd\theta)$.

For any point P on the hyperboloid, $r^2 - t^2 = -R^2$, and for any vector tangent to the hyperboloid at P, $dt = (r/t)\, dr$. After some simplification using these substitutions, we find that $D\pi(P)(dr, rd\theta, (r/t)\, dr) = \frac{2R}{R-t}(-(R/t)dr, rd\theta)$. The reader can now check that for two tangent vectors at P, $D\pi(P)(v_P) \cdot D\pi(P)(w_P) = [4R^2/(R-t)^2]\langle v_P, w_P\rangle_L$, as stated.

Sequences of stereographic projections Stereographic projection is an indispensable tool for transforming geometric models without changing the angles between geodesics. In particular, it provides us with another celebrated model of the hyperbolic plane called the *Poincaré half-plane*. To obtain it, first project the Poincaré disk of radius two from a horizontal plane onto the southern hemisphere of the unit sphere (and its complement, including the point at ∞, onto the other hemisphere). Then project onto any vertical plane tangent to the equator. The image of the circle at infinity under this sequence of projections is called the *line at infinity*. Each half-plane is an image of a hyperbolic plane, as in FIGURE 11.

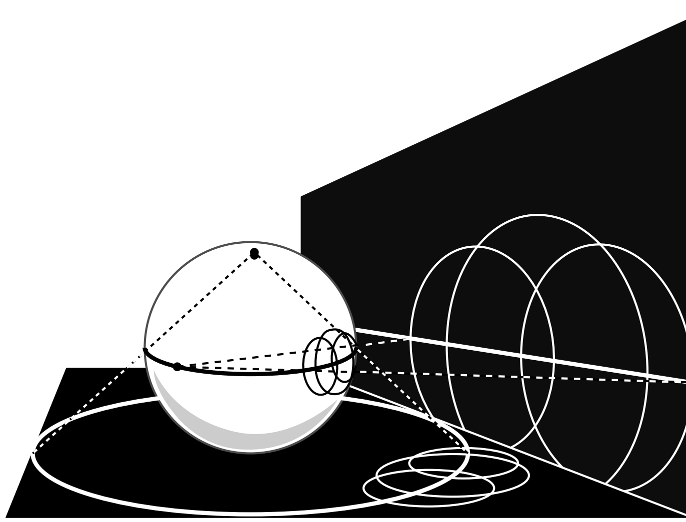


Figure 11 Conformal transformations between the Poincaré disk, hemisphere, and Poincaré half-plane models of hyperbolic geometry

By a sequence of two stereographic projections from different points, we also obtain a conformal map of the Euclidean plane in which the images of the geodesics are either lines or circles. See FIGURE 12.

In summary, we now have maps of the plane, sphere, and pseudosphere in which the geodesics are represented by lines and circles. Just as a map of the earth must be distorted in order to print it on the flat pages of an atlas, our maps distort the true distances between points in the geometric objects they depict. The art of map-making revolves around choosing a projection whose particular type of distortion allows the map to be useful for its intended purpose. For example, the famous Mercator projection of the earth's surface is useful to navigators because it accurately depicts the compass

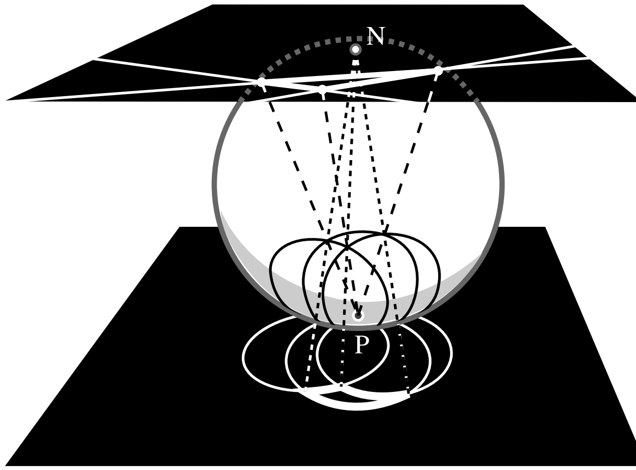


Figure 12 Euclidean geometry on the sphere

bearing between any two points. (McCleary [**6**, chapter 8^{bis}] gives a nice discussion of map projections.)

Since we are interested in the properties of angle measure, we have chosen to view our circles through conformal maps, which render the true angles between smooth curves. The question remains: Given a trio of circles that intersect in pairs, how can we interpret the configuration as a geodesics on a map of one of the geometries we have studied? Is such an interpretation even possible?

Choosing a geometry: a proof of the theorem

Recall that P is the intersection of the three lines determined by the pairwise intersections of the three circles. We will show that, depending on the location of P relative to the circles, there is a conformal map that takes the circles to lines of a standard geometric model—a model for which we know a great deal about angle sums.

Case 1: *P* lies on all three circles. In this case consider a sphere tangent to the plane with South pole at P, and a second plane tangent to this sphere at the North pole, as in FIGURE 12. The images of our three circles on the sphere are circles through the South pole. If we project again, this time from the South pole of the sphere onto the second plane, these circles are taken to straight lines. Thus, c_1, c_2, and c_3 are geodesics in a conformal planar model of Euclidean geometry, and the sum of the angles of the triangle they form is 180°.

Case 2: *P* lies outside all three circles. We already studied the special case where the three lines determined by pairwise intersections of the circles are parallel; this happens when the centers of all three circles are collinear, and that line was seen to be the boundary line of a hyperbolic plane. Having dispensed with that case, we assume that the lines are concurrent at a point P lying outside of all three circles. In this case, the Lemma presents us with a circle, d, that is orthogonal to all three circles. Thus, each arc of the circles lying inside or outside d is a geodesic in a model of a hyperbolic plane, and thus the sum of the angles of any triangle they form is less than 180°.

Case 3: *P* lies inside all three circles. Consider the family of spheres (of varying size) tangent to the plane with South pole at P. Consider how the area of the stereographic projection of the disk bounded by c_1 onto each of these spheres compares to the surface area of the sphere: if the sphere is very small, the projected region will have area more than half that of the sphere (in fact, it will include the entire Southern hemisphere); if the sphere is very large, the projected area will be less than half the surface. This is just a matter of having c_1 lie inside or outside the preimage of the equator on the plane. (See FIGURE 13.) Allowing the sphere to vary continuously, we see that there is a unique sphere, $\mathcal{S}$, such that the projected area on $\mathcal{S}$ is exactly half the surface area. (For an alternative argument, see the third remark below.) Consequently, the image on $\mathcal{S}$ of c_1 is a great circle.

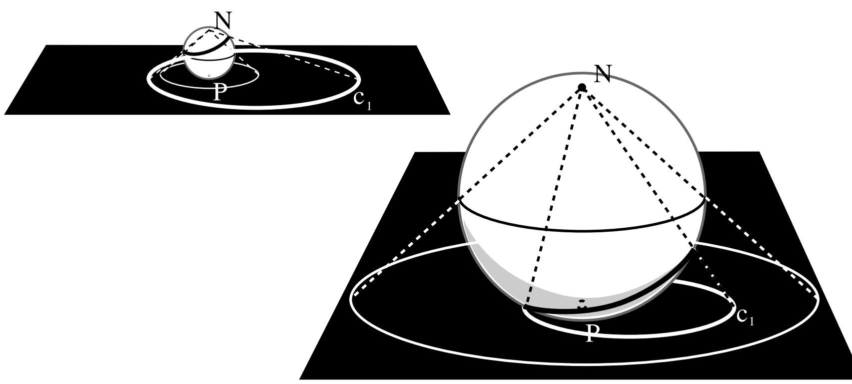


Figure 13 The family of spheres tangent to the plane at P

We claim that the images on $\mathcal{S}$ of c_2 and c_3 are also great circles. It suffices to prove this for c_2, as the same argument works for c_3. To do so, observe that the image of l_{12} under stereographic projection is a *meridian*, that is, a great circle passing through the North and South poles of the sphere. Since the image of c_1 is also a great circle, the images of the points of intersection of c_1 with l_{12} are antipodal. Since the points of $c_1 \cap l_{12}$ also lie on c_2, it follows that the image of c_2 is a great circle. FIGURE 14 illustrates this.

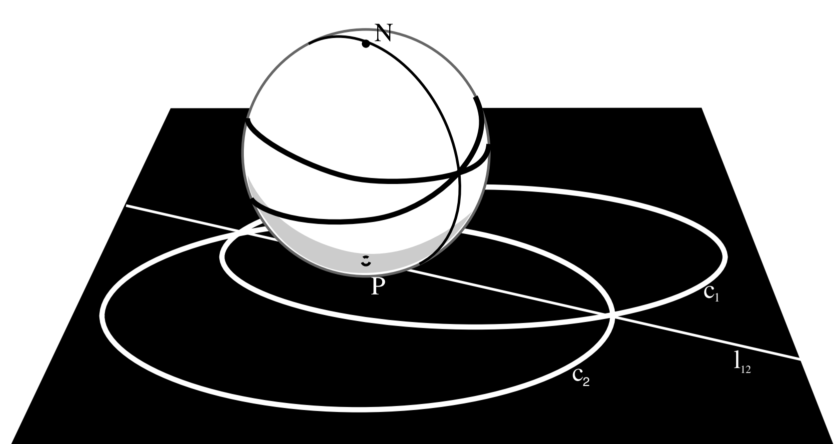


Figure 14 The images of the circles on the sphere $\mathcal{S}$ are great circles

We have thus shown that the images on $\mathcal{S}$ of the three circles are great circles; that is, they are spherical geodesics. Thus c_1, c_2, and c_3 are geodesics in a conformal planar model of spherical geometry, and the sum of the angles of any triangle formed by them is greater than 180°. ■

REMARK. That Euclidean geometry occurs only in the instance that point P lies exactly on the circles is an illustration of the fact that our familiar flat geometry is just a single point in the spectrum of geometries, from the sphere of large radius R, to the flat plane, where R is effectively infinite, to the pseudosphere, whose radius is imaginary in the model we have shown.

REMARK. Using the sequence of stereographic projections described earlier, we may transform the general picture for Case 2 into the half-plane picture.

REMARK. Although we like the continuity argument in Case 3 of the Theorem, it may be avoided as follows. Let $s = \sqrt{-\mathcal{P}(P)}$. The sphere $\mathcal{S}$ in the above proof is the one that has radius $s/2$. To show this, consider, for each of the circles c_1, c_2, and c_3, the chord through P that is perpendicular to the radius through P. Each of these chords has length $2s$ and midpoint P. Let e be the circle centered at P with radius s. This circle projects stereographically to the equator of $\mathcal{S}$, and c_1, c_2, and c_3, which intersect e at diametrically opposite points, also project to great circles. This alternative argument emphasizes the parallels between Cases 2 and 3.

REMARK. The theorem remains true if *circle* is understood in its more general sense to include straight lines as well.

We invite the reader to extend the theorem to the limiting cases in which two or more of the circles are tangent. The possibilities are illustrated in FIGURE 15. Case I

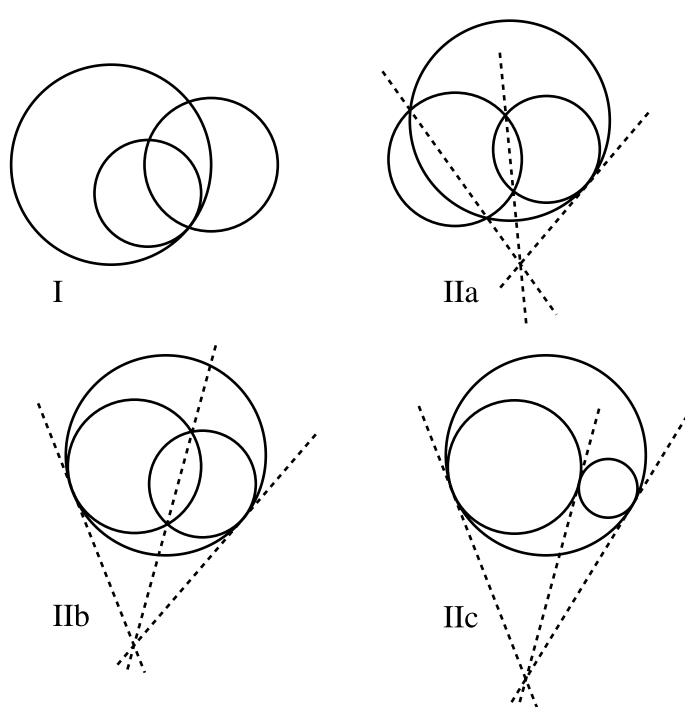


Figure 15 The possibilities for two circles to be tangent

of the figure leads to Euclidean geometry, when viewed with the right map: One vertex of the triangle is thrown out to infinity by the map that takes the circles to straight lines; hence, two sides of the triangle become parallel lines. The possibilities in Case II are hyperbolic, with one or more vertices of the triangle lying on the circle (or line) at infinity. What is the measure of an angle whose vertex lies on the circle at infinity? Note that none of the limiting cases is spherical.

Changing your point of view: a transformational approach

We are indebted to Keith Burns for pointing out the following very elegant formulation and proof of the Theorem. His proof uses the group of Möbius transformations of the extended complex plane (that is, the plane, regarded as the field of complex numbers, together with a point at ∞). Möbius transformations are invertible, bicontinuous, and conformal, and take generalized circles (with lines regarded as circles through ∞) to circles. Moreover, given any two ordered sets consisting of three points each, there is a (unique) Möbius transformation taking each point of one set to the corresponding point of the other. (The excellent, very readable text on geometry from a Kleinian point of view by Brannan, Esplen, and Gray contains a thorough discussion of the Möbius group and its geometric properties [**1**, Chapter 5].)

In particular, and this is all we will need, there is a Möbius transformation taking any given point to ∞. As we have seen, a transformation with the required properties can be constructed by composing a pair of stereographic projections.

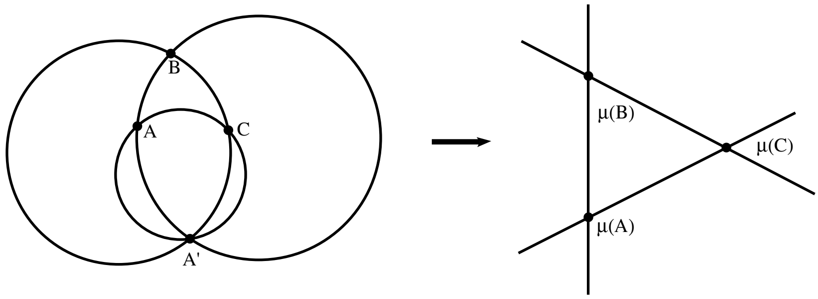
Consider a curvilinear triangle ABC formed by the three circles c_1, c_2, and c_3. Without loss of generality, assume that A lies on c_1 and c_2, and let A' be the other point of intersection of c_1 and c_2. We then have three possibilities for the positions of A and A' with respect to the third circle, c_3, which correspond to the three cases of the Lemma and Theorem: A' lies on c_3; A' lies on the opposite side of c_3 from A; or A' lies on the same side of c_3 as A. By simply changing our point of view, placing A' at ∞, we can discern by inspection how the angle sum of triangle ABC compares to 180°. Although this approach does not demonstrate the underlying global geometry in which the circles are geodesics, it does have the advantage of being delightfully simple and direct.

THEOREM. (ALTERNATIVE FORMULATION) Let c_1, c_2, and c_3 be three circles in the plane, with each pair intersecting in two distinct points. Then exactly one of the following three conditions holds and determines the sum of the angles of any triangle formed by the three circles:

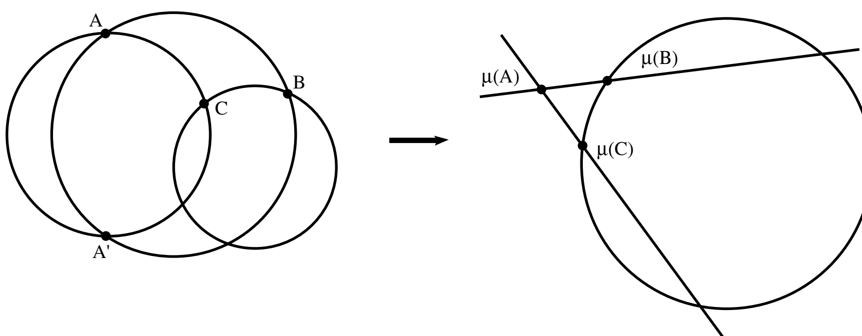
1. the intersection of each pair of circles contains a point of the third circle, in which case the sum of the angles is 180°;
2. the intersection of each pair of circles lies entirely inside or outside the third circle, in which case the sum of the angles is less than 180°; or,
3. the intersection of each pair of circles contains one point inside and one point outside the third circle, in which case the sum of the angles is greater than 180°.

Proof. Let A, B, and C be the vertices of a triangle formed by c_1, c_2, and c_3, with $\{A, A'\} = c_1 \cap c_2$, as in the paragraph preceding the statement of the theorem. The angle sum of the triangle does not depend on which pair of circles we consider, so it suffices to show that this sum is determined by the positions of A and A' relative to c_3.

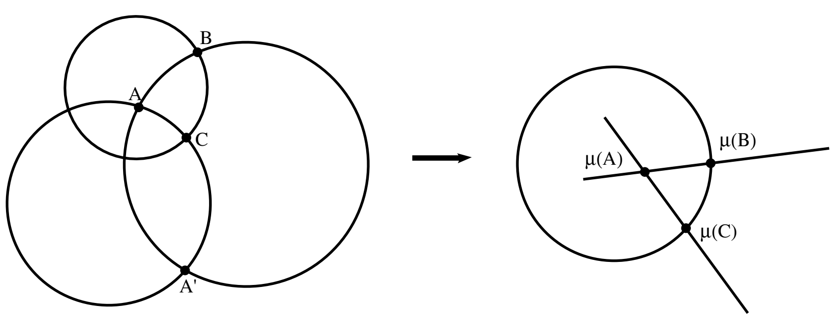
Apply a Möbius transformation, μ, that takes A' to ∞. Under this transformation, the images of c_1 and c_2, which pass through A', are lines. If A' lies on c_3, then the image of c_3 is also a line, hence the angle sum of the image of triangle ABC is 180°.



Case 1



Case 2



Case 3

Figure 16 The effect of a Möbius transformation, μ, taking A' to ∞

If A and A' lie on the same side of c_3, then the image of c_3 is a circle with the image of A *outside* it, since ∞ is outside; hence the angle sum of the image of triangle ABC is less than 180°. If A and A' lie on opposite sides of c_3, then the image of c_3 is a circle with the image of A *inside* it; hence the angle sum of the image of triangle ABC is greater than 180°. The three possibilities are illustrated in FIGURE 16. ■

Conclusion

Each of the classical geometries, Euclidean, spherical, and hyperbolic, has a variety of conformal representations on the plane, obtained by stereographic projection. Recognizing these representations enabled us to classify the generalized triangles whose sides are either segments or circular arcs as belonging to one of three types of geometries, allowing us to make the correct conclusion about the sum of the angles. Finally, by using a group of conformal transformations of the extended plane, we were able to refine our solution to be very simple, although perhaps less informative. Our success is just one example of the value of studying noneuclidean geometries and transformation groups.

Acknowledgments. In addition to Keith Burns, we would like to thank Duane Broline, Leo Comerford, Bob Foote, Yuri Ionin, and Rosemary Schmalz for their helpful comments. We would also like to thank an early referee for pointing out the argument in the third remark.

REFERENCES

1. D. Brannan, M. Esplen, and J. Gray, *Geometry*, Cambridge University Press, Cambridge, 1999.
2. H. S. M. Coxeter and S. L. Greitzer, *Geometry Revisited*, Random House, New York, 1967.
3. B. A. Dubrovin, A. T. Fomenko, and S. P. Novikov, *Modern Geometry—Methods and Applications. Part I. The geometry of surfaces, transformation groups, and fields*, Graduate Texts in Mathematics # 93, Springer-Verlag, New York, 1992.
4. M. J. Greenberg, *Euclidean and non-Euclidean Geometries*, Freeman, New York, 1993.
5. D. Hilbert and S. Cohn-Vossen, *Geometry and the Imagination*, Chelsea, New York, 1952.
6. J. McCleary, *Geometry from a Differentiable Viewpoint*, Cambridge University Press, Cambridge, 1994.
7. E. Moise, *Elementary Geometry from an Advanced Standpoint*, Addison-Wesley, Reading, 1990.
8. B. O'Neill, *Semi-Riemannian Geometry, with Applications to Relativity*, Academic Press, New York, 1983.

Pete, Repete, and One Bagel

Pete: Do you want a bagel?

Repete: Oh, no, I couldn't eat that much.

Pete: I could halve the bagel.

Repete: Yes, you should have it.

Pete: No, if we halve the bagel, we could each have half a bagel.

Repete: Oh, OK. But that brings up a tricky question.

Pete: Yes?

Repete: Before you halve a bagel, a bagel will have a hole. After you halve the bagel, does half a bagel have half a bagel hole, or it is a whole hole?

Pete: A whole!

Repete: But if half a bagel is to have a whole hole, then when you halve a whole bagel, you don't, in fact, halve a bagel hole, since in each half you have a whole hole, which is two holes.

Pete: Mysterious! So when you halve a bagel, you get to have a bagel half and a bagel hole.

Repete: Right! This means that when you have a whole bagel, you get half the bagel holes you get when you halve a whole bagel!

Pete: Holy Cow!

—GLENN APPLEBY
SANTA CLARA UNIVERSITY
SANTA CLARA, CA 95053