Flavors of Geometry
MSRI Publications
Volume **31**, 1997

Hyperbolic Geometry

JAMES W. CANNON, WILLIAM J. FLOYD, RICHARD KENYON,
AND WALTER R. PARRY

Contents

1. Introduction

Hyperbolic geometry was created in the first half of the nineteenth century in the midst of attempts to understand Euclid's axiomatic basis for geometry. It is one type of *non-Euclidean geometry*, that is, a geometry that discards one of Euclid's axioms. Einstein and Minkowski found in non-Euclidean geometry a

This work was supported in part by The Geometry Center, University of Minnesota, an STC funded by NSF, DOE, and Minnesota Technology, Inc., by the Mathematical Sciences Research Institute, and by NSF research grants.

geometric basis for the understanding of physical time and space. In the early part of the twentieth century every serious student of mathematics and physics studied non-Euclidean geometry. This has not been true of the mathematicians and physicists of our generation. Nevertheless with the passage of time it has become more and more apparent that the negatively curved geometries, of which hyperbolic non-Euclidean geometry is the prototype, are the generic forms of geometry. They have profound applications to the study of complex variables, to the topology of two- and three-dimensional manifolds, to the study of finitely presented infinite groups, to physics, and to other disparate fields of mathematics. A working knowledge of hyperbolic geometry has become a prerequisite for workers in these fields.

These notes are intended as a relatively quick introduction to hyperbolic geometry. They review the wonderful history of non-Euclidean geometry. They give five different analytic models for and several combinatorial approximations to non-Euclidean geometry by means of which the reader can develop an intuition for the behavior of this geometry. They develop a number of the properties of this geometry that are particularly important in topology and group theory. They indicate some of the fundamental problems being approached by means of non-Euclidean geometry in topology and group theory.

Volumes have been written on non-Euclidean geometry, which the reader must consult for more exhaustive information. We recommend [Iversen 1993] for starters, and [Benedetti and Petronio 1992; Thurston 1997; Ratcliffe 1994] for more advanced readers. The latter has a particularly comprehensive bibliography.

2. The Origins of Hyperbolic Geometry

Except for Euclid's five fundamental postulates of plane geometry, which we paraphrase from [Kline 1972], most of the following historical material is taken from Felix Klein's book [1928]. Here are Euclid's postulates in contemporary language (compare [Euclid 1926]):

1. Each pair of points can be joined by one and only one straight line segment.
2. Any straight line segment can be indefinitely extended in either direction.
3. There is exactly one circle of any given radius with any given center.
4. All right angles are congruent to one another.
5. If a straight line falling on two straight lines makes the interior angles on the same side less than two right angles, the two straight lines, if extended indefinitely, meet on that side on which the angles are less than two right angles.

Of these five postulates, the fifth is by far the most complicated and unnatural. Given the first four, the fifth postulate can easily be seen to be equivalent to the following *parallel postulate*, which explains why the expressions "Euclid's fifth postulate" and "the parallel parallel" are often used interchangeably:

5′. Given a line and a point not on it, there is exactly one line going through the given point that is parallel to the given line.

For two thousand years mathematicians attempted to deduce the fifth postulate from the four simpler postulates. In each case one reduced the proof of the fifth postulate to the conjunction of the first four postulates with an additional natural postulate that, in fact, proved to be equivalent to the fifth:

Proclus (ca. 400 A.D.) used as additional postulate the assumption that the points at constant distance from a given line on one side form a straight line.

The Englishman John Wallis (1616–1703) used the assumption that to every triangle there is a similar triangle of each given size.

The Italian Girolamo Saccheri (1667–1733) considered quadrilaterals with two base angles equal to a right angle and with vertical sides having equal length and deduced consequences from the (non-Euclidean) possibility that the remaining two angles were not right angles.

Johann Heinrich Lambert (1728–1777) proceeded in a similar fashion and wrote an extensive work on the subject, posthumously published in 1786.

Göttingen mathematician Kästner (1719–1800) directed a thesis of student Klügel (1739–1812), which considered approximately thirty proof attempts for the parallel postulate.

Decisive progress came in the nineteenth century, when mathematicians abandoned the effort to find a contradiction in the denial of the fifth postulate and instead worked out carefully and completely the consequences of such a denial. It was found that a coherent theory arises if instead one assumes that

Given a line and a point not on it, there is more than one line going through the given point that is parallel to the given line.

This postulate is to hyperbolic geometry as the parallel postulate 5′ is to Euclidean geometry.

Unusual consequences of this change came to be recognized as fundamental and surprising properties of non-Euclidean geometry: equidistant curves on either side of a straight line were in fact not straight but curved; similar triangles were congruent; angle sums in a triangle were not equal to π, and so forth.

That the parallel postulate fails in the models of non-Euclidean geometry that we shall give will be apparent to the reader. The unusual properties of non-Euclidean geometry that we have mentioned will all be worked out in Section 13, entitled "Curious facts about hyperbolic space".

History has associated five names with this enterprise, those of three professional mathematicians and two amateurs.

The amateurs were jurist Schweikart and his nephew Taurinus (1794–1874). By 1816 Schweikart had developed, in his spare time, an "astral geometry" that was independent of the fifth postulate. His nephew Taurinus had attained a non-Euclidean hyperbolic geometry by the year 1824.

The professionals were Carl Friedrich Gauss (1777–1855), Nikolaĭ Ivanovich Lobachevskiĭ (1793–1856), and János (or Johann) Bolyai (1802–1860). From the papers of his estate it is apparent that Gauss had considered the parallel postulate extensively during his youth and at least by the year 1817 had a clear picture of non-Euclidean geometry. The only indications he gave of his knowledge were small comments in his correspondence. Having satisfied his own curiosity, he was not interested in defending the concept in the controversy that was sure to accompany its announcement. Bolyai's father Fárkás (or Wolfgang) (1775–1856) was a student friend of Gauss and remained in correspondence with him throughout his life. Fárkás devoted much of his life's effort unsuccessfully to the proof of the parallel postulate and consequently tried to turn his son away from its study. Nevertheless, János attacked the problem with vigor and had constructed the foundations of hyperbolic geometry by the year 1823. His work appeared in 1832 or 1833 as an appendix to a textbook written by his father. Lobachevskiĭ also developed a non-Euclidean geometry extensively and was, in fact, the first to publish his findings, in 1829. See [Lobachevskiĭ 1898; Bolyai and Bolyai 1913].

Gauss, the Bolyais, and Lobachevskiĭ developed non-Euclidean geometry axiomatically on a synthetic basis. They had neither an analytic understanding nor an analytic model of non-Euclidean geometry. They did not prove the *consistency* of their geometries. They instead satisfied themselves with the conviction they attained by extensive exploration in non-Euclidean geometry where theorem after theorem fit consistently with what they had discovered to date. Lobachevskiĭ developed a non-Euclidean trigonometry that paralleled the trigonometric formulas of Euclidean geometry. He argued for the consistency based on the consistency of his analytic formulas.

The basis necessary for an analytic study of hyperbolic non-Euclidean geometry was laid by Leonhard Euler, Gaspard Monge, and Gauss in their studies of curved surfaces. In 1837 Lobachevskiĭ suggested that curved surfaces of constant negative curvature might represent non-Euclidean geometry. Two years later, working independently and largely in ignorance of Lobachevskiĭ's work, yet publishing in the same journal, Minding made an extensive study of surfaces of constant curvature and verified Lobachevskiĭ suggestion. Bernhard Riemann (1826–1866), in his vast generalization [Riemann 1854] of curved surfaces to the study of what are now called Riemannian manifolds, recognized all of these relationships and, in fact, to some extent used them as a springboard for his studies. All of the connections among these subjects were particularly pointed out by Eugenio Beltrami in 1868. This analytic work provided specific analytic models for non-Euclidean geometry and established the fact that non-Euclidean geometry was precisely as consistent as Euclidean geometry itself.

We shall consider in this exposition five of the most famous of the analytic models of hyperbolic geometry. Three are conformal models associated with the name of Henri Poincaré. A conformal model is one for which the metric is a point-by-point scaling of the Euclidean metric. Poincaré discovered his models in the process of defining and understanding Fuchsian, Kleinian, and general automorphic functions of a single complex variable. The story is one of the most famous and fascinating stories about discovery and the work of the subconscious mind in all of science. We quote from [Poincaré 1908]:

> For fifteen days I strove to prove that there could not be any functions like those I have since called Fuchsian functions. I was then very ignorant; every day I seated myself at my work table, stayed an hour or two, tried a great number of combinations and reached no results. One evening, contrary to my custom, I drank black coffee and could not sleep. Ideas rose in crowds; I felt them collide until pairs interlocked, so to speak, making a stable combination. By the next morning I had established the existence of a class of Fuchsian functions, those which come from the hypergeometric series; I had only to write out the results, which took but a few hours.
>
> Then I wanted to represent these functions by the quotient of two series; this idea was perfectly conscious and deliberate, the analogy with elliptic functions guided me. I asked myself what properties these series must have if they existed, and I succeeded without difficulty in forming the series I have called theta-Fuchsian.
>
> Just at this time I left Caen, where I was then living, to go on a geological excursion under the auspices of the school of mines. The changes of travel made me forget my mathematical work. Having reached Coutances, we entered an omnibus to go some place or other. At the moment when I put my foot on the step the idea came to me, without anything in my former thoughts seeming to have paved the way for it, that the transformations I had used to define the Fuchsian functions were identical with those of non-Euclidean geometry. I did not verify the idea; I should not have had time, as, upon taking my seat in the omnibus, I went on with a conversation already commenced, but I felt a perfect certainty. On my return to Caen, for conscience' sake I verified the result at my leisure.

3. Why Call it Hyperbolic Geometry?

The non-Euclidean geometry of Gauss, Lobachevskiĭ, and Bolyai is usually called *hyperbolic geometry* because of one of its very natural analytic models. We describe that model here.

Classically, space and time were considered as independent quantities; an event could be given coordinates $(x_1, \dots, x_{n+1}) \in \mathbb{R}^{n+1}$, with the coordinate x_{n+1} representing time, and the only reasonable metric was the Euclidean metric with the positive definite square-norm $x_1^2 + \cdots + x_{n+1}^2$.

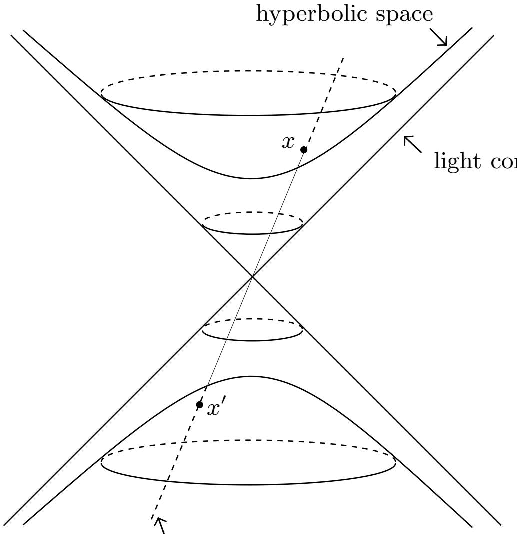


Figure 1. Minkowski space.

Relativity changed all that; in flat space-time geometry the speed of light should be constant as viewed from any inertial reference frame. The Minkowski model for space-time geometry is again $\mathbb{R}^{n+1}$ but with the indefinite norm $x_1^2 + \cdots + x_n^2 - x_{n+1}^2$ defining distance. The *light cone* is defined as the set of points of norm 0. For points $(x_1, \ldots, x_n, x_{n+1})$ on the light cone, the Euclidean space-distance

$$(x_1^2 + \cdots + x_n^2)^{1/2}$$

from the origin is equal to the time x_{n+1} from the origin; this equality expresses the constant speed of light starting at the origin.

These norms have associated inner products, denoted $\cdot$ for the Euclidean inner product and $*$ for the non-Euclidean.

If we consider the set of points at constant squared distance from the origin, we obtain in the Euclidean case the spheres of various radii and in Minkowski space hyperboloids of one or two sheets. We may thus define the unit *n-dimensional sphere* in Euclidean space $\mathbb{R}^{n+1}$ by the formula $S^n = \{x \in \mathbb{R}^{n+1} : x \cdot x = 1\}$ and *n-dimensional hyperbolic space* by the formula $\{x \in \mathbb{R}^{n+1} : x * x = -1\}$. Thus hyperbolic space is a hyperboloid of two sheets that may be thought of as a "sphere" of squared radius -1 or of radius $i = \sqrt{-1}$; hence the name hyperbolic geometry. See Figure 1.

Usually we deal only with one of the two sheets of the hyperboloid or identify the two sheets projectively.

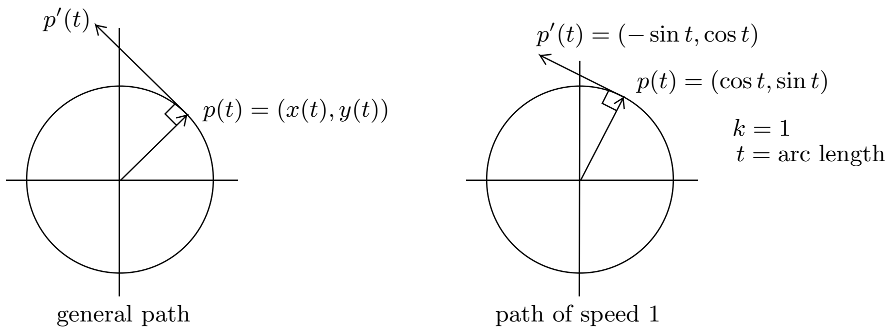


Figure 2. The circle S^1.

4. Understanding the One-Dimensional Case

The key to understanding hyperbolic space H^n and its intrinsic metric coming from the indefinite Minkowski inner product $*$ is to first understand the case $n = 1$. We argue by analogy with the Euclidean case and prepare the analogy by recalling the familiar Euclidean case of the circle S^1.

Let $p : (-\infty, \infty) \to S^1$ be a smooth path with $p(0) = (1, 0)$. If we write in coordinates $p(t) = (x(t), y(t))$ where $x^2 + y^2 = 1$, then differentiating this equation we find

$$2x(t)x'(t) + 2y(t)y'(t) = 0,$$

or in other words $p(t) \cdot p'(t) = 0$. That is, the velocity vector $p'(t)$ is Euclidean-perpendicular to the position vector $p(t)$. In particular we may write $p'(t) = k(t)(-y(t), x(t))$, since the tangent space to S^1 at $p(t)$ is one-dimensional and $(-y(t), x(t))$ is Euclidean-perpendicular to $p = (x, y)$. See Figure 2.

If we assume in addition that $p(t)$ has *constant speed* 1, then

$$1 = |p'(t)| = |k(t)|\sqrt{(-y)^2 + x^2} = |k(t)|,$$

and so $k \equiv \pm 1$. Taking $k \equiv 1$, we see that $p = (x, y)$ travels around the unit circle in the Euclidean plane at constant speed 1. Consequently we may by definition identify t with Euclidean arclength on the unit circle, $x = x(t)$ with $\cos t$ and $y = y(t)$ with $\sin t$, and we see that we have given a complete proof of the fact from beginning calculus that the derivative of the cosine is minus the sine and that the derivative of the sine is the cosine, a proof that is conceptually simpler than the proofs usually given in class.

In formulas, taking $k = 1$, we have shown that x and y (the cosine and sine) satisfy the system of differential equations

$$x'(t) = -y(t), \qquad y'(t) = x(t),$$

with initial conditions $x(0) = 1$, $y(0) = 0$. We then need only apply some elementary method such as the method of undetermined coefficients to easily discover the classical power series for the sine and cosine:

$$\cos t = 1 - t^2/2! + t^4/4! - \cdots,$$
$$\sin t = t - t^3/3! + t^5/5! - \cdots.$$

The hyperbolic calculation in H^1 requires only a new starting point $(0,1)$ instead of $(1,0)$, the replacement of S^1 by H^1, the replacement of the Euclidean inner product $\cdot$ by the hyperbolic inner product $*$, an occasional replacement of $+1$ by -1, the replacement of Euclidean arclength by hyperbolic arclength, the replacement of cosine by hyperbolic sine, and the replacement of sine by the hyperbolic cosine. Here is the calculation.

Let $p : (-\infty,\infty) \to H^1$ be a smooth path with $p(0) = (0,1)$. If we write in coordinates $p(t) = (x(t), y(t))$ where $x^2 - y^2 = -1$, then differentiating this equation we find

$$2x(t)x'(t) - 2y(t)y'(t) = 0;$$

in other words $p(t) * p'(t) = 0$. That is, the velocity vector $p'(t)$ is hyperbolic-perpendicular to the position vector $p(t)$. In particular we may write $p'(t) = k(t)(y(t), x(t))$, since the tangent space to H^1 at $p(t)$ is one-dimensional and the vector $(y(t), x(t))$ is hyperbolic-perpendicular to $p = (x, y)$. See Figure 3.

If we assume in addition that $p(t)$ has *constant speed* 1, then $1 = |p'(t)| = |k(t)|\sqrt{y^2 - x^2} = |k(t)|$, and so $k \equiv \pm 1$. Taking $k \equiv 1$, we see that $p = (x, y)$ travels to the right along the "unit" hyperbola in the Minkowski plane at constant hyperbolic speed 1. Consequently we may by definition identify t with hyperbolic arclength on the unit hyperbola H^1, $x = x(t)$ with $\sinh t$ and $y = y(t)$ with $\cosh t$, and we see that we have given a complete proof of the fact from beginning calculus that the derivative of the hyperbolic cosine is the hyperbolic sine and that the derivative of the hyperbolic sine is the hyperbolic cosine, a proof that is conceptually simpler than the proofs usually given in class.

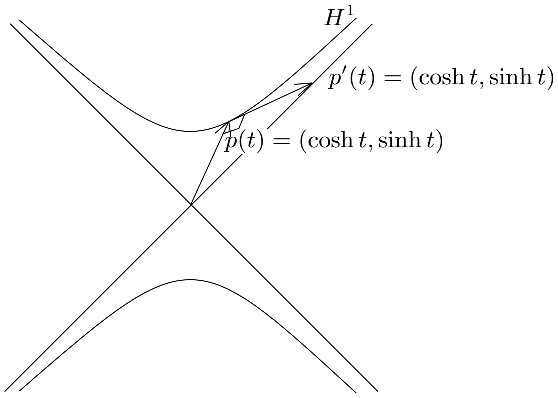


Figure 3. The hyperbolic line H^1.

In formulas, taking $k = 1$, we have shown that x and y (the hyperbolic sine and cosine) satisfy the system of differential equations

$$x'(t) = y(t), \qquad y'(t) = x(t),$$

with initial conditions $x(0) = 0$, $y(0) = 1$. We then need only apply some elementary method such as the method of undetermined coefficients to easily discover the classical power series for the hyperbolic sine and cosine:

$$\cosh t = 1 + t^2/2! + t^4/4! + \cdots,$$
$$\sinh t = t + t^3/3! + t^5/5! + \cdots.$$

It seems to us a shame that these analogies, being as easy as they are, are seldom developed in calculus classes. The reason of course is that the analogies become forced if one is not willing to leave the familiar Euclidean plane for the unfamiliar Minkowski plane.

Note the remarkable fact that our calculation showed that a nonzero tangent vector to H^1 has *positive square norm* with respect to the indefinite inner product $*$; that is, the indefinite inner product on the Minkowski plane restricts to a positive definite inner product on hyperbolic one-space. We shall find that the analogous result is true in higher dimensions and that the formulas we have calculated for hyperbolic length in dimension one apply in the higher-dimensional setting as well.

5. Generalizing to Higher Dimensions

In higher dimensions, H^n sits inside $\mathbb{R}^{n+1}$ as a hyperboloid. If $p : (-\infty, \infty) \to H^n$ again describes a smooth path, then from the defining equations we still have $p(t) * p'(t) = 0$. By taking paths in any direction running through the point $p(t)$, we see that the tangent vectors to H^n at $p(t)$ form the hyperbolic orthogonal complement to the vector $p(t)$ (vectors are hyperbolically orthogonal if their inner product with respect to $*$ is 0).

We can show that the form $*$ restricted to the tangent space is positive definite in either of two instructive ways.

The first method uses the Cauchy–Schwarz inequality $(x \cdot y)^2 \le (x \cdot x)(y \cdot y)$. Suppose that $p = (\hat{p}, p_{n+1})$ is in H^n and $x = (\hat{x}, x_{n+1}) \ne 0$ is in the tangent space of H^n at p, where $\hat{p}, \hat{x} \in \mathbb{R}^n$. If $x_{n+1} = 0$, then $x * x = x \cdot x$. Hence $x * x > 0$ if $x_{n+1} = 0$, so we may assume that $x_{n+1} \ne 0$. Then $0 = x * p = \hat{x} \cdot \hat{p} - x_{n+1} p_{n+1}$, and $-1 = p * p = \hat{p} \cdot \hat{p} - p_{n+1}^2$. Hence, Cauchy–Schwarz gives

$$(\hat{x} \cdot \hat{x})(\hat{p} \cdot \hat{p}) \ge (\hat{x} \cdot \hat{p})^2 = (x_{n+1} p_{n+1})^2 = x_{n+1}^2 (\hat{p} \cdot \hat{p} + 1).$$

Therefore, $(x * x)(\hat{p} \cdot \hat{p}) \ge x_{n+1}^2$, which implies $x * x > 0$ if $x \ne 0$.

The second method analyzes the inner product $*$ algebraically. (For complete details, see [Weyl 1919], for example.) Take a basis $p, p_1, \ldots, p_n$ for $\mathbb{R}^{n+1}$ where p is the point of interest in H^n and the remaining vectors span the n-dimensional tangent space to H^n at p. Now apply the Gram–Schmidt orthogonalization process to this basis. Since $p * p = -1$ by the defining equation for H^n, the vector p, being already a unit vector, is unchanged by the process and the remainder of the resulting basis spans the orthogonal complement of p, which is the tangent space to H^n at p. Since the inner product $*$ is nondegenerate, the resulting matrix is diagonal with entries of ± 1 on the diagonal, one of the -1's corresponding to the vector p. By Sylvester's theorem of inertia, the number of $+1$'s and -1's on the diagonal is an invariant of the inner product (the number of 1's is the dimension of the largest subspace on which the metric is positive definite). But with the standard basis for $\mathbb{R}^{n+1}$, there is exactly one -1 on the diagonal and the remaining entries are $+1$. Hence the same is true of our basis. Thus the matrix of the inner product when restricted to our tangent space is the identity matrix of order n; that is, the restriction of the metric to the tangent space is positive definite.

Thus the inner product $*$ restricted to H^n defines a genuine Riemannian metric on H^n.

6. Rudiments of Riemannian Geometry

Our analytic models of hyperbolic geometry will all be differentiable manifolds with a Riemannian metric.

One first defines a Riemannian metric and associated geometric notions on Euclidean space. A Riemannian metric ds^2 on Euclidean space $\mathbb{R}^n$ is a function that assigns at each point $p \in \mathbb{R}^n$ a positive definite symmetric inner product on the tangent space at p, this inner product varying differentiably with the point p. Given this inner product, it is possible to define any number of standard geometric notions such as the length $|x|$ of a vector x, where $|x|^2 = x \cdot x$, the angle θ between two vectors x and y, where $\cos\theta = (x \cdot y)/(|x| \cdot |y|)$, the length element $ds = \sqrt{ds^2}$, and the area element dA, where dA is calculated as follows: if $x_1, \ldots,$ x_n are the standard coordinates on $\mathbb{R}^n$, then ds^2 has the form $\sum_{i,j} g_{ij}\, dx_i\, dx_j$, and the matrix (g_{ij}) depends differentiably on x and is positive definite and symmetric. Let $\sqrt{|g|}$ denote the square root of the determinant of (g_{ij}). Then $dA = \sqrt{|g|}\, dx_1\, dx_2 \cdots dx_n$. If $f : \mathbb{R}^k \to \mathbb{R}^n$ is a differentiable map, one can define the pullback $f^*(ds^2)$ by the formula

$$f^*(ds^2)(v, w) = ds^2(Df(v), Df(w))$$

where v and w are tangent vectors at a point u of $\mathbb{R}^k$ and Df is the derivative map that takes tangent vectors at u to tangent vectors at $x = f(u)$. One can also calculate the pullback formally by replacing $g_{ij}(x)$ with $x \in \mathbb{R}^n$ by $g_{ij} \circ f(u)$, where $u \in \mathbb{R}^k$ and $f(u) = x$, and replacing dx_i by $\sum_j (\partial f_i/\partial u_j) du_j$. One can

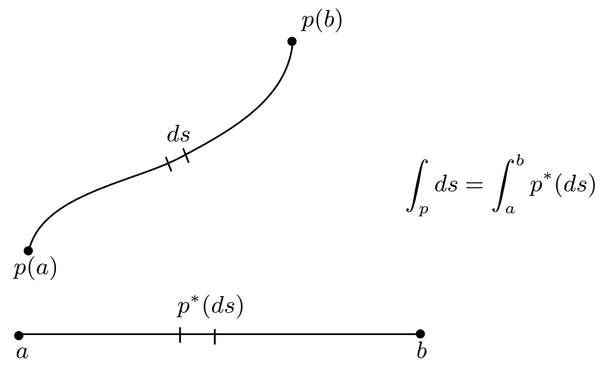


Figure 4. The length of a path.

calculate the length of a path $p : [a, b] \to \mathbb{R}^n$ by integrating ds over p:

$$\int_p ds = \int_a^b p^*(ds).$$

See Figure 4. The Riemannian distance $d(p, q)$ between two points p and q in $\mathbb{R}^n$ is defined as the infimum of path length over all paths joining p and q.

Finally, one generalizes all of these notions to manifolds by requiring the existence of a Riemannian metric on each coordinate chart with these metrics being invariant under pullback on transition functions connecting these charts; that is, if ds_1^2 is the Riemannian metric on chart one and if ds_2^2 is the Riemannian metric on chart two and if f is a transition function connecting these two charts, then $f^*(ds_2^2) = ds_1^2$. The standard change of variables formulas from calculus show that path lengths and areas are invariant under chart change.

7. Five Models of Hyperbolic Space

We describe here five analytic models of hyperbolic space. The theory of hyperbolic geometry could be built in a unified way within a single model, but with several models it is as if one were able to turn the object that is hyperbolic space about in one's hands so as to see it first from above, then from the side, and finally from beneath or within; each view supplies its own natural intuitions. Each model has its own metric, geodesics, isometries, and so on. Here are our mnemonic names for the five models:

- H, the Half-space model.
- I, the Interior of the disk model.
- J, the Jemisphere model (pronounce the J as in Spanish).
- K, the Klein model.
- L, the 'Loid model (short for hyperboloid).

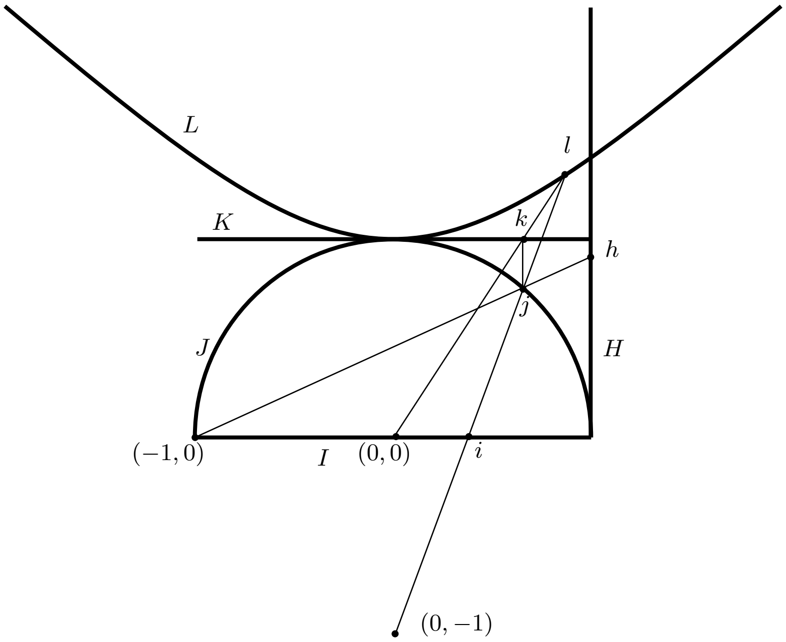


Figure 5. The five analytic models and their connecting isometries. The points $h \in H$, $i \in I$, $j \in J$, $k \in K$, and $l \in L$ can be thought of as the same point in (synthetic) hyperbolic space.

Each model is defined on a different subset of $\mathbb{R}^{n+1}$, called its *domain;* for $n = 1$ these sets are schematically indicated in Figure 5, which can also be regarded as a cross section of the picture in higher dimensions. Here are the definitions of the five domains:

$$
\begin{aligned}
H &= \{(1, x_2, \ldots, x_{n+1}) : x_{n+1} > 0\};\\
I &= \{(x_1, \ldots, x_n, 0) : x_1^2 + \cdots + x_n^2 < 1\};\\
J &= \{(x_1, \ldots, x_{n+1}) : x_1^2 + \cdots + x_{n+1}^2 = 1 \text{ and } x_{n+1} > 0\};\\
K &= \{(x_1, \ldots, x_n, 1) : x_1^2 + \cdots + x_n^2 < 1\};\\
L &= \{(x_1, \ldots, x_n, x_{n+1}) : x_1^2 + \cdots + x_n^2 - x_{n+1}^2 = -1 \text{ and } x_{n+1} > 0\}.
\end{aligned}
$$

The associated Riemannian metrics ds^2 that complete the analytic description of the five models are:

$$\begin{aligned}
ds_H^2 &= \frac{dx_2^2 + \cdots + dx_{n+1}^2}{x_{n+1}^2};\\
ds_I^2 &= 4\frac{dx_1^2 + \cdots + dx_n^2}{(1 - x_1^2 - \cdots - x_n^2)^2};\\
ds_J^2 &= \frac{dx_1^2 + \cdots + dx_{n+1}^2}{x_{n+1}^2};\\
ds_K^2 &= \frac{dx_1^2 + \cdots + dx_n^2}{(1 - x_1^2 - \cdots - x_n^2)} + \frac{(x_1\,dx_1 + \cdots + x_n\,dx_n)^2}{(1 - x_1^2 - \cdots - x_n^2)^2};\\
ds_L^2 &= dx_1^2 + \cdots + dx_n^2 - dx_{n+1}^2.
\end{aligned}$$

To see that these five models are isometrically equivalent, we need to describe isometries among them. We use J as the central model and describe for each of the others a simple map to or from J.

The map $\alpha : J \to H$ is central projection from the point $(-1, 0, \ldots, 0)$:

$$\alpha : J \to H, \quad (x_1, \ldots, x_{n+1}) \mapsto (1, 2x_2/(x_1 + 1), \ldots, 2x_{n+1}/(x_1 + 1)).$$

The map $\beta : J \to I$ is central projection from $(0, \ldots, 0, -1)$:

$$\beta : J \to I, \quad (x_1, \ldots, x_{n+1}) \mapsto (x_1/(x_{n+1} + 1), \ldots, x_n/(x_{n+1} + 1), 0).$$

The map $\gamma : K \to J$ is vertical projection:

$$\gamma : K \to J, \quad (x_1, \ldots, x_n, 1) \mapsto \big(x_1, \ldots, x_n, \sqrt{1 - x_1^2 - \cdots - x_n^2}\,\big).$$

The map $\delta : L \to J$ is central projection from $(0, \ldots, 0, -1)$:

$$\delta : L \to J, \quad (x_1, \ldots, x_{n+1}) \mapsto (x_1/x_{n+1}, \ldots, x_n/x_{n+1}, 1/x_{n+1}).$$

Each map can be used in the standard way to pull back the Riemannian metric from the target domain to the source domain and to verify thereby that the maps are isometries. Among the twenty possible connecting maps among our models, we have chosen the four for which we find the calculation of the metric pullback easiest. It is worth noting that the metric on the Klein model K, which has always struck us as particularly ugly and unintuitive, takes on obvious meaning and structure relative to the metric on J from which it naturally derives via the connecting map $\gamma : K \to J$. We perform here two of the four pullback calculations as examples and recommend that the reader undertake the other two.

Here is the calculation that shows that $\alpha^*(ds_H^2) = ds_J^2$. Set

$$y_2 = 2x_2/(x_1 + 1),\ \ldots,\ y_{n+1} = 2x_{n+1}/(x_1 + 1).$$

Then

$$dy_i = \frac{2}{x_1 + 1}(dx_i - \frac{x_i}{x_1 + 1}\,dx_1).$$

Since $x_1^2+\cdots+x_{n+1}^2=1$,

$$x_1\,dx_1=-(x_2\,dx_2+\cdots+x_{n+1}\,dx_{n+1})$$

and

$$x_2^2+\cdots+x_{n+1}^2=1-x_1^2.$$

These equalities justify the following simple calculation:

$$\begin{aligned}\alpha^*(ds_H^2)&=\frac{1}{y_{n+1}^2}(dy_2^2+\cdots+dy_{n+1}^2)\\&=\frac{(x_1+1)^2}{4x_{n+1}^2}\,\frac{4}{(x_1+1)^2}\left(\sum_{i=2}^{n+1}dx_i^2-\frac{2dx_1}{x_1+1}\sum_{i=2}^{n+1}x_i\,dx_i+\frac{dx_1^2}{(x_1+1)^2}\sum_{i=2}^{n+1}x_i^2\right)\\&=\frac{1}{x_{n+1}^2}\left(\sum_{i=2}^{n+1}dx_i^2+\frac{2}{(x_1+1)}\cdot x_1\,dx_1^2+\frac{dx_1^2}{(x_1+1)^2}(1-x_1^2)\right)\\&=\frac{1}{x_{n+1}^2}\sum_{i=1}^{n+1}dx_i^2\\&=ds_J^2.\end{aligned}$$

Here is the calculation that shows that $\gamma^*(ds_J^2)=ds_K^2$. Set $y_1=x_1,\ \dots,\ y_n=x_n$, and $y_{n+1}^2=1-y_1^2-\cdots-y_n^2=1-x_1^2-\cdots-x_n^2$. Then $dy_i=dx_i$ for $i=1,\dots,n$ and $y_{n+1}\,dy_{n+1}=-(x_1\,dx_1+\cdots+x_n\,dx_n)$. Thus

$$\begin{aligned}\gamma^*(ds_J^2)&=\frac{1}{y_{n+1}^2}(dy_1^2+\cdots+dy_n^2)+\frac{1}{y_{n+1}^2}dy_{n+1}^2\\&=\frac{1}{(1-x_1^2-\cdots-x_n^2)}(dx_1^2+\cdots+dx_n^2)+\frac{(x_1\,dx_1+\cdots+x_n\,dx_n)^2}{(1-x_1^2-\cdots-x_n^2)^2}\\&=ds_K^2.\end{aligned}$$

The other two pullback computations are comparable.

8. Stereographic Projection

In order to understand the relationships among these models, it is helpful to understand the geometric properties of the connecting maps. Two of them are *central* or *stereographic* projection from a sphere to a plane. In this section we develop some important properties of stereographic projection. We begin with the definition and then establish the important properties that stereographic projection *preserves angles* and *takes spheres to planes or spheres.* We give a geometric proof in dimension three and an analytic proof in general.

DEFINITION. Let S^n denote a sphere of dimension n in Euclidean $(n+1)$-dimensional space $\mathbb{R}^{n+1}$. Let P denote a plane tangent to the sphere S^n at the point S, which we think of as the south pole of S^n. Let N denote the point of S^n opposite S, a point that we think of as the north pole of S^n. If x is any point of $S^n \setminus \{N\}$, there is a unique point $\pi(x)$ of P on the line that contains N and x. It is called the *stereographic projection* from x into P. See Figure 6. Note that π has a natural extension, also denoted by π, which takes all of $\mathbb{R}^{n+1}$ except for the plane $\{x : x_{n+1} = 1\}$ into P.

Theorem 8.1 (Conformality, or the preservation of angles). *Let $S^n \subset \mathbb{R}^{n+1}$, P, S, N, and π (extended) be as in the definition. Then π preserves angles between curves in $S^n \setminus \{N\}$. Furthermore, if $x \in S^n \setminus \{N, S\}$ and if $T = xy$ is a line segment tangent to S^n at x, then the angles $\pi(x)\,x\,y$ and $x\,\pi(x)\,\pi(y)$ are either equal or complementary whenever $\pi(y)$ is defined.*

Proof. We first give the analytic proof in arbitrary dimensions that π preserves angles between curves in $S^n \setminus \{N\}$.

We may clearly normalize everything so that S^n is in fact the unit sphere in $\mathbb{R}^{n+1}$, S is the point with coordinates $(0, \dots, 0, -1)$, N is the point with coordinates $(0, \dots, 0, 1)$, P is the plane $x_{n+1} = -1$, and $\pi : S^n \to P$ is given by the formula $\pi(x) = (y_1, \dots, y_n, -1)$, where

$$y_i = \frac{-2}{x_{n+1} - 1}\, x_i.$$

We take the Euclidean metric $ds^2 = dy_1^2 + \cdots + dy_n^2$ on P and pull it back to a metric $\pi^*(ds^2)$ on S^n. The pullback of dy_i is the form

$$\frac{-2}{x_{n+1} - 1}\Bigl(dx_i - \frac{x_i}{x_{n+1} - 1} dx_{n+1}\Bigr).$$

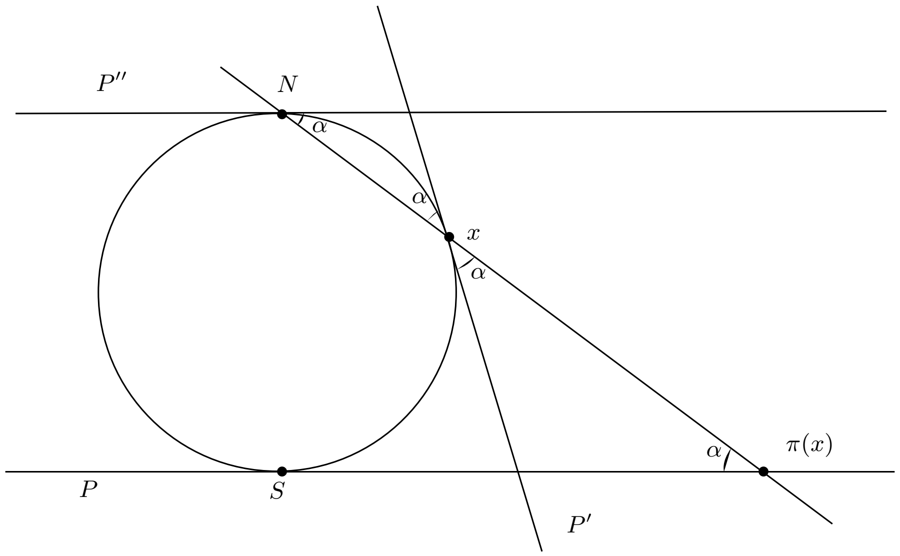


Figure 6. Stereographic projection.

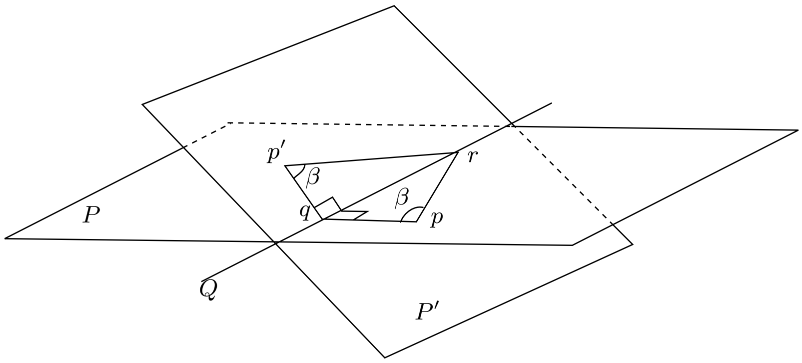


Figure 7. The angles qpr and $qp'r$.

Because $x \in S^n$, we have the two equations

$$x_1^2 + \cdots + x_n^2 + x_{n+1}^2 = 1$$

and

$$x_1\,dx_1 + \cdots + x_n\,dx_n + x_{n+1}\,dx_{n+1} = 0.$$

From these equations it is easy to deduce that

$$\pi^*(ds^2) = \frac{4}{(x_{n+1}-1)^2}(dx_1^2 + \cdots + dx_n^2 + dx_{n+1}^2);$$

the calculation is essentially identical with one that we have performed above. We conclude that at each point the pullback of the Euclidean metric on P is a positive multiple of the Euclidean metric on S^n. Since multiplying distances in a tangent space by a positive constant does not change angles, the map $\pi : S^n \setminus \{N\} \to P$ preserves angles.

For the second assertion of the theorem we give a geometric proof that, in the special case of dimension $n+1=3$, also gives an alternative geometric proof of the fact that we have just proved analytically. This proof is taken from [Hilbert and Cohn-Vossen 1932].

In preparation we consider two planes P and P' of dimension n in Euclidean $(n+1)$-space $\mathbb{R}^{n+1}$ that intersect in a plane Q of dimension $n-1$. We then pick points $p \in P$, $q \in Q$, and $p' \in P'$ such that the line segments pq and $p'q$ are of equal length and are at right angles to Q.

As can be seen in Figure 7, if $r \in Q$, the angles qpr and $qp'r$ are equal. Similarly, the angles $p'pr$ and $pp'r$ are equal.

To prove the second assertion of the theorem, first note that the case in which the line M containing x and y misses P follows by continuity from the case in which M meets P. So suppose that M meets P. Note that π maps the points of M for which π is defined to the line containing $\pi(x)$ and $\pi(y)$. This implies that we may assume that $y \in P$. See Figure 6. Now for the plane P of the obvious assertion we take the plane P tangent to the sphere S^n at the south pole S. For the plane P' of the obvious assertion we take the plane tangent to S^n at x. For the points $p' \in P'$ and $p \in P$ we take, respectively, the points $p = \pi(x) \in P$ and $p' = x \in P'$. For the plane Q we take the intersection of P and P'. For the point r we take y. Now the assertion that the angles $p'pr$ and $pp'r$ are equal proves the second assertion of the theorem.

In dimension three, the obvious fact that the angles qpr and $qp'r$ are equal shows that π preserves the angle between any given curve and certain reference tangent directions, namely pq and $p'q$. Since the tangent space is, in this dimension only, two-dimensional, preserving angle with reference tangent directions is enough to ensure preservation of angle in general. $\square$

Theorem 8.2 (Preservation of spheres). *Assume the setting of the previous theorem. If C is a sphere (C for circle) in S^n that passes through the north pole N of S^n and has dimension c, then the image $\pi(C) \subset P$ is a plane in P of dimension c. If on the other hand C misses N, then the image $\pi(C)$ is a sphere in P of dimension c.*

Proof. If $N \in C$, then the proof is easy; indeed C is contained in a unique plane P' of dimension $c+1$, and the image $\pi(C)$ is the intersection of P' and P, a c-dimensional plane.

If, on the other hand, C misses N, we argue as follows. We assume all normalized as in the analytic portion of the proof of the previous theorem so that S^n is the unit sphere. We can deal with the case where C is a union of great circles by continuity if we manage to prove the theorem in all other cases. Consequently, we may assume that the vector subspace of $\mathbb{R}^{n+1}$ spanned by the vectors in C has dimension $c+2$. We lose no generality in assuming that it is all of $\mathbb{R}^{n+1}$ (that is, $c = n-1$).

The tangent spaces to S^n at the points of C define a conical envelope with cone point y; one easy way to find y is to consider the two-dimensional plane R containing N and two antipodal points r and r' of C, and to consider the two tangent lines $t(r)$ to $C \cap R$ at r and $t(r')$ to $C \cap R$ at r'; then y is the point at which $t(r)$ and $t(r')$ meet. See Figure 8. By continuity we may assume that $\pi(y)$ is defined.

We assert that $\pi(y)$ is equidistant from the points of $\pi(C)$, from which the reader may deduce that $\pi(C)$ is a sphere centered at $\pi(y)$. By continuity it suffices to prove that $\pi(y)$ is equidistant from the points of $\pi(C) \setminus S$. Here is the argument that proves the assertion. Let $x \in C \setminus S$, and consider the two-dimensional plane containing N, x, and y. In this plane there is a point x' on the line through x and N such that the line segment yx' is parallel to the segment $\pi(y)\pi(x)$; that is, the angles $N\,\pi(x)\,\pi(y)$ and $N\,x'\,y$ are equal. By the final assertion of Theorem 8.1, the angles $\pi(y)\,\pi(x)\,x$ and $y\,x\,\pi(x)$ are either equal or complementary. Thus the triangle xyx' is isosceles so that sides xy and $x'y$ are equal. Thus considering proportions in the similar triangles $N\,x'\,y$ and

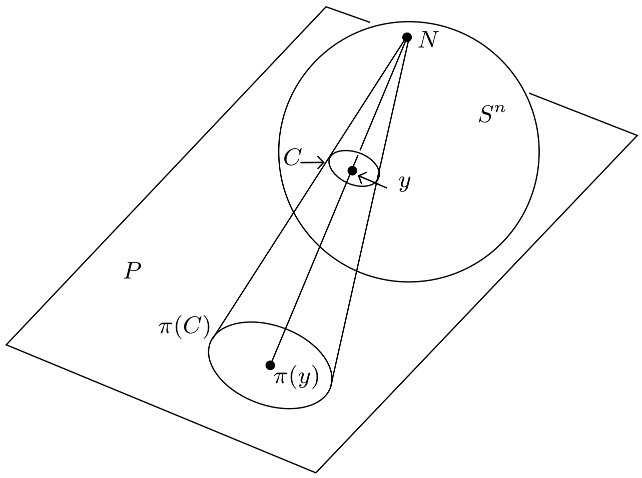


Figure 8. Stereographic projection maps spheres to spheres.

$N\,\pi(x)\,\pi(y)$, we have the equalities

$$d(\pi(x), \pi(y)) = \frac{d(N, \pi(y))}{d(N, y)}\, d(x', y) = \frac{d(N, \pi(y))}{d(N, y)}\, d(x, y).$$

Of course, the fraction is a constant since N, y, and $\pi(y)$ do not depend on x; and the distance $d(x, y)$ is also a constant since $x \in C$, C is a sphere, and y is the center of the tangent cone of C. We conclude that the distance $d(\pi(x), \pi(y))$ is constant. □

Definition. Let S^n denote a sphere of dimension n in $\mathbb{R}^{n+1}$ with north pole N and south pole S as above. Let P denote a plane through the center of S^n and orthogonal to the line through N and S. If x is any point of $S^n \setminus \{N\}$, then there is a unique point $\pi'(x)$ of P on the line that contains N and x. This defines a map $\pi' : S^n \setminus \{N\} \to P$, *stereographic projection* from $S^n \setminus \{N\}$ to P.

Theorem 8.3. *The map π' preserves angles between curves in $S^n \setminus \{N\}$, and π' maps spheres to planes or spheres.*

Proof. We normalize so that S^n is the unit sphere in $\mathbb{R}^{n+1}$, $N = (0, \dots, 0, 1)$, and $S = (0, \dots, 0, -1)$. From the proof of Theorem 8.1 we have for every $x \in S^n \setminus \{N\}$ that $\pi(x) = (y_1, \dots, y_n, -1)$, where

$$y_i = \frac{-2}{x_{n+1} - 1}\, x_i.$$

In the same way $\pi'(x) = (y'_1, \dots, y'_n, -1)$, where

$$y'_i = \frac{-1}{x_{n+1} - 1}\, x_i = \frac{y_i}{2}.$$

Thus π' is the composition of π with a translation and a dilation. Since π preserves angles and maps spheres to planes or spheres, so does π'. □

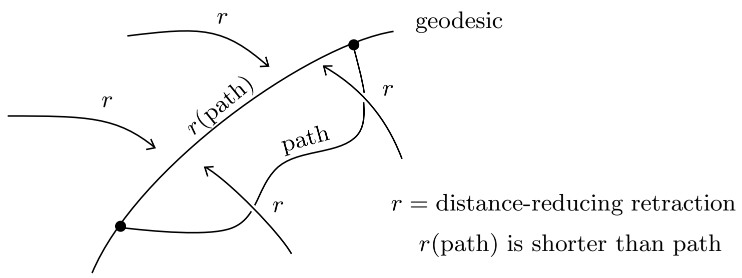


Figure 9. The retraction principle.

9. Geodesics

Having established formulas for the hyperbolic metric in our five analytic models and having developed the fundamental properties of stereographic projection, it is possible to find the straight lines or *geodesics* in our five models with a minimal amount of effort. Though geodesics can be found by solving differential equations, we shall not do so. Rather, we establish the existence of one geodesic in the half-space model by means of what we call the retraction principle. Then we deduce the nature of all other geodesics by means of simple symmetry properties of the hyperbolic metrics. Here are the details. We learned this argument from Bill Thurston.

THEOREM 9.1 (THE RETRACTION PRINCIPLE). *Suppose that X is a Riemannian manifold, that $C : (a, b) \to X$ is an embedding of an interval (a, b) in X, and that there is a retraction $r : X \to \mathrm{image}(C)$ that is distance-reducing in the sense that, if one restricts the metric of X to $\mathrm{image}(C)$ and pulls this metric back via r to obtain a new metric on all of X, then at each point the pullback metric is less than or equal to the original metric on X. Then the image of C contains a shortest path (geodesic) between each pair of its points.*

The proof is left as an exercise. (Take an arbitrary path between two points of the image and show that the retraction of that path is at least as short as the original path. See Figure 9.)

THEOREM 9.2 (EXISTENCE OF A FUNDAMENTAL GEODESIC IN HYPERBOLIC SPACE). *In the half-space model of hyperbolic space, all vertical lines are geodesic. Such a line is the unique shortest path between any pair of points on it.*

PROOF. Let $C : (0, \infty) \to H$, where $C(t) = (1, x_2, \ldots, x_n, t) \in H$ and where the numbers $x_2, \ldots, x_n$ are fixed constants; that is, C is an arbitrary vertical line in H.

Define a retraction $r : H \to \mathrm{image}(C)$ by the formula

$$r(1, x_1', \ldots, x_n', t) = (1, x_1, \ldots, x_n, t).$$

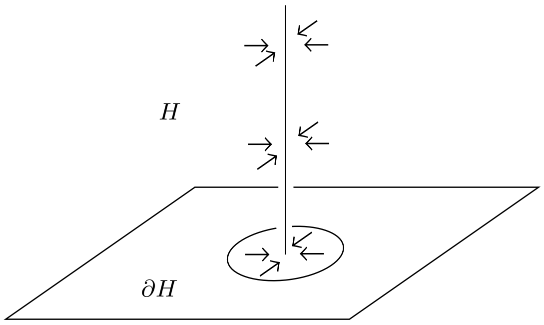


Figure 10. A fundamental hyperbolic geodesic and a distance-reducing retraction

See Figure 10. The original hyperbolic metric was $ds^2 = (dx_2^2 + \cdots + dx_{n+1}^2)/x_{n+1}^2$. The pullback metric is dx_{n+1}^2/x_{n+1}^2. Thus, by the retraction principle, the image of C contains a shortest path between each pair of its points.

It remains only to show that there is only one shortest path between any pair of points on the image of C. If one were to start with an arbitrary path between two points of the image of C that does not in fact stay in the image of C, then at some point the path is not vertical; hence the pullback metric is actually smaller than the original metric at that point since the original metric involves some dx_i^2 with $i \neq n+1$. Thus the retraction is actually strictly shorter than the original path. It is clear that there is only one shortest path between two points of the image that stays in the image. ☐

Theorem 9.3 (Classification of geodesics in H). *The geodesics in the half-space model H of hyperbolic space are precisely the vertical lines in H and the Euclidean metric semicircles whose endpoints lie in and intersect the boundary $\{(1, x_2, \ldots, x_n, 0)\}$ of hyperbolic space H orthogonally.*

Proof. See Figure 11 for the two types of geodesics. We need to make the following observations:

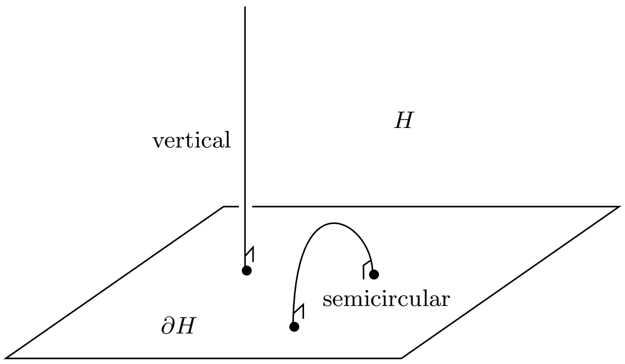


Figure 11. The two types of geodesics in H.

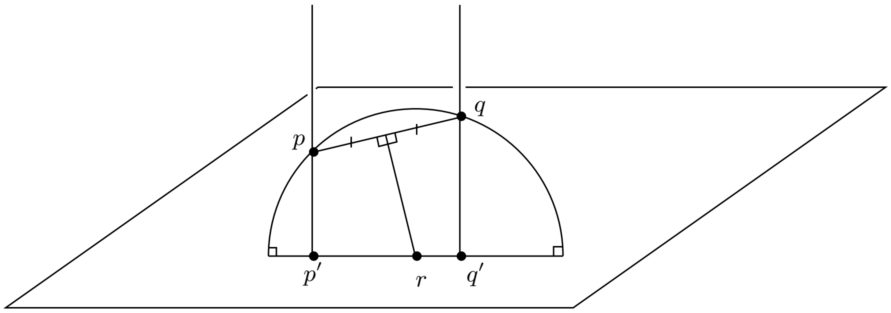


Figure 12. Finding the hyperbolic geodesic between points of H not on a vertical line.

(1) Euclidean isometries of H that take the boundary $\{(1, x_2, \ldots, x_n, 0)\}$ of H to itself are hyperbolic isometries of H. Similarly, the transformations of H that take $(1, x_1, \ldots, x_n, t)$ to $(1, rx_1, \ldots, rx_n, rt)$ with $r > 0$ are hyperbolic isometries. (Proof by direct, easy calculation.)

(2) Euclidean isometries of J are hyperbolic isometries of J. (Proof by direct, easy calculation.)

(3) If p and q are arbitrary points of H, and if p and q do not lie on a vertical line, then there is a unique boundary-orthogonal semicircle that contains p and q. Indeed, to find the center of the semicircle, take the Euclidean segment joining p and q and extend its Euclidean perpendicular bisector in the vertical plane containing p and q until it touches the boundary of H. See Figure 12.

(4) If C and C' are any two boundary-orthogonal semicircles in H, then there is a hyperbolic isometry taking C onto C'. (The proof is an easy application of (1) above.)

We now complete the proof of the theorem as follows. By the previous theorem and (1), all vertical lines in H are geodesic and hyperbolically equivalent, and each contains the unique shortest path between each pair of its points. Now map the vertical line in H with infinite endpoint $(1, 0, \ldots, 0)$ into J via the connecting stereographic projection. Then the image is a great semicircle. Rotate J, a hyperbolic isometry by (2), so that the center of the stereographic projection is not an infinite endpoint of the image. Return the rotated semicircle to H via stereographic projection. See Figure 13. By the theorems on stereographic projection, the image is a boundary-orthogonal semicircle in H. Since it is the image under a composition of isometries of a geodesic, this boundary-orthogonal semicircle is a geodesic. But all boundary-orthogonal semicircles in H are hyperbolically equivalent by (4) above. Hence each is a geodesic. Since there is a unique geodesic joining any two points of a vertical line, we find that there is a unique geodesic joining any two points of H (see (3)). This completes the proof of the theorem. $\square$

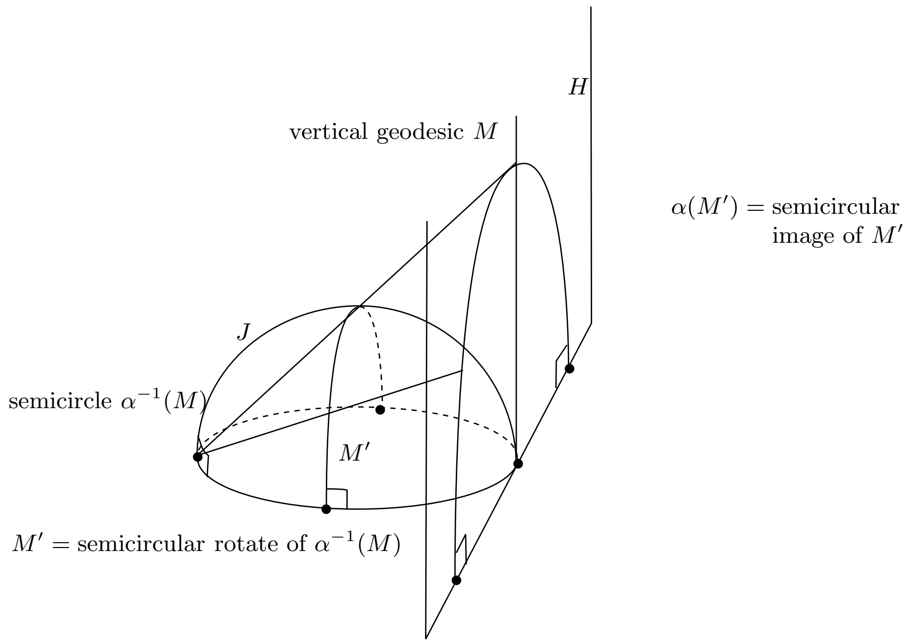


Figure 13. Geodesics in H.

By Theorems 8.1 and 8.2, the boundary-orthogonal semicircles in J correspond precisely to the boundary-orthogonal semicircles and vertical lines in H. Hence the geodesics in J are the boundary-orthogonal semicircles in J.

By Theorem 8.3, the boundary-orthogonal semicircles in J correspond to the diameters and boundary-orthogonal circular segments in I. Hence the diameters and boundary-orthogonal circular segments in I are the geodesics in I. See Figure 14.

The boundary-orthogonal semicircles in J clearly correspond under vertical projection to straight line segments in K. Hence the latter are the geodesics in K. See Figure 15.

The straight line segments in K clearly correspond under central projection from the origin to the intersections with L of two-dimensional vector subspaces of $\mathbb{R}^{n+1}$ with L; hence the latter are the geodesics of L. See Figure 16.

10. Isometries and Distances in the Hyperboloid Model

We begin our study of the isometries of hyperbolic space with the hyperboloid model L where all isometries, as we shall see, are restrictions of linear maps of $\mathbb{R}^{n+1}$.

DEFINITION. A *linear isometry* $f : L \to L$ of L is the restriction to L of a linear map $F : \mathbb{R}^{n+1} \to \mathbb{R}^{n+1}$ that preserves the hyperbolic inner product $*$ (that is,

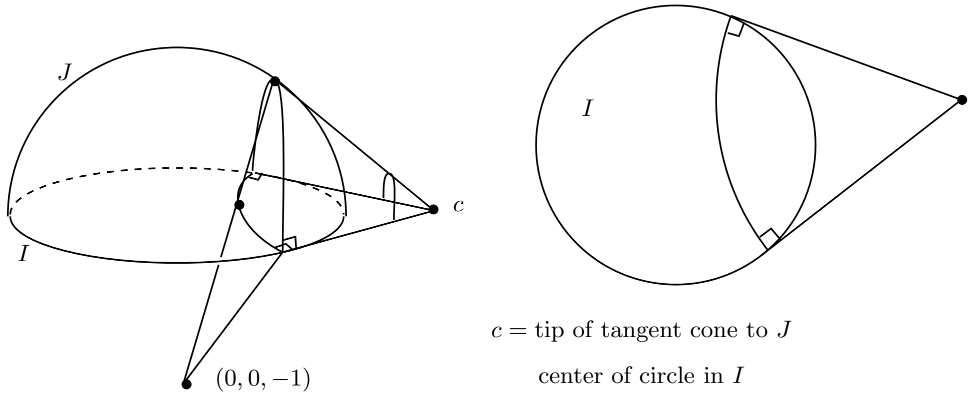


Figure 14. Geodesics in I and J and their stereographic relationship.

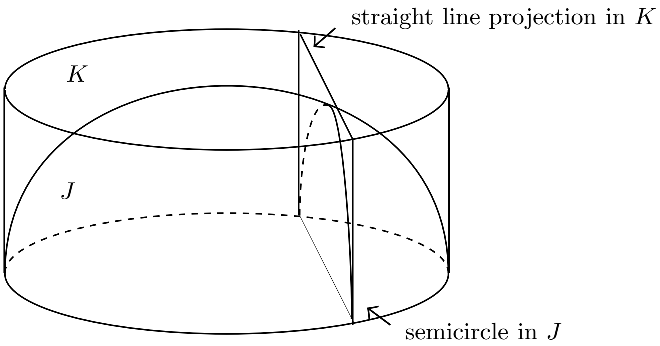


Figure 15. Geodesics in J and K.

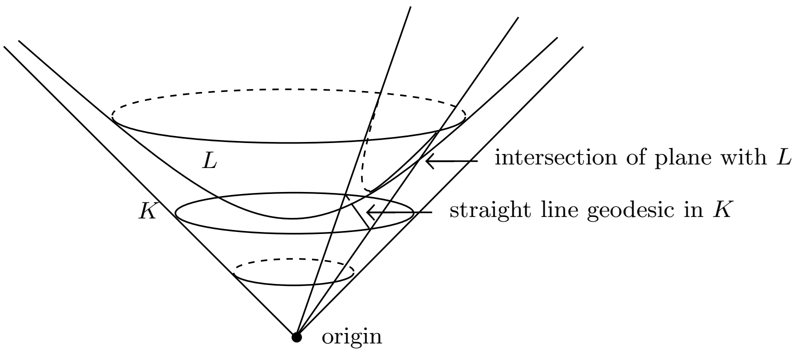


Figure 16. Geodesics in K and L.

for each pair v and w of vectors from $\mathbb{R}^{n+1}$, $Fv * Fw = v * w$) and that takes the upper sheet of the hyperboloid L into itself.

DEFINITION. A *Riemannian isometry* $f : L \to L$ of L is a diffeomorphism of L that preserves the Riemannian metric (that is, $f^*(ds^2) = ds^2$).

DEFINITION. A *topological isometry* $f : L \to L$ of L is a homeomorphism of L that preserves the Riemannian distance between each pair of points of L (that is, if d is the Riemannian distance function and if x and y are points of L, then $d(f(x), f(y)) = d(x, y)$).

THEOREM 10.1. *A square matrix M with columns $m_1, \ldots, m_n, m_{n+1}$ induces a linear isometry of L if and only if the following two conditions are satisfied.*

1. *For each pair of indices i and j, we have $m_i * m_j = e_i * e_j$, where $e_1, \ldots, e_n$, e_{n+1} is the standard basis for $\mathbb{R}^{n+1}$.*
2. *The last entry of the last column m_{n+1} is positive.*

Condition 1 *is satisfied if and only if M is invertible with $M^{-1} = JM^tJ$, where J is the diagonal matrix with diagonal entries $J_{11} = \cdots = J_{nn} = -J_{n+1,n+1} = 1$.*

PROOF. Let J denote the diagonal matrix with diagonal entries $J_{11} = \cdots = J_{nn} = -J_{n+1,n+1} = 1$. Then for each $x, y \in \mathbb{R}^{n+1}$, $x*y = x^tJy$. Thus $Mx*My = x^tM^tJMy$. Consequently, M preserves $*$ if and only if $M^tJM = J$; but the ij entry of M^tJM is $m_i * m_j$ while that of J is $e_i * e_j$. Thus M preserves $*$ if and only if condition 1 of the theorem is satisfied. Since J is invertible, condition 1 implies that M is also invertible and that it takes the hyperboloid of two sheets, of which L is the upper sheet, homeomorphically onto itself. Condition 2 is then just the statement that the image of e_{n+1} lies in L, that is, that M takes the upper sheet L of the hyperboloid onto itself.

Finally, the equality $M^{-1} = JM^tJ$ is clearly equivalent to the equality $M^tJM = J$ since $J^{-1} = J$. □

THEOREM 10.2. *A map $f : L \to L$ that satisfies any of the three definitions of isometry—linear, Riemannian, or topological—satisfies the other two as well.*

PROOF. We first prove the two easy implications, linear $\Rightarrow$ Riemannian $\Rightarrow$ topological, then connect the hyperbolic inner product $x * y$ with Riemannian distance $d(x, y)$ in preparation for the more difficult implication, topological $\Rightarrow$ linear.

Linear isometry $\Rightarrow$ Riemannian isometry: Let $F : \mathbb{R}^{n+1} \to \mathbb{R}^{n+1}$ be a linear map that preserves the hyperbolic inner product $*$ and takes the upper sheet L of the hyperboloid of two sheets into itself and thereby induces a linear isometry $f : L \to L$. The Riemannian metric ds^2 is at each point x of L simply a function of two variables that takes as input two tangent vectors v and w at x and delivers as output the hyperbolic inner product $v * w$. We calculate the pullback metric

$f^*(ds^2)$ in the following manner:

$$\begin{aligned} f^*(ds^2)(v,w) &= ds^2(Df(v), Df(w)) = ds^2(DF(v), DF(w)) \\ &= ds^2(F(v), F(w)) = F(v) * F(w) = v * w = ds^2(v,w). \end{aligned}$$

We conclude that $f^*(ds^2) = ds^2$, so that f is a Riemannian isometry.

Riemannian isometry ⇒ topological isometry: Riemannian distance is calculated by integrating the Riemannian metric. Since a Riemannian isometry preserves the integrand, it preserves the integral as well.

LEMMA. *If $a, b \in L$, then $a * b = -\cosh(d(a,b))$.*

PROOF. Let t denote the Riemannian distance $d(a,b)$ between a and b. One obtains this distance by integrating the Riemannian metric along the unique geodesic path joining a and b, or, since this integral is invariant under linear isometry, one can translate a and b to a standard position in L as follows and then perform the integration. Let m_1 be the unit tangent vector at a in the direction of the geodesic from a to b. Let $m_{n+1} = a$. By the Gram–Schmidt orthonormalization process from elementary linear algebra we may extend the orthonormal set $\{m_1, m_{n+1}\}$ to an orthonormal basis $m_1, \dots, m_n, m_{n+1}$ for $\mathbb{R}^{n+1}$; that is, $m_i * m_j = e_i * e_j$. By Theorem 10.1, the matrix M with columns $m_1, \dots, m_n, m_{n+1}$ gives a linear isometry of L as does its inverse M^{-1}. The inverse takes a to e_{n+1} and takes the two-dimensional subspace spanned by a and b to the space P spanned by e_1 and e_{n+1}. The intersection of P with L is one branch of a standard hyperbola that passes through $M^{-1}(a)$ and $M^{-1}(b)$ and is the unique hyperbolic geodesic through those two points. Since $M^{-1}(a) = (0, \dots, 0, 1)$ and since $t = d(a,b) = d(M^{-1}(a), M^{-1}(b))$, we may assume that $M^{-1}(b) = (\sinh(t), \dots, 0, \cosh(t))$. (See Section 4.) Thus we may calculate:

$$\begin{aligned} a * b &= M^{-1}(a) * M^{-1}(b) = (0, \dots, 0, 1) * (\sinh(t), \dots, 0, \cosh(t)) \\ &= -\cosh(t) = -\cosh(d(a,b)). \end{aligned}$$ □

Topological isometry ⇒ linear isometry: Let $f : L \to L$ denote a topological isometry. Let $v_1, \dots, v_n, v_{n+1}$ denote a basis for $\mathbb{R}^{n+1}$ such that each v_i lies in L. Let F denote the linear map that takes v_i to $f(v_i)$ for each i. We claim that F preserves $*$; to see this, write $e_i = \sum_j a_{ij} v_j$ and compute:

$$\begin{aligned} F(e_i) * F(e_j) &= \sum_{k,l} a_{ik} a_{jl} f(v_k) * f(v_l) \\ &= \sum_{k,l} a_{ik} a_{jl} (-\cosh(d(f(v_k), f(v_l)))) \\ &= \sum_{k,l} a_{ik} a_{jl} (-\cosh(d(v_k, v_l))) = e_i * e_j. \end{aligned}$$

Moreover, F agrees with f on L. To prove this, it suffices to replace f by $F^{-1} \circ f$, so that we can assume $f(v_i) = v_i$; then we must prove that $f = \mathrm{id}$, which we can do by showing $f(x) * e_i = x * e_i$ for each $x \in L$ and for each index i. Here is the calculation:

$$\begin{aligned} f(x) * e_i &= f(x) * \sum_j a_{ij} v_j = \sum_j a_{ij} (f(x) * f(v_j)) \\ &= \sum_j a_{ij} (-\cosh(d(f(x), f(v_j)))) \\ &= \sum_j a_{ij} (-\cosh(d(x, v_j))) = x * e_i. \end{aligned} \qquad \square$$

11. The Space at Infinity

It is apparent from all of our analytic models with the possible exception of the hyperboloid model L that there is a natural space at infinity. In the half-space model H it is the bounding plane of dimension $n-1$ that we compactify by adding one additional point; we visualize the additional point as residing at the top of the collection of vertical geodesics in H. In the disk model I, in the hemisphere model J, and in the Klein model K it is the bounding $(n-1)$-sphere. If we reinterpret the hyperboloid model as lying in projective space (each point of L is represented by the unique one-dimensional vector subspace of $\mathbb{R}^{n+1}$ that contains that point), then the space at infinity becomes apparent in that model as well: it consists of those lines that lie in the light cone $\{x \in \mathbb{R}^{n+1} : x * x = 0\}$. Furthermore, it is apparent that not only the models but also the unions of those models with their spaces at infinity correspond homeomorphically under our transformations connecting the models. That is, the space at infinity is a sphere of dimension $n-1$ and the union of the model with the space at infinity is a ball of dimension n.

Having analyzed the isometries of the hyperboloid model, we see that each isometry of L actually extends naturally not only to the space at infinity but to the entirety of projective n-space. That is, each linear mapping of $\mathbb{R}^{n+1}$ defines a continuous mapping of projective n-space P^n.

12. The Geometric Classification of Isometries

We recall from the previous sections that every isometry f of L extends to a linear homeomorphism F of $\mathbb{R}^{n+1}$, hence upon passage to projective space P^n induces a homeomorphism $f \cup f_\infty : L \cup \partial L \to L \cup \partial L$ of the ball that is the union of hyperbolic space L and its space ∂L at infinity. Every continuous map from a ball to itself has a fixed point by the Brouwer fixed point theorem. There is a very useful and beautiful geometric classification of the isometries of hyperbolic space that refers to the fixed points of this extended map. Our analysis of these maps requires that we be able to normalize them to some extent by moving given fixed points into a standard position. To that end we note that we have already shown how to move any point in L and nonzero tangent vector at that point so that the point is at e_{n+1} and the tangent points in the direction of e_1. As a consequence we can move any pair of points in $L \cup \partial L$ so that they lie in any given geodesic; and by conjugation we find that we may assume that any pair of fixed points of an isometry lies in a given geodesic. Indeed, let f be an isometry with fixed point x, let g be an isometry that takes x into a geodesic line L, and note that $g(x)$ is a fixed point of gfg^{-1}. Here are the three possible cases.

The elliptic case occurs when the extended map has a fixed point in L itself: conjugating by a linear isometry of L, we may assume that the isometry $f : L \to L$ fixes the point $e_{n+1} = (0, \dots, 0, 1)$. Let $F : \mathbb{R}^{n+1} \to \mathbb{R}^{n+1}$ be the linear extension of f. The representing matrix M has as last column m_{n+1} the vector e_{n+1}. The remaining columns must be $*$-orthogonal to m_{n+1}, hence Euclidean or $\cdot$-orthogonal to e_{n+1}. On the orthogonal complement of e_{n+1}, the hyperbolic and the Euclidean inner products coincide. Hence the remaining columns form not only a hyperbolic orthonormal basis but also a Euclidean orthonormal basis. We conclude that the matrix M defining F is actually Euclidean orthogonal. We call such a transformation of hyperbolic space *elliptic*.

The hyperbolic case occurs when the extended map has no fixed point in L itself but has two fixed points at infinity: we examine this transformation in the half-space model H for hyperbolic space. We ignore the initial constant coordinate 1 in H and identify H with the half-space $\{x = (x_1, \dots, x_n) \in \mathbb{R}^n : x_n > 0\}$. Conjugating by an isometry, we may assume that the fixed points of the map f of $H \cup \partial H$ are the infinite endpoints of the hyperbolic geodesic $(0, \dots, t)$, where $t > 0$. Let $(0, \dots, k)$ denote the image under f of $(0, \dots, 1)$. Then $(1/k) \cdot f$ is an isometry that fixes every point of the hyperbolic geodesic $(0, \dots, t)$. By the previous paragraph, the transformation $(1/k) \cdot f$ is an orthogonal transformation O. It follows easily that $f(x) = k\,O(x)$, the composite of a Euclidean orthogonal transformation O, which preserves the boundary plane at infinity and which is simultaneously a hyperbolic isometry, with the hyperbolic translation $x \mapsto kx$ along the geodesic $(0, \dots, t)$. Such a transformation is called *hyperbolic* or *loxodromic*. The invariant geodesic $(0, \dots, t)$ is called the *axis* of the hyperbolic transformation. See Figure 17. Often one preserves the name *hyperbolic* for the case where the orthogonal transformation is trivial and the name *loxodromic* for the case where the orthogonal transformation is nontrivial.

The parabolic case occurs when the extended map has only one fixed point and that fixed point is at infinity: we examine this transformation in the half-space model H for hyperbolic space. We may assume that the fixed point of the map f of $H \cup \partial H$ is the upper infinite endpoint of the hyperbolic geodesic $(0, \dots, t)$, where $t > 0$. The transformation $g : x \mapsto f(x) - f((0, \dots, 0))$ fixes both ends of the same geodesic. Hence g may be written as a composite $x \mapsto k\,O(x)$ where $k > 0$ and O have the significance described in the previous paragraph. Thus $f(x) = k\,O(x) + v$, where $k > 0$, O is Euclidean orthogonal preserving the boundary plane of H, and $v = f((0, \dots, 0))$ is a constant vector. We claim that

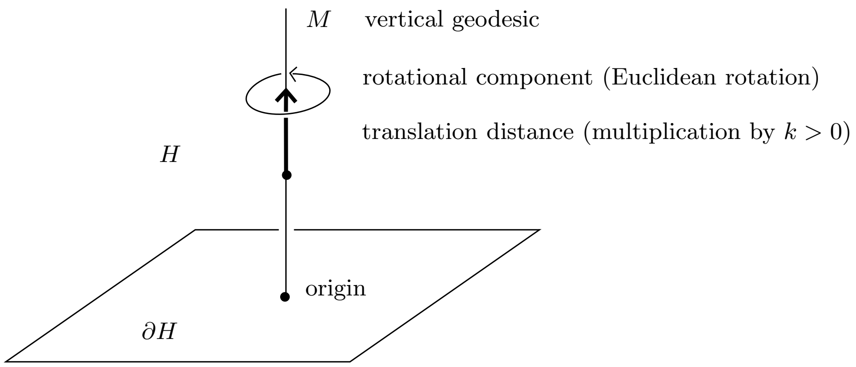


Figure 17. Hyperbolic or loxodromic isometry.

$k = 1$ so that f is a Euclidean isometry preserving the boundary plane of H; such a map, without fixed points in the boundary plane, is called *parabolic*. If $k \neq 1$, we claim that f has another fixed point. We find such a fixed point in the following way. The fixed point will be a solution of the equation $(I - kO)x = v$. The eigenvalues of $I - kO$ have the form $1 - k\lambda$, where λ is an eigenvalue of O. Since O is orthogonal, its eigenvalues have absolute value 1. Hence if $k \neq 1$, then $I - kO$ is invertible, and the equation $(I - kO)x = v$ does indeed have a solution.

13. Curious Facts about Hyperbolic Space

FACT 1. *In the three conformal models for hyperbolic space, hyperbolic spheres are also Euclidean spheres; however, Euclidean and hyperbolic sphere centers need not coincide.*

PROOF. We work in the hemisphere model J for hyperbolic space and consider the point $p = (0, \ldots, 0, 1) \in J$. The Riemannian metric ds_J^2 is clearly rotationally symmetric around p so that a hyperbolic sphere centered at p is a Euclidean sphere.

We project such a sphere from J into the half-space model H for hyperbolic space via stereographic projection. See Figure 18. Since stereographic projection takes spheres that miss the projection point to spheres in H, we see that there is one point of H, namely the image of p, about which hyperbolic spheres are Euclidean spheres. But this point can be taken to any other point of H by a composition of Euclidean translations and Euclidean similarities that are hyperbolic isometries as well. Since these Euclidean transformations preserve both the class of hyperbolic spheres and the class of Euclidean spheres, we see that the hyperbolic spheres centered at each point of H are also Euclidean spheres.

We project this entire class of spheres back into J and from thence into I by stereographic projections that preserve this class of Euclidean (and hyperbolic)

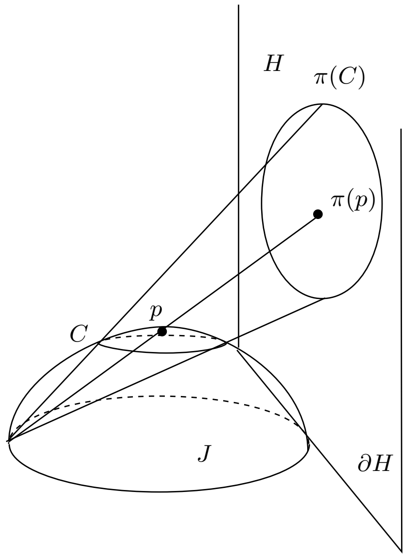


Figure 18. The projection of a sphere from J to H. C is both a Euclidean and a hyperbolic sphere; $\pi(p)$ is the hyperbolic center of the circle $\pi(C)$.

spheres. We conclude that all hyperbolic spheres in these three models are also Euclidean spheres, and conversely.

Finally, we give a geometric construction for the hyperbolic center of a Euclidean sphere S in the half-space model H. See Figure 19. Draw the vertical geodesic line M through the center of S until it meets the plane at infinity at some point p. Draw a tangent line to S from p meeting S at a tangency point q. Draw the circle C through q that is centered at p and lies in the same plane as M. The circle C then meets the line M at the hyperbolic center of S (proof, an exercise for the reader). Note that this center is not the Euclidean center of S. $\square$

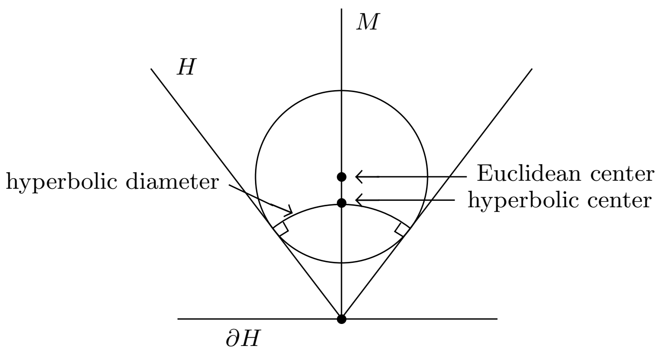


Figure 19. Constructing the hyperbolic center of a circle

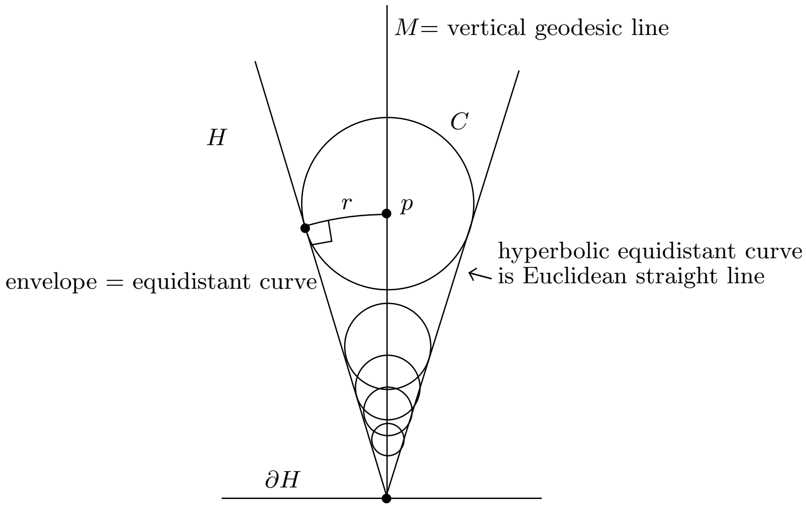


Figure 20. Equidistant curves in H.

Fact 2. *In the hyperbolic plane, the two curves at distance r on either side of a straight line are not straight.*

Proof. We can use the preceding result to analyze the curves equidistant from a hyperbolic geodesic in the hyperbolic plane. We work in the half-space model $H \subset \mathbb{R}^2$ of the two-dimensional hyperbolic plane and take as geodesic line the vertical line M that passes through the origin of $\mathbb{R}^2$. Put a hyperbolic circle C of hyperbolic radius r about a point p of M. Then we obtain the set of all such circles centered at points of M by multiplying C by all possible positive scalars. The union of these spheres $t \cdot C$ is a cone, or angle, D of which the origin is the vertex and whose central axis is M. The envelope or boundary of this cone or angle is a pair of Euclidean straight lines, the very equidistant lines in which we are interested. See Figure 20. Since these straight lines are not vertical, they are not hyperbolic straight lines. $\square$

Fact 3. *Triangles in hyperbolic space have angle sum less than π; in fact, the area of a triangle with angles α, β, and γ is $\pi - \alpha - \beta - \gamma$ (the Gauss–Bonnet theorem). Given three angles α, β, and γ whose sum is less than π, there is one and only one triangle up to congruence having those angles. Consequently, there are no nontrivial similarities of hyperbolic space.*

Proof. Any triangle in hyperbolic space lies in a two-dimensional hyperbolic plane. Hence we may work in the half-space model H for the hyperbolic plane. Assume that we are given a triangle $\Delta = pqr$ with angles α, β, and γ. We may arrange via an isometry of hyperbolic space that the side pq lies in the unit circle. Then by a hyperbolic isometry of hyperbolic space that has the unit circle as its invariant axis and translates along the unit circle we may arrange that the side pr points vertically upward. The resulting picture is in Figure 21.

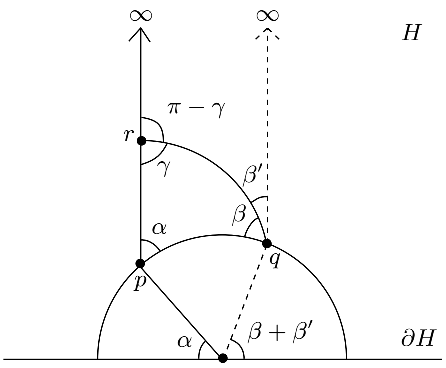


Figure 21. The Gauss–Bonnet theorem.

We note that the triangle $\Delta = pqr$ is the difference of two ideal triangles $pq\infty$ and $rq\infty$. ("Ideal" means that at least one vertex is at infinity.) We first prove the Gauss–Bonnet theorem for such an ideal triangle, then deduce the desired formula by taking a difference.

The element of area is $dA = dx\,dy/y^2$. It is easy to verify that the area of $pq\infty$ is therefore

$$\int_{pq\infty} dA = \int_{x=\cos(\beta+\beta')}^{x=\cos(\pi-\alpha)} dx \int_{y=\sqrt{1-x^2}}^{y=\infty} \frac{dy}{y^2}.$$

Straightforward evaluation leads to the value $\pi - \alpha - \beta - \beta'$ for the integral. Similar evaluation gives the value $\pi - (\pi - \gamma) - \beta'$ for the area of $rq\infty$. The difference of these two values is $\pi - \alpha - \beta - \gamma$ as claimed. This proves the Gauss–Bonnet theorem.

We now construct a triangle with given angles. Suppose therefore that three angles α, β, and γ are given whose sum is less than π, a necessary restriction in view of the Gauss–Bonnet theorem. Pick a model of the hyperbolic plane, say the disk model I. Pick a pair Q and R of geodesic rays (radii) from the origin p meeting at the Euclidean (= hyperbolic) angle of α. See Figure 22. Note that any pair of geodesic rays meeting at angle α is congruent to this pair. Pick points q and r on these rays and consider the triangle pqr. Let β' denote the angle at q and let γ' denote the angle at r. Let A' denote the area of the triangle pqr. We will complete the construction by showing that there is a unique choice for q on Q and for r on R such that $\beta' = \beta$ and $\gamma' = \gamma$. The argument will be variational.

We first consider the effect of fixing a value of q and letting r vary from ∞ to p along R. At ∞, the angle γ' is 0. At (near) p the angle γ' is (almost) $\pi - \alpha$. As r moves inward toward p along R, both β' and A' clearly decrease monotonically. Hence, by the Gauss–Bonnet theorem, $\gamma' = \pi - \alpha - \beta' - A'$ increases monotonically. In particular, there is a unique point $r(q)$ at which

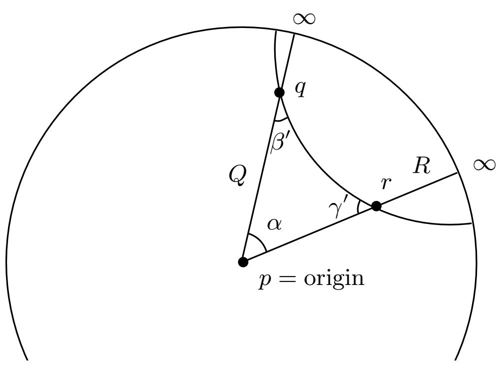


Figure 22. Constructing a triangle with angles α, β, and γ, with $\alpha+\beta+\gamma<\pi$

$\gamma' = \gamma$. Now fix q, fix r at $r(q)$, and move inward along Q from q to a point q'. Note that the angle of $pq'r$ at r is smaller than the angle γ, which is the angle of pqr at r. We conclude that $r(q')$ must be closer to p than is $r(q)$. That is, as q moves inward toward p, so also does $r(q)$. We conclude that the areas of the triangles $pqr(q)$ decrease monotonically as q moves inward along Q toward p.

We are ready for the final variational argument. We work with the triangles $pqr(q)$. We start with q at ∞ and note that the area A' is equal to $\pi-\alpha-0-\gamma > \pi-\alpha-\beta-\gamma$. We move q inward along Q, and consequently move $r(q)$ inward along R, until q reaches p and $A' = 0 < \pi-\alpha-\beta-\gamma$. As noted in the previous paragraph, the area A' decreases monotonically. Hence there is a unique value of q at which the area is $\pi-\alpha-\beta-\gamma$. At that value of q the angle β' must equal β by the Gauss–Bonnet theorem. □

FACT 4. *If $\Delta = pqr$ is a triangle in hyperbolic space, and if x is a point of the side pq, then there is a point $y \in pr \cup qr$ such that the hyperbolic distance $d(x,y)$ is less than $\ln(1+\sqrt{2})$; that is, triangles in hyperbolic space are uniformly thin.*

PROOF. We need two observations. First, if P and Q are two vertical geodesics in the half-space model H for hyperbolic space, and if a point p moves monotonically downward along P, then the distance $d(p,Q)$ increases monotonically to infinity. See Figure 23. Second, if p and q are two points on the same boundary-orthogonal semicircle (geodesic) in H, say on the unit circle with coordinates $p = (\cos(\phi), \sin(\phi))$ and $q = (\cos(\theta), \sin(\theta))$ with $\theta > \phi$, then the hyperbolic distance between the two is given by the formula

$$d(p,q) = \int_\phi^\theta \frac{d\psi}{\sin(\psi)} = \ln \frac{\sin(\psi)}{1+\cos(\psi)}\bigg|_\phi^\theta.$$

See Figure 24. Actually, the radius of the semicircle is irrelevant because scaling is a hyperbolic isometry. Only the beginning and ending angles are important.

We are now ready for the proof that triangles are thin. Let $\Delta = pqr$ denote a triangle in the hyperbolic plane. We view Δ in the half-space model of the

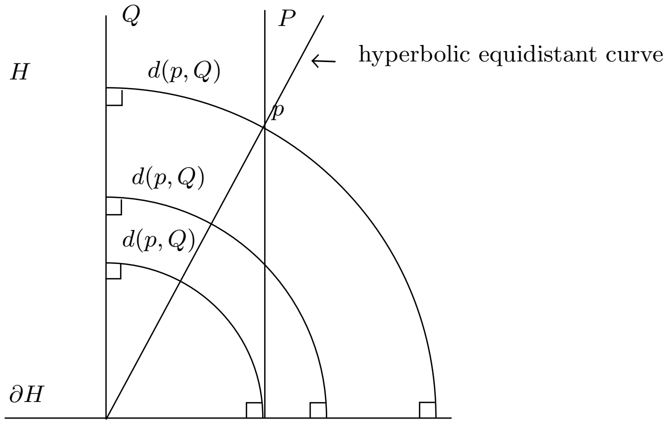
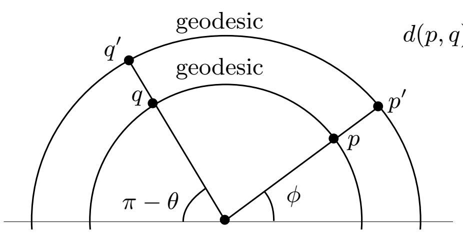


Figure 23. The monotonicity of $d(p,Q)$.

hyperbolic plane. We may assume that the side pq lies in the unit circle with p to the left of q, and we may assume that the side pr is vertical with r above p. We assume a point $x \in pq$ given. See Figure 25. We want to find an upper bound for the distance $d(x, pr \cup qr)$. The following operations simply expand the triangle Δ and hence increase the distance that we want to bound above. First we may move r upward until it moves to ∞. We may then slide p leftward along the unit circle until it meets infinity at $p' = -1$. We may then slide q rightward along the unit circle until it meets infinity at $q' = 1$. We now have an ideal triangle $p'q'\infty$ with $x \in p'q'$. See Figure 26. The pair of sides $p'q'$ and $p'\infty$ are congruent as a pair to a pair of vertical geodesics (simply move p' to ∞ by an isometry of H). Hence as we move x toward q', the distance $d(x, p'\infty)$ increases monotonically. Similarly, as we move x toward p', the distance $d(x, q'\infty)$ increases monotonically. We conclude that the maximum distance to $p'\infty \cup q'\infty$ is realized when x is at the topmost point of the unit circle. The distances to the two vertical geodesics $p'\infty$ and $q'\infty$ are then equal and the shortest path is realized by a boundary-orthogonal semicircle that passes through x and meets, say, $p'\infty$ orthogonally (if it did not meet orthogonally, then a shortcut near the vertical geodesic would reduce the length of the path). It is clear from the geometry that this shortest



$$d(p,q) = d(p',q') = \int_{\text{geodesic}} ds = \int_{\phi}^{\theta} \frac{d\psi}{\sin\psi}$$

Figure 24. The formula for $d(p,q)$.

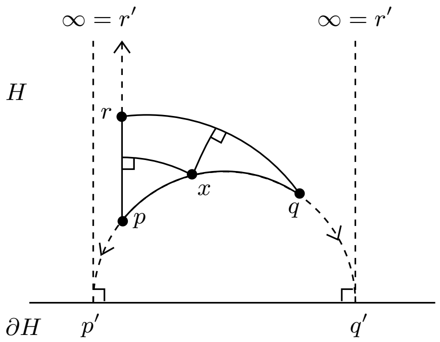


Figure 25. Triangles are thin.

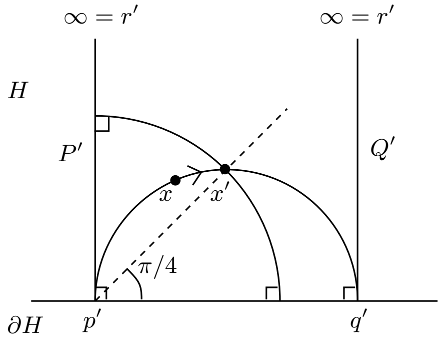


Figure 26. The ideal triangle $p'q'\infty$.

path travels through the angle interval $[\pi/4, \pi/2]$ in going from x to the vertical geodesic $p'\infty$. Hence, by our calculation above, the distance between the point and the opposite sides is

$$\ln \frac{\sin(\pi/2)}{1+\cos(\pi/2)} - \ln \frac{\sin(\pi/4)}{1+\cos(\pi/4)} = \ln\bigl(1+\sqrt{2}\,\bigr) = 0.88\ldots.$$

We conclude that triangles are uniformly thin, as claimed. $\square$

Fact 5. *For a circular disk in the hyperbolic plane, the ratio of area to circumference is less than* 1 *and approaches* 1 *as the radius approaches infinity. That is, almost the entire area of the disk lies very close to the circular edge of the disk. Both area and circumference are exponential functions of hyperbolic radius.*

Proof. We do our calculations in the disk model I of the hyperbolic plane. The Riemannian metric is, as we recall,

$$ds_I^2 = 4(dx_1^2 + \cdots + dx_n^2)/(1 - x_1^2 - \cdots x_n^2)^2.$$

We are considering the case $n = 2$. Using polar coordinates (see Section 6) we can easily compute the distance element along a radial arc, namely

$$ds = 2\,\frac{dr}{1-r^2},$$

while the area element is

$$dA = \frac{4}{(1-r^2)^2} r\,dr\,d\theta.$$

We fix a Euclidean radius R with associated circular disk centered at the origin in I and calculate the hyperbolic radius ρ, area A, and circumference C (see

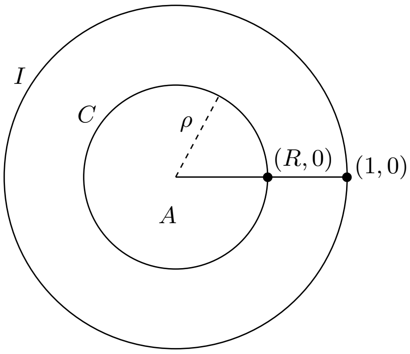


Figure 27. The hyperbolic radius ρ, the area A, and the circumference C.

Figure 27):

$$\begin{aligned} \rho &= \int_0^R 2\,\frac{dr}{1-r^2} = \ln\frac{1+R}{1-R}; \\ A &= \int_{\theta=0}^{2\pi}\int_{r=0}^{R} \frac{4}{(1-r^2)^2}\, r\,dr\,d\theta = \frac{4\pi R^2}{1-R^2}; \\ C &= \int_{\theta=0}^{2\pi} \frac{2R}{1-R^2}\, d\theta = \frac{4\pi R}{1-R^2}. \end{aligned}$$

Therefore

$$\begin{aligned} R &= \frac{e^\rho - 1}{e^\rho + 1} = \frac{\cosh\rho - 1}{\sinh\rho}; \\ A &= 2\pi(\cosh\rho - 1) = 2\pi\Big(\frac{\rho^2}{2!} + \frac{\rho^4}{4!} + \cdots\Big) \approx \pi\rho^2 \text{ for small } \rho; \\ C &= 2\pi\sinh\rho = 2\pi\Big(\rho + \frac{\rho^3}{3!} + \frac{\rho^5}{5!} + \cdots\Big) \approx 2\pi\rho \text{ for small } \rho. \end{aligned}$$

Note that the formulas are approximately the Euclidean formulas for small ρ. This is apparent in the half-space model if one works near a point at unit Euclidean distance above the bounding plane; for at such a point the Euclidean and hyperbolic metrics coincide, both for areas and lengths. □

FACT 6. *In the half-space model H of hyperbolic space, if S is a sphere centered at a point at infinity $x \in \partial H$, then inversion in the sphere S induces a hyperbolic isometry of H that interchanges the inside and outside of S in H.*

PROOF. Consider a Euclidean sphere S centered at a point p of the bounding plane at infinity. Let x be an arbitrary point of H, and let M be the Euclidean straight line through p and x. There is a unique point $x' \in M \cap H$ on the opposite side of S such that the two Euclidean straight line segments $x(S \cap M)$ and $x'(S \cap M)$ have the same *hyperbolic* length. See Figure 28. The points x and x' are said to be *mirror images* of one another with respect to S. We claim that

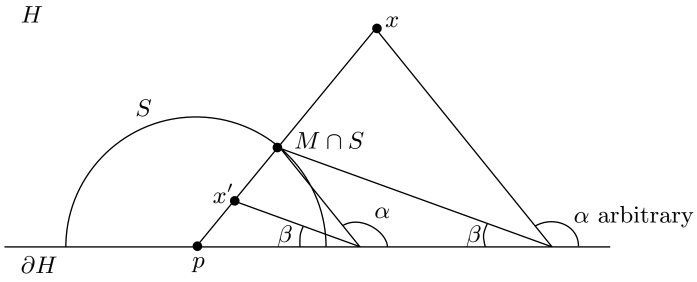


Figure 28. Inversion in S.

the map of H that interchanges all of the inverse pairs x and x' is a hyperbolic isometry. We call this map *inversion in* S.

Note that all such spheres S are congruent via hyperbolic isometries that are Euclidean similarities. Inversion is clearly invariant under such isometries. We shall make use of this fact both in giving formulas for inversion and in proving that inversion is a hyperbolic isometry.

Though our proof will make no use of formulas, we nevertheless describe inversion in S by means of a formula. We lose no generality in assuming that S is centered at the origin of Euclidean space. If S has radius r and if x has length t, then multiplication of H by the positive constant r/t is a hyperbolic isometry that takes M onto itself and takes x to the point $M \cap S$. A second multiplication by r/t takes the Euclidean segment $x(M \cap S)$ to a segment of the same hyperbolic length on the opposite side of S, hence takes $M \cap S$ to x'. That is, $x' = (r/t)^2 x$.

We now prove that inversion is a hyperbolic isometry. For that purpose we consider the hemisphere model J for hyperbolic space. Consider the n-dimensional plane $P = \{x \in \mathbb{R}^{n+1} : x_1 = 0\}$ through the origin of $\mathbb{R}^{n+1}$ that is parallel to the half-space model $H = \{x \in \mathbb{R}^{n+1} : x_1 = 1\}$ of hyperbolic space. See Figure 29. This plane intersects the hemisphere model J in one half of a sphere of dimension $n-1$, which we denote by S'. The entire model J is filled by circular segments that begin at the point $(-1, 0, \dots, 0)$, end at the point $(1, 0, \dots, 0)$, and intersect S' at right angles. The hyperbolic metric ds_J^2 is clearly symmetric with respect to the plane P and its intersection S' with J. Euclidean reflection in that plane therefore induces a hyperbolic isometry of J that takes a point on any of our circular segments to the point on the same circular segment but on the opposite side of S'. The symmetry of the hyperbolic metric clearly implies that the hyperbolic length of the two corresponding circular segments joining the point and its image to S' have the same hyperbolic length.

Now map J to H by stereographic projection. Then S' goes to one of our admissible spheres $S \cap H$ and our circular segments go to the family of lines M through the origin. We see therefore that our hyperbolic reflection isometry

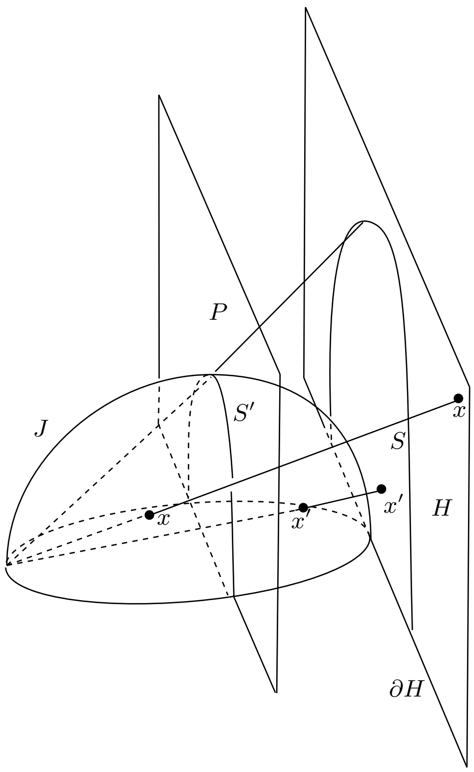


Figure 29. Inversion is a hyperbolic isometry.

of J goes precisely to our inversion of H in the sphere S. This completes the proof. □

14. The Sixth Model

We now turn to yet another model of the hyperbolic plane. This sixth model is only a combinatorial approximation to the half-space model, rather than a model in the sense of the other five. Consider the (infinite) family of "squares" sitting in the half-plane model, part of which is shown in Figure 30. This family is the image of the unit square, with vertical and horizontal sides and whose lower left corner is at $(0, 1)$, under the maps $p \mapsto 2^j\big(p + (k, 0)\big)$ with $(j, k) \in \mathbb{Z}^2$. Since horizontal translation and homotheties are hyperbolic isometries in H, each "square" is isometric to every other square. (We've called them squares even though in the hyperbolic metric they bear no resemblance to squares.)

Moving around in this family of squares is essentially like moving around in the hyperbolic plane. The advantage of the squares is that you can see combinatorially many of the aspects of hyperbolic space.

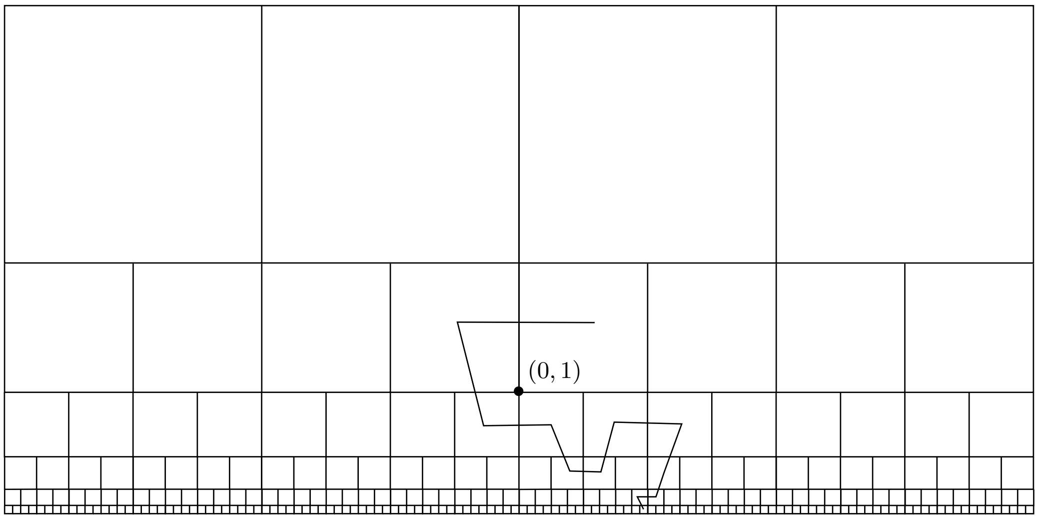


Figure 30. The sixth model and a random walk in the dual graph.

For example, note that a random walk on the dual graph will tend almost surely to infinity: from inside any square, the probability of exiting downwards is twice as great as the probability of exiting upwards.

Let p and q be vertices of the dual graph. Then one geodesic from p to q is gotten by taking a a path as in Figure 31, which rises initially straight upwards, goes horizontally a length at most 5, and then descends to q. More generally, let γ be a geodesic from p to q. Then there exists a geodesic δ from p to q that rises initially straight upwards, goes horizontally a length at most 5, and then descends to q in such a way that the distance from any vertex of γ is at most one from some vertex of δ, and vice versa.

Another aspect of the hyperbolic plane that can be illustrated in this model is the "thin triangles" property. Given that we understand what geodesics look like from the previous paragraph, we first consider only a triangle with geodesic sides as in Figure 32.

The combinatorial lengths of the bottom two horizontal arcs are at most 5. Since the combinatorial length divides by approximately 2 as you ascend one level up, it follows that the combinatorial vertical distance from the middle horizontal arc to the top horizontal arc is at most 3. Hence it follows that every point on one side of the triangle is within distance at most 8 of the union of the two opposite sides of the triangle. Thus triangles in this model are said to be 8-*thin*. (In hyperbolic space, we saw that triangles are $\log(1+\sqrt{2})$-thin in this sense.)

A consequence of the thin triangles property in a metric space is the exponential divergence of geodesics. Consider once again the half-space model H. Recall that a hyperbolic sphere (the set of points at a fixed distance from a point) is in fact also a Euclidean sphere.

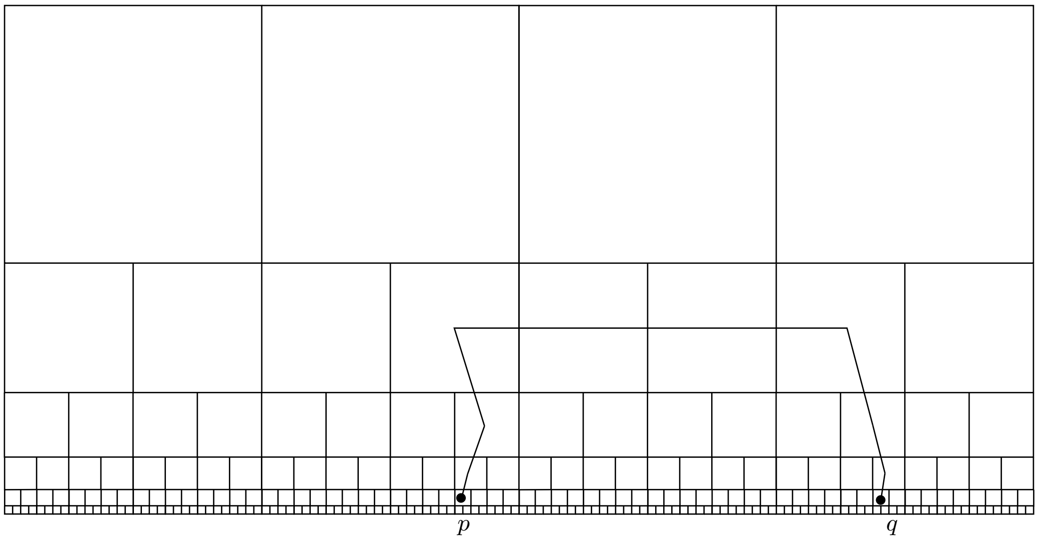


Figure 31. A geodesic in the dual graph.

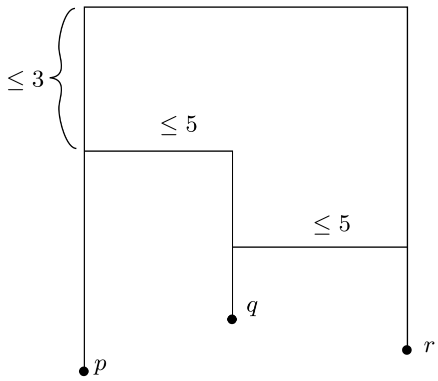


Figure 32. A triangle with geodesic sides.

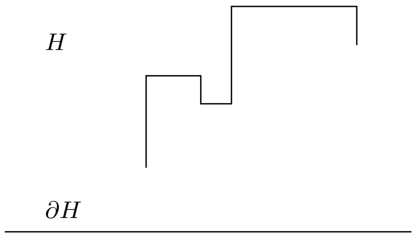


Figure 33. A taxicab path. Integrate with respect to $ds = \sqrt{dx^2 + dy^2}/y$ to find the taxicab distance.

As in the proof of Fact 5 in Section 13, the area of a disk of radius r_h is

$$A_h = 2\pi(\cosh(r_h) - 1),$$

whereas the length of the boundary of a disk of radius r_h is $2\pi\sinh(r_h)$. For large r_h, these are both quite close to πe^{r_h}, so in particular we see that the circumference is exponential in the radius. This phenomenon will be known as the *exponential explosion*, and is true in any metric space satisfying the thin triangles condition.

Before we go on, we leave the reader with two exercises.

EXERCISE 1. Take a "taxicab" metric on H^2 in which the allowed paths are polygonal paths that have horizontal or vertical edges. See Figure 33. Analyze the geodesics in this new metric, and prove the thin triangles property.

EXERCISE 2. Generalize the previous exercise to H^3: let the allowed paths be polygonal paths that are vertical (in the z-direction) or horizontal (lie parallel to the xy-plane). Define the length of a horizontal line segment to be $\max\{\Delta x, \Delta y\}/z$.

15. Why Study Hyperbolic Geometry?

Hyperbolic geometry and its geometric insights have application in diverse areas of mathematics. The next three sections informally introduce some of those applications. The material is intended to be skimmed, since the reader may be unfamiliar with some of the prerequisite background material.

Hyperbolic geometry arises in three main areas:

(i) Complex variables and conformal mappings. In fact, work on automorphic functions was Poincaré's original motivation for defining hyperbolic space; see the quotation on page 63.
(ii) Topology (of three-manifolds in particular). More on this later regarding Thurston's surprising geometrization conjecture.
(iii) Group theory, in particular combinatorial group theory à la Gromov.

Historically, hyperbolic geometry lies at the center of a triangle around which revolve these three topics. See Figure 34. By using hard theorems in one domain and hard connections between domains, one can prove surprising results.

One such example is the Mostow rigidity theorem [1973], which we quote here in its simplest, lowest-dimensional form, for oriented, connected, compact three-dimensional manifolds. To state the theorem we need a definition: a *hyperbolic structure* on an n-dimensional manifold M is a Riemannian metric on M such that the universal covering space of M is isometric to hyperbolic space H^n.

THEOREM 15.1. *Suppose the three-manifolds M_1 and M_2 having hyperbolic structures are homotopy equivalent. Then M_1 and M_2 are in fact isometric.*

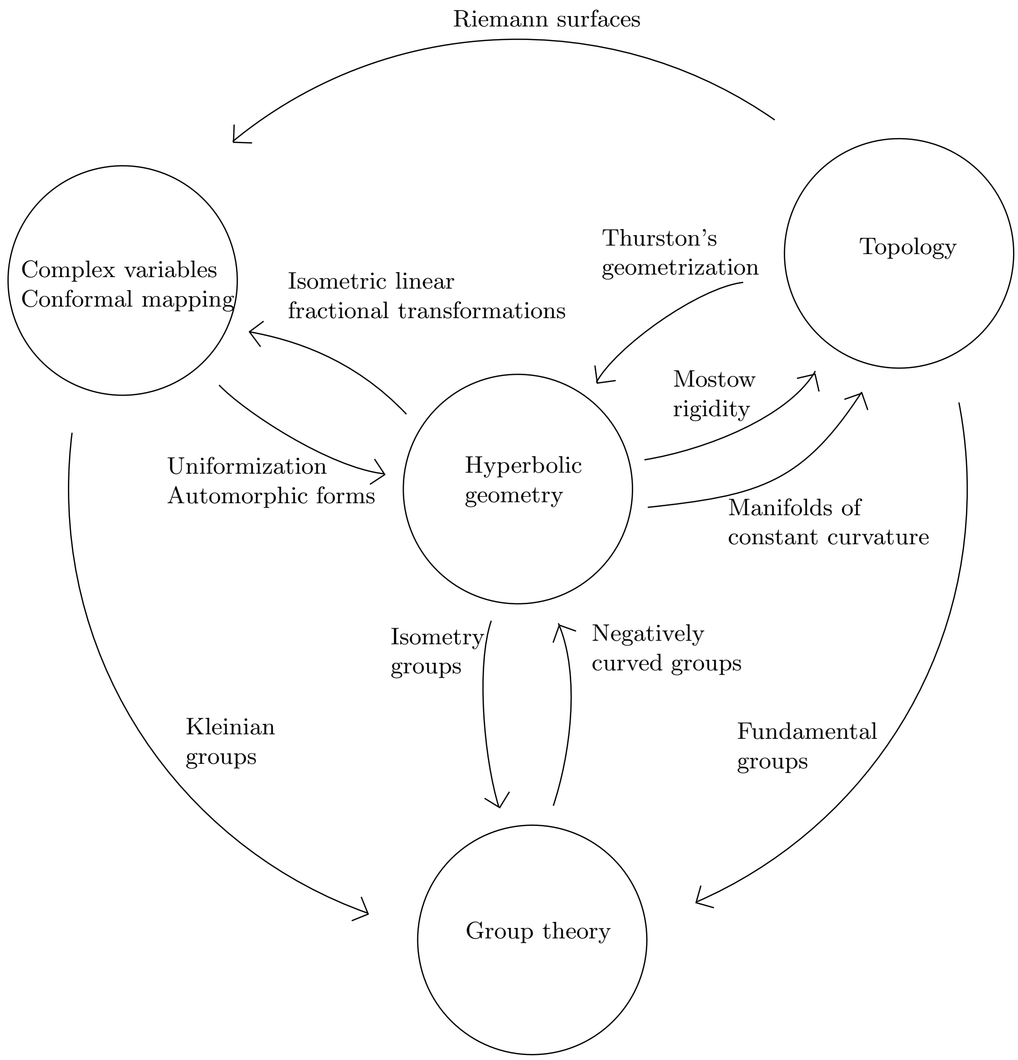


Figure 34. Connections between hyperbolic geometry and the three areas.

Here is a very abbreviated outline of the proof. There is an approximate metric correspondence between the universal covering space of M_1, as realized by hyperbolic space H^3, and the fundamental group $\pi_1(M_1)$, as realized geometrically by its Cayley group graph. Likewise the universal cover of M_2 and the Cayley graph of $\pi_1(M_2)$ are approximately isometric. The isomorphism between the fundamental groups $\pi_1(M_1)$ and $\pi_1(M_2)$ relates the two Cayley graphs metrically, and consequently gives an approximate geometric mapping from H^3, viewed as one universal cover, back to H^3 itself, viewed as the other universal cover. This mapping relates the actions of the two groups on their universal covers with enough precision that it matches their actions on the two-sphere at infinity homeomorphically. It remains to prove that this correspondence at infinity is not only homeomorphic but exactly conformal; the proof uses the theory of quasiconformal mappings and ergodicity of the group actions on the two-sphere at infinity. When the most general theorem is proved, the details fill a book.

A corollary to this result is that the hyperbolic structure on a three-manifold is a *topological invariant*. Consequently, so also are hyperbolic volume, lengths of geodesics, etc.

The proof of Mostow Rigidity suggests that many important properties of hyperbolic geometry are retained by spaces that only *approximate* hyperbolic geometry. The case in point involves the Cayley graphs of the fundamental groups, which are only one-dimensional and are not even manifolds, yet give a good approximation to three-dimensional hyperbolic geometry in the large. We shall return to this point when we describe Gromov's word-hyperbolic groups. But first we make precise the notion of spaces that are metrically comparable in the large, or *quasi-isometric*. We start by recalling some definitions.

A *group action* of a group G on a space X is a map $\alpha : G \times X \to X$, denoted $\alpha(g, x) = g(x)$, such that $1(x) = x$ for all $x \in X$, and $(g_1 g_2)(x) = g_1(g_2(x))$ for all $g_1, g_2 \in G$ and $x \in X$. In other words, α is a homomorphism from G into $\mathrm{Homeo}(X)$.

A *geometry* is a path metric space in which metric balls are compact.

A *geometric action* of a group G on a geometry X is a group action that satisfies the following conditions:

(i) G acts by isometries of X.
(ii) The action is properly discontinuous; that is, for every compact set $Y \subset X$, the set $\{g \in G : g(Y) \cap Y \neq \varnothing\}$ has finite cardinality.
(iii) The quotient $X/G = \{xG : x \in X\}$ is compact in the quotient topology.

Two geometries X_1, X_2 are *quasi-isometric* if there exist (not necessarily continuous) functions $R : X_1 \to X_2$, $S : X_2 \to X_1$ and a positive real number M such that

(i) $S \circ R : X_1 \to X_1$ and $R \circ S : X_2 \to X_2$ are within M of the identities.
(ii) For all $x_1, y_1 \in X_1$, $d(R(x_1), R(y_1)) \leq M d(x_1, y_1) + M$ and likewise for X_2.

Theorem 15.2 (Quasi-isometry Theorem). *If a group G acts geometrically on geometries X_1 and X_2, then X_1 and X_2 are quasi-isometric.*

Here are some exercises to challenge your understanding of these concepts.

Exercise 3. Let $G = \mathbb{Z}^2$, let X_1 be the Cayley graph of G with standard generators, and let $X_2 = \mathbb{R}^2$. Show that X_1 and X_2 are quasi-isometric.

Exercise 4. If G acts geometrically on any geometry, then G is finitely generated.

Exercise 5. (Harder) If G acts geometrically on any simply connected geometry, then G is finitely presented.

Exercise 6. (Harder) If G acts geometrically on any n-connected geometry, then G has a $K(G, 1)$ with finite $(n + 1)$-skeleton.

(For proofs of 2, 3, and 4, see [Cannon 1991].)

The Gromov–Rips thin triangles condition. The most important *approximate* metric property of hyperbolic geometry is the *thin triangles condition:* let δ be a nonnegative constant; geodesic triangles in a path-metric space X are δ*-thin* if each point on each side of each geodesic triangle in X lies within δ of the union of the other two sides of the triangle. A group is *negatively curved* (in the large) or *word-hyperbolic* if it is finitely generated, and with respect to some finite generating set, there is a nonnegative constant δ such that all of the geodesic triangles of the Cayley graph of the group are δ-thin. Gromov has outlined a beautiful theory of negatively curved groups, and many mathematicians have helped to work out the details.

Many of the results are suggested directly by the classical case of hyperbolic geometry and the Kleinian groups which act geometrically on hyperbolic geometry.

We shall outline only one aspect of the Gromov program, namely the *space at infinity.*

The space at infinity. We have already noted that for each of our models H, I, J, K, L, there is a natural space at infinity: in the model I, for example, it is the unit $(n-1)$-sphere that bounds I. This space at infinity can be seen from within the models themselves, as we indicated in the outline of Mostow's proof and in more detail now explain.

To each point "at infinity", there is a family of geodesic rays within the model that "meet" at the given point at infinity in a well-defined sense. Namely, define a *point at infinity* as an equivalence class of geodesic *rays*, any two being equivalent if they are asymptotically near one another (remain within a bounded distance of one another). See Figure 35. We let S_∞ denote the set of such equivalence classes and call it the *space at infinity.*

We can define an intrinsic topology on the space at infinity as follows: given a single geodesic ray, the orthogonal complement at a point on the ray determines

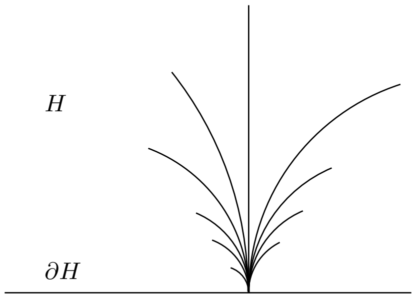


Figure 35. Geodesics with a common endpoint at infinity.

a hyperplane that bounds two hyperbolic half-spaces of hyperbolic space. One of these two half-spaces, the one containing the terminal subray of our ray, cuts off a disk on the sphere at infinity, and determines thereby a basic or fundamental neighborhood of the endpoint of the geodesic ray. See Figure 36. It is easy to see that this topology is invariant under hyperbolic isometries, and that the group of isometries acts as homeomorphisms of S_∞.

Gromov [1987] has shown that an analogous space at infinity can always be defined for a space where triangles are uniformly thin. Though his construction is not exactly analogous to what we have just described, it is nevertheless possible to obtain exactly the Gromov space by a construction that *is* exactly analogous to what we have described [Cannon 1991]. In particular, one may define geodesic rays and equivalent rays, also half-spaces and fundamental "disks" at infinity. See Figure 37.

A special property of the classical spaces at infinity is that hyperbolic isometries act on the space at infinity not only as homeomorphisms but also as conformal mappings. This can be seen from the conformal models simply by the fact that the isometries preserve spheres in the ambient space $\mathbb{R}^{n+1}$, and so preserve spheres on S_∞. The same is true of Gromov boundaries only in a weak sense.

For a deeper understanding of negatively curved groups and the related word-hyperbolic group, the reader should turn to the expository articles [Alonso et al.

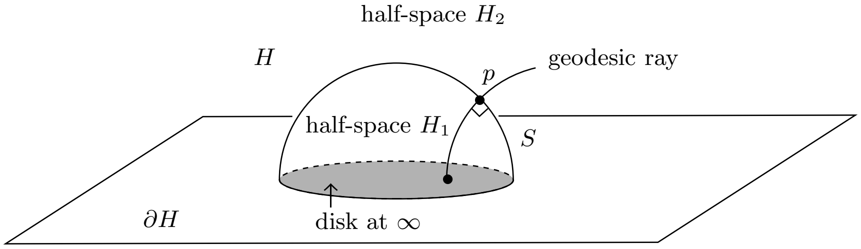


Figure 36. The disk at ∞ determined by a ray p and a point on that ray. S is the hyperbolic orthogonal complement to p.

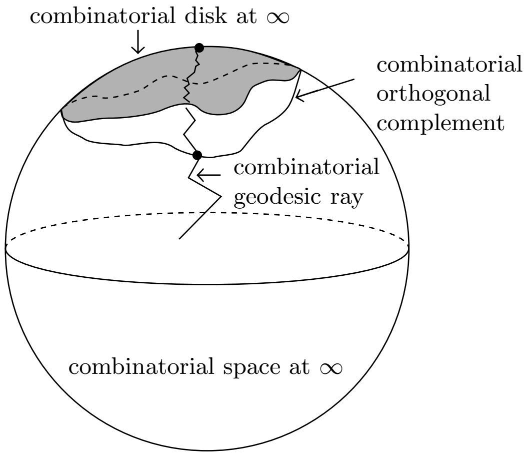


Figure 37. The combinatorial analogue.

1991; Cannon 1991; Coornaert, Delzant, and Papadopoulos 1990; Ghys and de la Harpe 1990] and Swenson's Ph.D. thesis [1993].

16. When Does a Manifold Have a Hyperbolic Structure?

Deciding when a manifold has a hyperbolic structure is a difficult problem. Much work has been done on this problem, and there are several hyperbolization conjectures and theorems. Let M be a closed (compact, without boundary) three-manifold. If M is hyperbolic, it is known that its fundamental group $\pi_1(M)$ satisfies these conditions:

- It is infinite.
- It does not contain a $\mathbb{Z} \oplus \mathbb{Z}$.
- It is not a free product.

Thurston's hyperbolization conjecture is that the converse is also true: these three conditions are also sufficient for M to be hyperbolic. Thurston has proved this under some additional assumptions.

We now describe one of several programs attempting to prove the hyperbolization conjecture. This program involves at various stages all the connections of Figure 34; in fact one can trace the line of proof in a spiral fashion around the diagram in Figure 34. We start in the upper right corner.

The first step, Mosher's *weak hyperbolization conjecture* [Mosher 1995; Mosher and Oertel $\geq$ 1997] states that if $G = \pi_1(M)$ satisfies the above three conditions, then it has thin triangles (by which we mean that its Cayley graph $\Gamma(G)$ for some choice of generators has the thin triangles property). This brings us from topology into the domain of combinatorial group theory.

Note that in a group G with the thin triangles property you can define the space at infinity ∂G, whose points are equivalence classes of geodesics (in $\Gamma(G)$) staying a bounded distance apart.

Assuming additionally that $\pi_1(M)$ has boundary homeomorphic to S^2, we attempt to equip this sphere with a conformal structure on which $\pi_1(M)$ acts uniformly quasiconformally. This would bring the problem into the domain of conformal mappings.

We would then apply a result of Sullivan and Tukia to conclude that the group acts conformally for another conformal structure, quasiconformally equivalent to this one.

Conformal self-maps of S^2 extend to hyperbolic isometries of H^3 (in the disk model I). This would then give us, by taking the quotient, a hyperbolic manifold (actually, an orbifold) M' homotopy equivalent to M.

Gabai and collaborators [Gabai 1994a; Gabai 1994b; Gabai, Meyerhoff, and Thurston 1996] are extending Mostow rigidity to show that a three-manifold homotopy equivalent to a hyperbolic three-manifold M' is in fact homeomorphic to M'.

So this would complete the program. Unfortunately many gaps remain to be bridged.

Our current focus is on the construction of a conformal structure assuming $\pi_1(M)$ has thin triangles and the space at infinity is homeomorphic to S^2. We have the following theorem (the converse of what we'd like to prove):

Theorem 16.1. *Suppose a group G acts geometrically on H^3. Then*:

1. *G is finitely generated.*
2. $\Gamma = \Gamma(G)$ (*the Cayley graph for some choice of generators*) *has thin triangles.*
3. $\partial\Gamma \cong S^2$.

Conjecture 16.2. *The converse holds.*

Here is the intuition behind parts 2 and 3 in Theorem 16.1. The group G acts geometrically on Γ and on H^3. By the quasi-isometry theorem, H^3 and Γ are quasi-isometric.

Consequently, the image in H^3 of a geodesic in Γ looks "in the large" like a geodesic with a *linear* factor of inefficiency. To avoid exponential inefficiency, it must stay within a bounded distance of some genuine geodesic.

Any triangle in Γ will map to a thin triangle in H^3, and hence is thin itself, which proves condition 2. Condition 3 is established similarly.

To understand the difficulty in proving the conjecture, we have to appreciate the difference between constant and variable negative sectional curvature.

Consider the following example, which illustrates a variable negative curvature space. In the space $K^3 = \{(x, y, z) : z > 0\}$, consider the paths that are piecewise vertical (in the z-direction) or horizontal (parallel to the xy-plane). Use the metric length element $|dz|/z$ for vertical paths, and the metric $\max\{|dx|/z^a,\ |dy|/z^b\}$ (where $a, b > 0$ are constants) for horizontal paths.

This metric is analogous to the Riemannian metric

$$ds^2 = \frac{dx^2}{z^{2a}} + \frac{dy^2}{z^{2b}} + \frac{dz^2}{z^2},$$

but the calculations are simpler. Note that the latter reduces to the hyperbolic metric when $a = b = 1$.

In a plane parallel to the xz-plane our metric is analogous to

$$\frac{dx^2}{z^{2a}} + \frac{dz^2}{z^2},$$

which, under the change of variables $X = ax$, $Z = z^a$, yields the metric

$$\frac{1}{a^2}\,\frac{dX^2 + dZ^2}{Z^2};$$

the latter is a scaled version of the hyperbolic metric. A similar formula holds for the planes parallel to the yz plane. If $a \neq b$ then these two sectional curvatures are indeed different.

It is not hard to figure out what the geodesics in K^3 look like. A shortest (piecewise horizontal and vertical) curve joining two points $p_1 = (x_1, y_1, z_1)$ and $p_2 = (x_2, y_2, z_2)$ goes straight up from p_1 to some height z_3, then goes horizontally and straight in the plane $z = z_3$ until it is above p_2, and then goes straight down to p_2. Since the length of such a path is

$$\ell(z_3) = \log(z_3/z_1) + \log(z_3/z_2) + \max\{|x_1 - x_2|z_3^{-a}, |y_1 - y_2|z_3^{-b}\},$$

we can then find the optimal z_3 by differentiating and considering the various cases.

Consider a geodesic line of the form $p(t) = (x_1, y_1, z_1e^{-t})$. The half space specified by that line and the point $p(0)$ turns out to be the box

$$B = \left\{(x, y, z) : |x - x_1| < \frac{2z_1^b}{b}\ ,\ |y - y_1| < \frac{2z_1^a}{a}\ ,\ 0 < z < z_1\right\}.$$

The footprint of this half-space on the space at infinity is the rectangle

$$\left\{(x, y) : |x - x_1| < 2\frac{z_1^b}{b}\ ,\ |y - y_1| < 2\frac{z_1^a}{a}\right\},$$

whose aspect ratio is az_1^{b-a}/b. These aspect ratios are not bounded, so the half spaces do not induce any reasonable conformal structure at infinity.

Note that the isometries of K^3 include horizontal translations and maps of the form $(x, y, z) \to (v^a x, v^b y, vz)$. The latter map acts linearly on the space at infinity. However, for large v, the quasiconformal distortion is unbounded (when $a \neq b$).

17. How to Get Analytic Coordinates at Infinity?

The previous example suggests that the task of finding analytic coordinates on S^2 for which the group acts uniformly quasiconformally may be difficult. Among the uncountably many quasiconformality classes of conformal structures on a topological S^2, one must select (the unique) one on which the group acts uniformly quasiconformally.

In order to accomplish this task, one needs to work with whatever structure on the sphere is a priori provided by the group. Let v_0 be the vertex of Γ corresponding to the identity of G. Fix some positive integer n. Consider the collection of all combinatorially defined half-spaces defined by any geodesic ray starting at v_0 and the vertex on the ray at distance n from v_0 [Cannon 1991]. These half-spaces cut off combinatorial "disks" at infinity and thereby give a finite covering of S^2. In the appropriate conformal structure on S^2 (if it exists), the sets in this cover are approximately round [Cannon and Swenson $\geq$ 1997]. Hence we should think of this cover as providing a sort of "discrete conformal structure" on S^2.

The uniformization theorem for S^2 says that any conformal structure on S^2 is equivalent to the standard Riemann sphere. Hence, once a conformal structure is constructed, analytic coordinates exist. This suggests that one should look for discrete generalizations of uniformization theorems, and in particular, of the Riemann mapping theorem.

The Riemann mapping theorem is a theorem about conformal mappings, and conformality is usually defined in terms of analytic derivatives. In the absence of a priori analytic coordinates, any discrete Riemann mapping theorem cannot begin with a well-defined notion of analytic derivative. Fortunately, there are variational formulations of the Riemann mapping theorem that avoid the mention of derivatives. One is based on *extremal length*.

Consider a quadrilateral Q in the plane $\mathbb{C}$. This is just a closed topological disk with four distinct points marked on the boundary. These marked points partition the boundary of the disk into four arcs, denoted a, b, c, d in Figure 38. Consider metrics on Q that are conformal to the metric that Q inherits from the plane. Conformal changes of metric are determined by positive weight functions $m : Q \to (0, \infty)$ that one should view as point-by-point scalings of the Euclidean metric. With such a weight function m one can define (weighted) lengths of paths γ by

$$\ell_m(\gamma) = \int_\gamma m\,|dz|$$

and (weighted) total areas by

$$a_m = \int_Q m^2\,dz\,d\bar{z}.$$

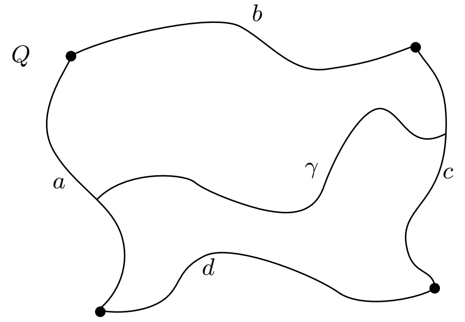


Figure 38. A quadrilateral Q.

Let d_m be the distance in the weighted metric between the edges a and c of Q. It turns out that there is an essentially unique weight function m_0 that maximizes the ratio d_m^2/a_m, and (Q, d_{m_0}) is isometric with a rectangle. (Actually, m_0 is unique up to a positive scalar multiple and a.e. equivalence. Here we take m_0 to be the continuous representative in its a.e. equivalence class.) The maximal ratio $d_{m_0}^2/a_{m_0}$ is also called the extremal length from a to c. It may be interpreted as the resistance to the flow of electricity between a and c if the quadrilateral is interpreted as a conducting metal plate.

This approach provides a uniformization theorem that does not resort to derivatives. It also has a discrete counterpart. (See Figure 39. For more information, see [Cannon 1994; Cannon, Floyd, and Parry 1994].)

A finite covering $C = \{C_j\}$ of a quadrilateral or annulus Q' provides us with a discrete extremal length. In this discrete setting, a weight function is just an assignment of a nonnegative number $m(C_j)$ to each set C_j in the covering. A length of a path γ in Q can be defined as just the sum of $m(C_j)$ over all $C_j \in C$ that intersect γ, and the area of m is defined as the sum of $m(C_j)^2$ over all sets $C_j \in C$. We can then solve a discrete version of the extremal length problem on Q', and use the solution to define an "approximate conformal structure".

This technique can be applied to find a conformal structure on $S^2 = \partial G$, if it exists: the half-spaces defined by G as n increases define a nested sequence of covers C^n of S^2; we get a sequence of "finite" conformal structures that must converge, in the appropriate sense, to a genuine quasiconformal structure if one exists.

In this respect, we close with the following theorem.

Theorem 17.1 (Cannon, Floyd, Parry). *There exists an invariant conformal structure on S^2 if and only if the sequence of covers C^n satisfies the following: for every $x \in S^2$ and for every neighborhood U of x, there is an annulus Q whose closure lies in $U \setminus \{x\}$ and that separates x from $S^2 \setminus U$, such that the discrete extremal lengths between the boundary components of Q with respect to the sequence of covers C^n are bounded away from* 0.

References

[Alonso et al. 1991] J. M. Alonso et al., “Notes on word hyperbolic groups”, pp. 3–63 in *Group theory from a geometrical viewpoint* (Trieste, 1990), edited by E. Ghys et al., World Scientific, River Edge, NJ, 1991.

[Benedetti and Petronio 1992] R. Benedetti and C. Petronio, *Lectures on hyperbolic geometry*, Universitext, Springer, Berlin, 1992.

[Bolyai and Bolyai 1913] F. Bolyai and J. Bolyai, *Geometrische Untersuchungen*, Teubner, Leipzig and Berlin, 1913. Edited by P. G. Stackel. Reprinted by Johnson Reprint Corp., New York and London, 1972. Historical and biographical materials.

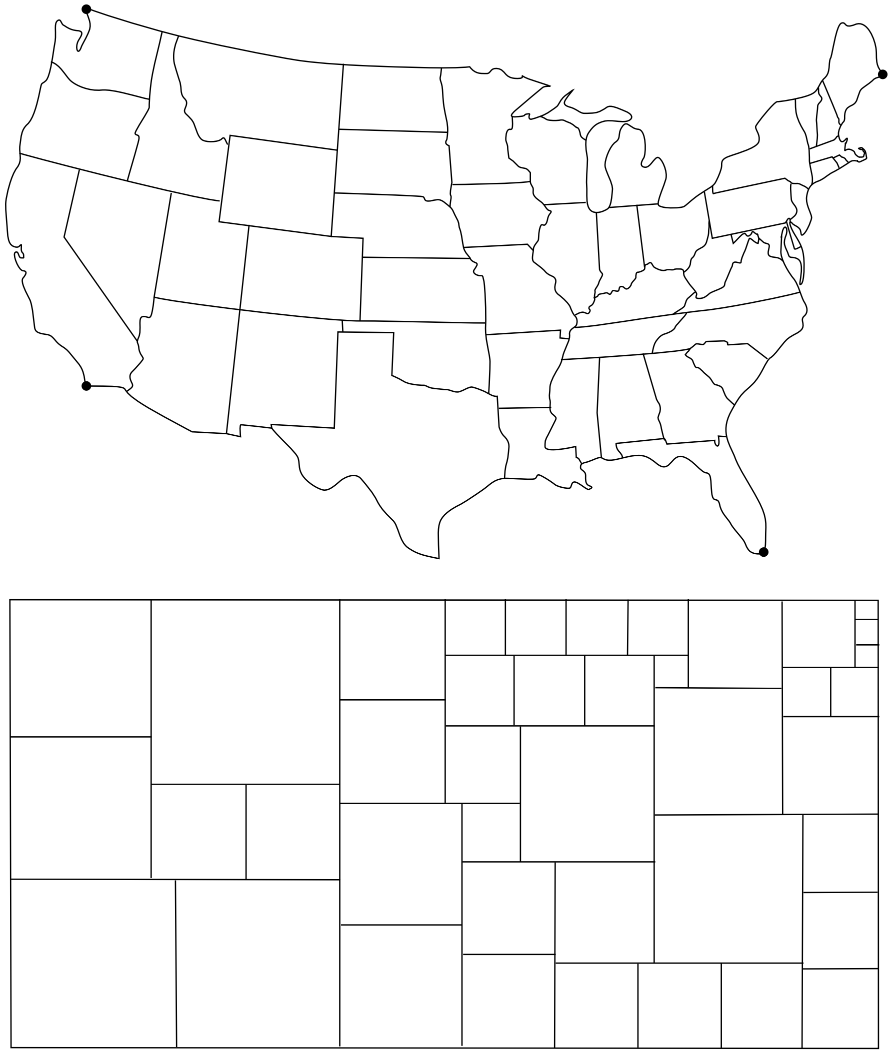


Figure 39. Combinatorial Riemann mapping.

[Cannon 1991] J. W. Cannon, "The theory of negatively curved spaces and groups", pp. 315–369 in *Ergodic theory, symbolic dynamics and hyperbolic spaces*, edited by T. Bedford et al., Oxford U. Press, Oxford, 1991.

[Cannon 1994] J. W. Cannon, "The combinatorial Riemann mapping theorem", *Acta Math.* **173** (1994), 155–234.

[Cannon et al. 1994] J. W. Cannon, W. J. Floyd, and W. R. Parry, "Squaring rectangles: the finite Riemann mapping theorem", pp. 133–212 in *The mathematical legacy of Wilhelm Magnus: groups, geometry and special functions* (Brooklyn, NY, 1992), edited by W. Abikoff et al., Contemporary Mathematics **169**, Amer. Math. Soc., Providence, 1994.

[Cannon and Swenson $\geq$ 1998] J. W. Cannon and E. L. Swenson, "Recognizing constant curvature groups in dimension 3". Preprint.

[Coornaert et al. 1990] M. Coornaert, T. Delzant, and A. Papadopoulos, *Géometrie et théorie des groupes: les groupes hyperboliques de Gromov*, Lecture Notes in Math. **1441**, Springer, Berlin, 1990.

[Euclid 1926] Euclid, *The Thirteen Books of Euclid's Elements*, second ed., Cambridge U. Press, Cambridge, 1926. Edited by T. L. Heath. Reprinted by Dover, New York, 1956.

[Gabai 1994a] D. Gabai, "Homotopy hyperbolic 3-manifolds are virtually hyperbolic", *J. Amer. Math. Soc.* **7** (1994), 193–198.

[Gabai 1994b] D. Gabai, "On the geometric and topological rigidity of hyperbolic 3-manifolds", *Bull. Amer. Math. Soc. (N.S.)* **31** (1994), 228–232.

[Gabai et al. 1996] D. Gabai, R. Meyerhoff, and N. Thurston, "Homotopy hyperbolic 3-manifolds are hyperbolic", 1996. Available at http://www.msri.org/MSRI-preprints/online/1996-058.

[Ghys and de la Harpe 1990] E. Ghys and P. de la Harpe, *Sur les groupes hyperboliques d'après Mikhael Gromov*, Birkhäuser, boston, 1990.

[Gromov 1987] M. L. Gromov, "Hyperbolic groups", pp. 75–263 in *Essays in group theory*, edited by S. Gersten, MSRI Publications **8**, Springer, New York, 1987. Perhaps the most influential recent paper in geometric group theory.

[Hilbert and Cohn-Vossen 1932] D. Hilbert and S. Cohn-Vossen, *Anschauliche Geometrie*, Grundlehren der math. Wissenschaften **37**, Springer, Berlin, 1932. Reprinted by Dover, New York, 1944. Translated by Paul Nemenyi as *Geometry and the Imagination*, Chelsea, New York, 1952. College-level exposition of rich ideas from low-dimensional geometry, with many figures.

[Iversen 1993] B. Iversen, *Hyperbolic geometry*, London Math. Soc. Student Texts **25**, Cambridge U. Press, Cambridge, 1993. Very clean algebraic approach to hyperbolic geometry.

[Klein 1928] F. Klein, *Vorlesungen über nicht-Euklidische Geometrie*, Grundlehren der math. Wissenschaften **26**, Springer, Berlin, 1928. Edited by W. Rosemann. Reprinted by Chelsea, New York, 1960. Mostly algebraic development of non-Euclidean geometry with respect to Klein and projective models. Beautiful figures. Elegant exposition.

[Kline 1972] M. Kline, *Mathematical thought from ancient to modern times*, Oxford U. Press, New York, 1972. A well-written and fairly concise history of mathematics (1200 pages in one or three volumes). Full of interesting material.

[Lobachevskiĭ 1898] N. I. Lobachevskiĭ, *Zwei geometrische Abhandlungen*, Bibliotheca math. Teubneriana **70**, Teubner, Leipzig, 1898. Reprinted by Johnson Reprint Corp., New York and London, 1972. Original papers.

[Mosher 1995] L. Mosher, "Geometry of cubulated 3-manifolds", *Topology* **34** (1995), 789–814.

[Mosher and Oertel ≥ 1998] L. Mosher and U. Oertel, "Spaces which are not negatively curved". Preprint.

[Poincaré 1908] H. Poincaré, *Science et méthode*, E. Flammarion, Paris, 1908. Translated by F. Maitland as *Science and method*, T. Nelson and Sons, London, 1914. Reprinted by Dover, New York, 1952, and by Thoemmes Press, Bristol, 1996. One of Poincaré's several popular expositions of science. Still worth reading after almost 100 years.

[Ratcliffe 1994] J. G. Ratcliffe, *Foundations of hyperbolic manifolds*, Graduate Texts in Math. **149**, Springer, New York, 1994. Fantastic bibliography, careful and unified exposition.

[Riemann 1854] G. F. B. Riemann, "Über die Hypothesen, welche der Geometrie zu Grunde liegen", *Königliche Gesellschaft der Wissenschaften zu Göttingen* **13** (1854), 1–20. Also pp. 272–287 in *Gesammelte Mathematische Werke*, edited by R. Dedekind and H. Weber, 2nd ed., Teubner, Leipzig, 1902 (reprint: Dover, New York, 1953).

[Swenson 1993] E. L. Swenson, *Negatively curved groups and related topics*, Ph.D. thesis, Brigham Young University, 1993.

[Thurston 1997] W. P. Thurston, *Three-dimensional geometry and Topology*, Princeton U. Press, Princeton, 1997. Thurston reintroduced hyperbolic geometry to the topologist in the 1970s. This richly illustrated book has an accessible beginning and progresses to great and exciting heights.

[Weyl 1919] H. Weyl, *Raum, Zeit, Materie: Vorlesungen über allgemeine Relativitätstheorie*, Springer, Berlin, 1919. Translated by H. L. Brose as *Space–Time–Matter*, Dover, New York, 1922; reprinted 1950. Weyl's exposition and development of relativity and gauge theory; begins at the beginning with motivation, philosophy, and elementary developments as well as advanced theory.

Index

James W. Cannon
Department of Mathematics
Brigham Young University
Provo, UT 84602-1001
cannon@math.byu.edu

William J. Floyd
Department of Mathematics
Virginia Polytechnical Institute and State University
Blacksburg, VA 24061-0123
floyd@math.vt.edu

Richard Kenyon
Département de Mathématiques
École Normale Supérieure de Lyon
46, allée d'Italie, 69364 Lyon
France
rkenyon@umpa.ens-lyon.fr

Walter R. Parry
Department of Mathematics
Eastern Michigan University
Ypsilanti, MI 48197
mth_parry@emuvax.emich.edu