CHAPTER 2

Hyperbolic Geometry

We now give an introduction to non-Euclidean hyperbolic geometry with its more common analytic description.

This material first appeared in *Flavors of Geometry* [**41**], Mathematical Sciences Research Institute Publications 31, Cambridge University Press, 1997, edited by Silvio Levy. Except for a comment or two, this chapter is simply a republication of the paper that appeared there. I thank the Mathematical Sciences Research Institute for permission to reprint it. I thank my coauthors (W. J. Floyd, R. Kenyon, and W. R. Parry[1]) for their hard work in bringing the chapter to completion. The references to this chapter appear in a separate portion of the annotated bibliography. This chapter considers non-Euclidean geometry in all dimensions, not just in dimension 2.

2.1. Introduction

Non-Euclidean, or 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. Einstein and Minkowski found in non-Euclidean geometry a 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 which are particularly important in topology and group theory. They indicate

[1]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.

some of the fundamental problems being approached by means of non-Euclidean geometry in topology and group theory.

Of course, volumes have been written on non-Euclidean geometry which the reader must consult for more exhaustive information.

2.2. The Origins of Hyperbolic Geometry

Except for Euclid's five fundamental postulates of plane geometry, which we paraphrase from Kline [**60**], most of the following historical material is taken from Felix Klein's book [**59**]. Other historical references appear in the bibliography. Here are Euclid's postulates:

(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 produced 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. For two thousand years mathematicians attempted to deduce postulate (5) from the four simpler postulates. In each case one reduced the proof of postulate (5) to the conjunction of the first four postulates with an additional natural postulate which, in fact, proved to be equivalent to postulate (5):

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

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

Italian **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.

Lambert (1728-1777) proceeded in a similar fashion and wrote an extensive work on the subject which was published after his death in 1786.

Göttingen mathematician **Kaestner** (1719-1800) directed a thesis of student **Klügel** (1739-1812) which considered approximately 30 proof attempts for the parallel postulate.

Decisive progress came in the 19th century when mathematicians abandoned the effort to find a contradiction in the denial of postulate (5) and instead worked out carefully and completely the consequences of such a denial. Unusual consequences of that denial 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, which we entitle *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). Schweikart by the year 1816, in his spare time, developed an "astral geometry" which was independent of the parallel axiom (5). His nephew Taurinus had attained a non-Euclidean hyperbolic geometry by the year 1824.

The three professional mathematicians were **C. F. Gauss** (1777-1855), **N. Lobachevskii** (1793-1856) (see [**61**]), and **Johann Bolyai** (1802-1860) (see [**44**]). From the papers of Gauss's 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. Johann Bolyai's father Wolfgang (1775-1856) was a student friend of Gauss and remained in correspondence with him throughout his life. Wolfgang devoted much of his life's effort unsuccessfully to the proof of the parallel postulate and consequently tried to turn his son Johann away from its study. Nevertheless, Johann 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. Lobachevskii also developed a non-Euclidean geometry extensively and was, in fact, the first to publish his work (1829).

Gauss, the Bolyais, and Lobachevskii 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. Lobachevskii developed a non-Euclidean **trigonometry** which 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 **Euler**, **Monge**, and **Gauss** in their studies of curved surfaces. In 1837 **Lobachevskii** suggested that curved surfaces of constant negative curvature might represent non-Euclidean geometry. Two years later, working independently and largely in ignorance of Lobachevskii's work, yet publishing in the same journal, **Minding** made an extensive study of surfaces of constant curvature and verified Lobachevskii's suggestion. **Riemann**, in his vast generalization (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 his jumping off point for his studies. But Riemann's work did not appear in print until after his death. All of the connections among these subjects were particularly pointed out by **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 of these models are conformal models associated with the name of **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 [**65**]:

> 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.

2.3. Why Call It Hyperbolic Geometry?

The non-Euclidean geometry of Gauss, Lobachevskii, 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 $(x_1, \ldots, x_{n+1}) \in \mathbb{R}^{n+1}$, with the coordinate x_{n+1} representing time, and the only reasonable metric was the Euclidean metric with the positive definite square-norm $x_1^2 + \ldots + x_{n+1}^2$.

Relativity changed all that; in flat spacetime geometry the speed of light should be constant as viewed from any inertial reference frame. The Minkowski model for spacetime geometry is again $\mathbb{R}^{n+1}$ but with the indefinite norm $x_1^2 + \ldots + x_n^2 - x_{n+1}^2$ 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.

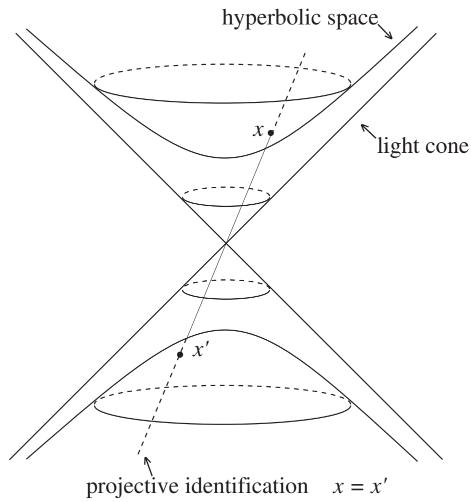

FIGURE 1. Minkowski space

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 standard **n-dimensional sphere** in Euclidean space $\mathbb{R}^{n+1}$ by the formula $S^n = \{x \in \mathbb{R}^{n+1} \mid x \cdot x = 1\}$ and **n-dimensional hyperbolic space** by the formula $\{x \in \mathbb{R}^{n+1} \mid x * x = -1\}$. Thus hyperbolic space is a hyperboloid of two sheets which 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.

2.4. Understanding the One-dimensional Case

The key to understanding $\mathbb{H}^n$ 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

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

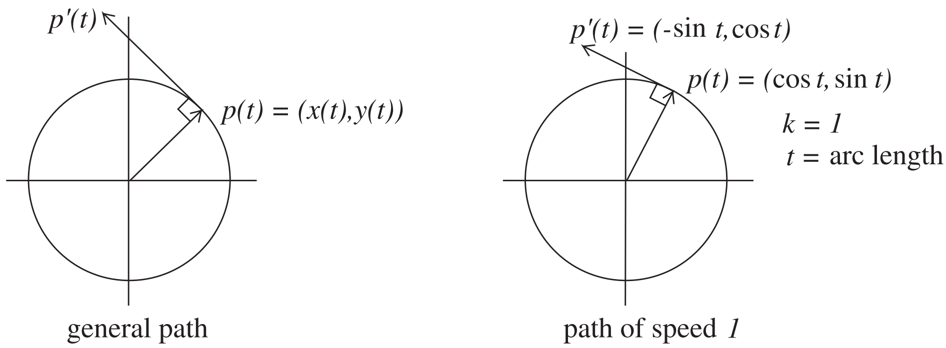

FIGURE 2. S^1, the circle

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

$$\begin{aligned} x'(t) &= -y(t) \\ y'(t) &= x(t) \end{aligned}$$

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:

$$\cos t = 1 - t^2/2! + t^4/4! - \cdots \text{ ; and}$$

$$\sin t = t - t^3/3! + t^5/5! - \cdots .$$

The hyperbolic calculation in $\mathbb{H}^1$ requires only a new starting point $(0, 1)$ instead of $(1, 0)$, the replacement of S^1 by $\mathbb{H}^1$, 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 $p : (-\infty, \infty) \to \mathbb{H}^1$ 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

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

or 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 $\mathbb{H}^1$ at $p(t)$ is one-dimensional and the vector $(y(t), x(t))$ is hyperbolic-perpendicular to $p = (x, y)$. See Figure 3.

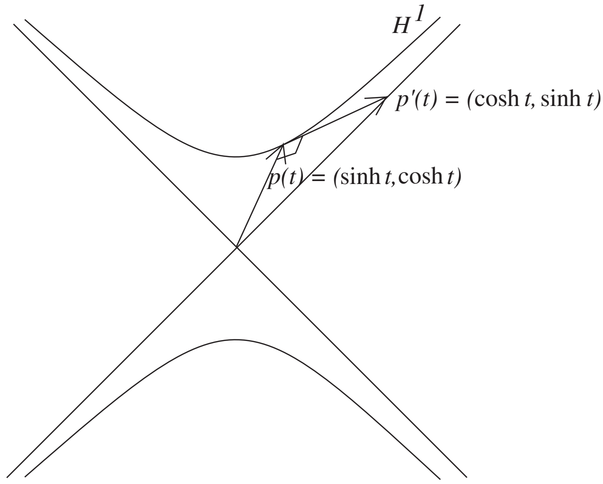

FIGURE 3. $\mathbb{H}^1$

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 $\mathbb{H}^1$, $x = x(t)$ 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.

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

$$\begin{aligned} x'(t) &= y(t) \\ y'(t) &= x(t) \end{aligned}$$

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:

$$\cosh t = 1 + t^2/2! + t^4/4! + \cdots; \text{ and}$$
$$\sinh t = t + t^3/3! + t^5/5! + \cdots.$$

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 $\mathbb{H}^1$ 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 1-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 1 apply in the higher-dimensional setting as well.

2.5. Generalizing to Higher Dimensions

In higher dimensions, $\mathbb{H}^n$ sits inside $\mathbb{R}^{n+1}$ as a hyperboloid. If $p : (-\infty, \infty) \to \mathbb{H}^n$ 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 $\mathbb{H}^n$ 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 $(x \cdot y)^2 \leq (x \cdot x)(y \cdot y)$. Suppose that $p = (\hat{p}, p_{n+1})$ is in $\mathbb{H}^n$ and $x = (\hat{x}, x_{n+1}) \neq 0$ is in the tangent space of $\mathbb{H}^n$ 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 $x_{n+1} \neq 0$. 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

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

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

The second method analyzes the inner product $*$ algebraically. (For complete details, see for example Weyl [**70**].) Take a basis p, p_1, ... , p_n for $\mathbb{R}^{n+1}$ where p is the point of interest in $\mathbb{H}^n$ and the remaining vectors span the n-dimensional tangent space to $\mathbb{H}^n$ at p. Now apply the Gram-Schmidt orthogonalization process to this basis. Since $p * p = -1$ by the defining equation for $\mathbb{H}^n$, 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 $\mathbb{H}^n$ 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 $\mathbb{H}^n$ defines a genuine Riemannian metric on $\mathbb{H}^n$.

2.6. 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 which 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

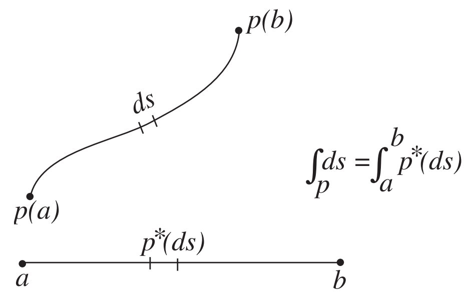

FIGURE 4. The length of a path

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, then 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 which 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 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.

2.7. 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 which 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. As mnemonic names for these analytic models we choose the following:

H, the **H**alf-space model.
I, the **I**nterior of the disk model.

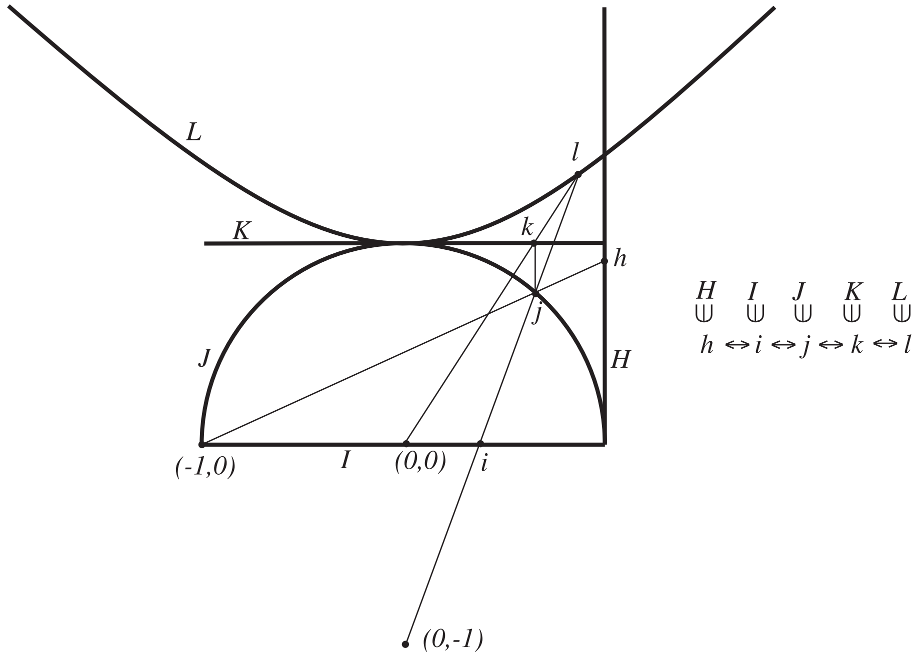

FIGURE 5. The five analytic models and their connecting isometries

J, the **J**emisphere model. (Pronounce with a south-of-the-border accent.)
K, the **K**lein model.
L, the hyperbo**L**oid model, or 'Loid model, for short.

Each model has its own metric, geodesics, isometries, and so on.

Here are set descriptions of the five analytic models (see Figure 5):

$$H = \{(1, x_2, \ldots, x_{n+1}) \mid x_{n+1} > 0\};$$

$$I = \{(x_1, \ldots, x_n, 0) \mid x_1^2 + \cdots + x_n^2 < 1\};$$

$$J = \{(x_1, \ldots, x_{n+1}) \mid x_1^2 + \cdots + x_{n+1}^2 = 1 \text{ and } x_{n+1} > 0\};$$

$$K = \{(x_1, \ldots, x_n, 1) \mid x_1^2 + \cdots + x_n^2 < 1\}; \text{ and}$$

$$L = \{(x_1, \ldots, x_n, x_{n+1}) \mid x_1^2 + \cdots + x_n^2 - x_{n+1}^2 = -1 \text{ and } x_{n+1} > 0\}.$$

The associated Riemannian metrics ds^2 which complete the analytic description of the five models are:
for H,

$$ds_H^2 = (dx_2^2 + \cdots + dx_{n+1}^2)/x_{n+1}^2;$$

for I,

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

for J,

$$ds_J^2 = (dx_1^2 + \cdots + dx_{n+1}^2)/x_{n+1}^2;$$

for K,

$$ds_K^2 = (dx_1^2+\cdots+dx_n^2)/(1-x_1^2-\cdots-x_n^2)+(x_1\,dx_1+\cdots+x_n\,dx_n)^2/(1-x_1^2-\cdots-x_n^2)^2;$$

and for L,

$$ds_L^2 = dx_1^2 + \cdots + dx_n^2 - dx_{n+1}^2.$$

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:

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

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

$$\gamma : K \to J, \quad (x_1, \ldots, x_n, 1) \mapsto (x_1, \ldots, x_n, x_{n+1}),$$

with $x_1^2 + \cdots + x_{n+1}^2 = 1$ and $x_{n+1} > 0$; and

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

The geometry of these mappings is the following:

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

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

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

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

Each map can be used in the standard way to pull back the Riemannian metric from the target space to the domain space 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 personally found 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 which 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:

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

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

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

The other two pullback computations are comparable.

2.8. 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 (1) preserves angles and (2) 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 point of tangency S which we think of as the south pole of S^n. Let N denote the point of S^n opposite S, a point which we think of as the north pole of S^n. If x is any point of $S^n\setminus\{N\}$, then there is a unique point $\pi(x)$ of P on the line which 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 $\{x\mid x_{n+1}=1\}$ into P.

THEOREM 2.1 (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 $\pi(x)\cdot x\cdot y$ and $x\cdot\pi(x)\cdot\pi(y)$ 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 $(0,\ldots,0,-1)$, N is the point with coordinates $(0,\ldots,0,1)$, P is the plane $x_{n+1}=-1$, and $\pi:S^n\to P$ is given by the formula

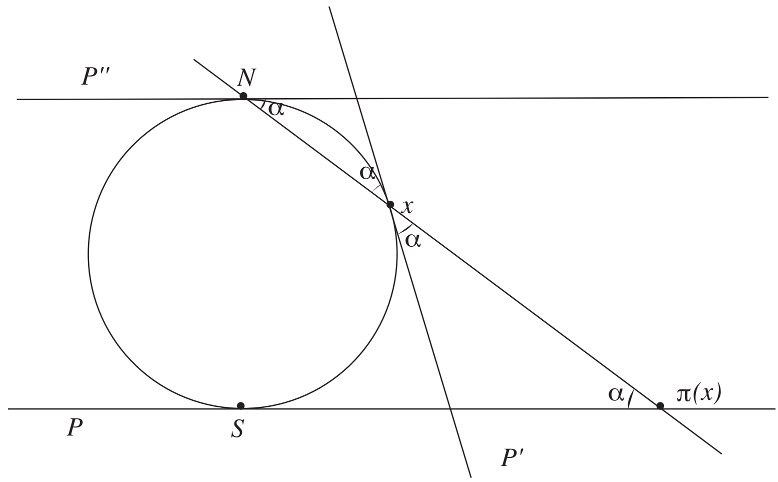

FIGURE 6. Stereographic projection

$\pi(x) = (y_1, \ldots, y_n, -1)$ 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

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

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

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

and

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

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 which 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 which, 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—Cohn-Vossen [**57**].

[Comment added, 2016: Although the proof given above is very attractive and works in the dimension considered in Hilbert—Cohn-Vossen, it seems to need modification in higher dimensions.]

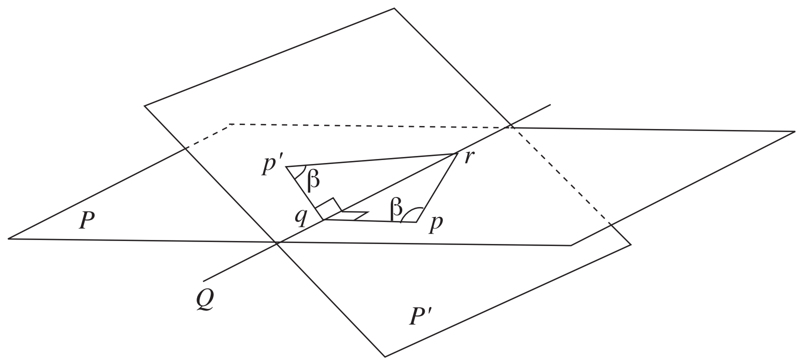

FIGURE 7. The angles qpr and $qp'r$

In preparation we consider two planes P and P' of dimension n in Euclidean (n+1)-space $\mathbb{R}^{n+1}$ which 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.

Obvious assertion: If $r \in Q$, then the angles qpr and $qp'r$ are equal. See Figure 7. Similarly, the angles $p'pr$ and $pp'r$ are equal.

To prove the second assertion, 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 3, the obvious assertion 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. □

THEOREM 2.2 (Preservation of spheres). *Assume the setting of the previous theorem. If C is a sphere (C for circle) in S^n which 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

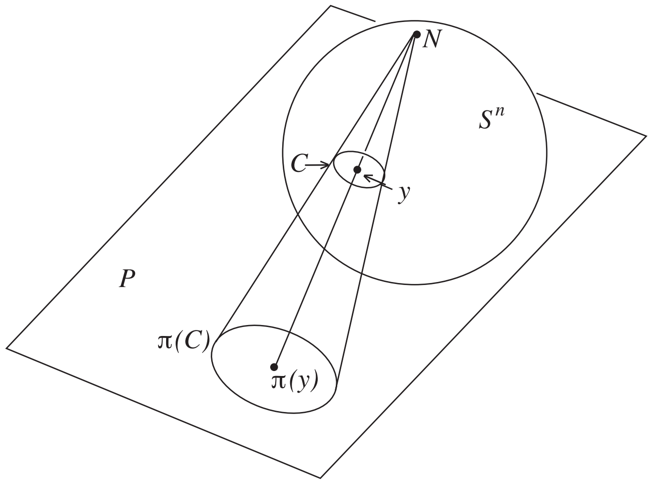

FIGURE 8. The spheres to spheres property of stereographic projection

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 2-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 which proves the assertion. Let $x \in C \setminus S$, and consider the 2-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 $N \cdot \pi(x) \cdot \pi(y)$ and $N \cdot x' \cdot y$ are equal. By the final assertion of Theorem 2.1, the angles $\pi(y) \cdot \pi(x) \cdot x$ and $y \cdot x \cdot \pi(x)$ 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 $N \cdot x' \cdot y$ and $N \cdot \pi(x) \cdot \pi(y)$, 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 which 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 2.3. *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 2.1 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'. □

2.9. 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 upper 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 2.4 (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 $r : X \to$ image(C) which is distance reducing in the sense that, if one restricts the metric of X to 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.*

Proof. 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 2.5 (Existence of a fundamental geodesic in hyperbolic space). *In the upper half-space model of hyperbolic space, all vertical lines are geodesic. In fact they contain the unique shortest path between any pair of points of the line.*

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

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

$$r(1, x'_1, \dots, x'_n, t) = (1, x_1, \dots, x_n, t).$$

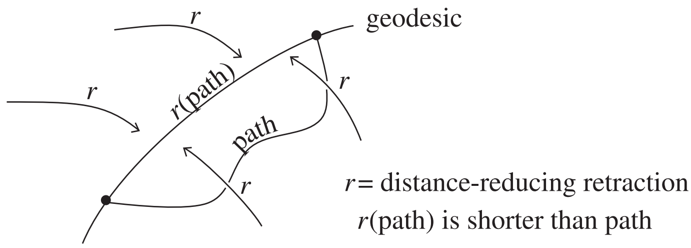

FIGURE 9. The retraction principle

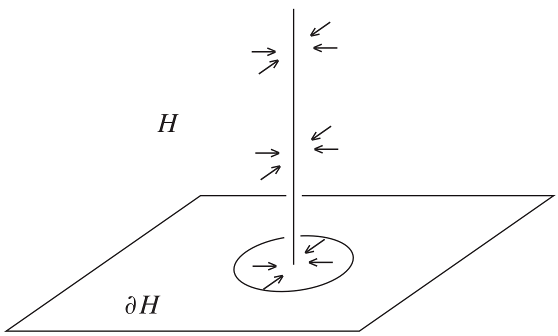

FIGURE 10. A fundamental hyperbolic geodesic and a distance-reducing retraction

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 which 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 which stays in the image. □

THEOREM 2.6 (Classification of geodesics in H). *The geodesics in the upper 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:

(1) Euclidean isometries of H which take the boundary $\{(1, x_2, \ldots, x_n, 0)\}$ of H to itself are hyperbolic isometries of H. Similarly, the transformations of H

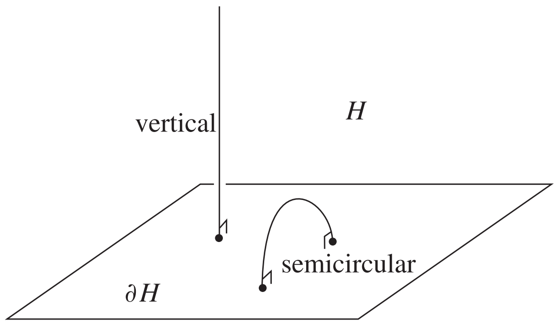

FIGURE 11. The two types of geodesics in H

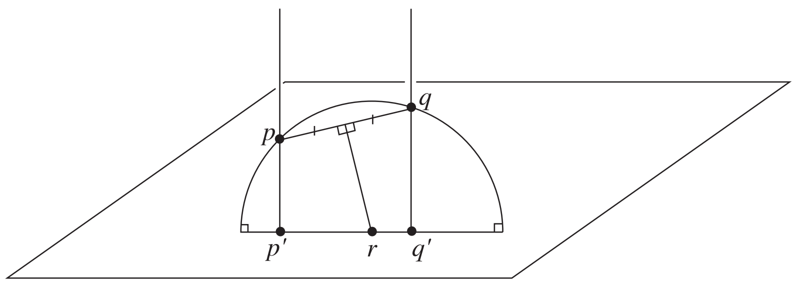

FIGURE 12. Finding the hyperbolic geodesic between points of H not on a vertical line

which take $(1, x_1, \ldots, x_n, t)$ to $(1, r \cdot x_1, \ldots, r \cdot x_n, r \cdot t)$ 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 which 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

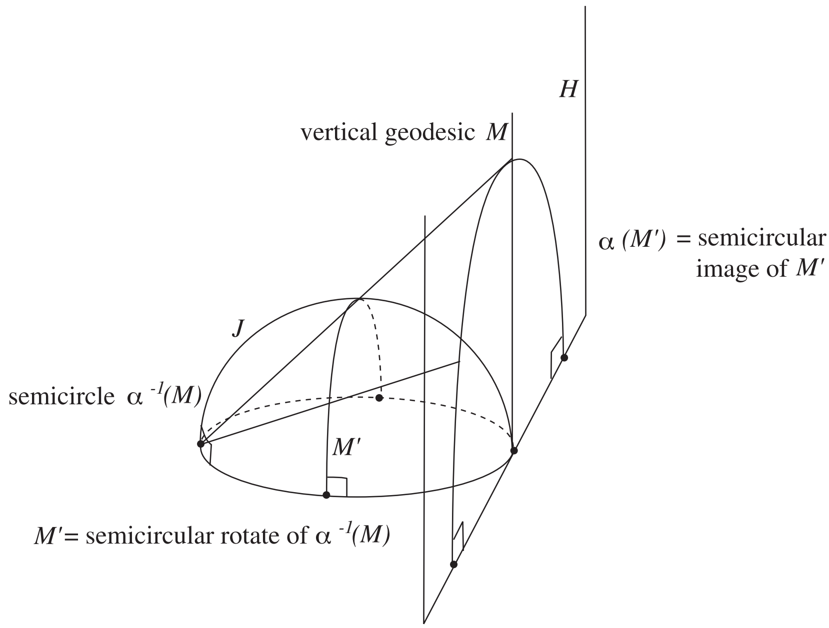

FIGURE 13. Geodesics in H

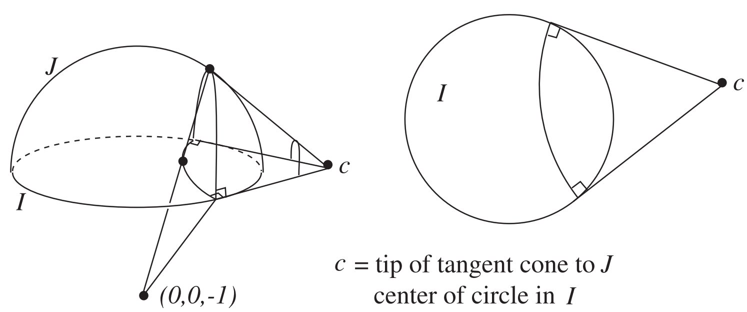

FIGURE 14. Geodesics in I and J and their stereographic relationship

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. □

Discussion of geodesics in the other analytic models. By Theorems 2.1 and 2.2, 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.

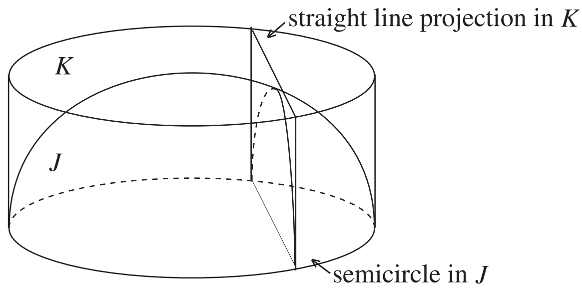

FIGURE 15. Geodesics in J and K

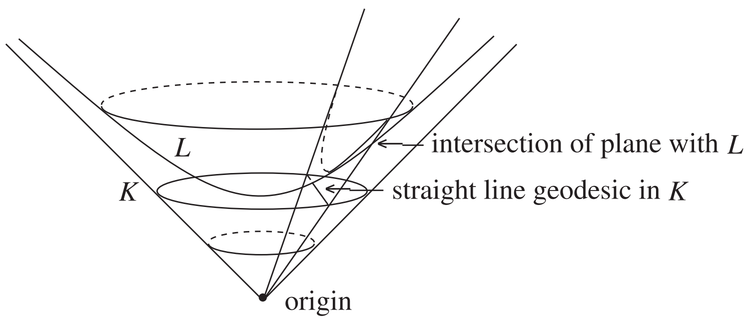

FIGURE 16. Geodesics in K and L

By Theorem 2.3, 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.

2.10. 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 $F : \mathbb{R}^{n+1} \to \mathbb{R}^{n+1}$ which preserves the hyperbolic inner product $*$ (that is, for each pair v and w of vectors from $\mathbb{R}^{n+1}$, $Fv * Fw = v * w$) and which 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 which preserves the Riemannian metric (that is, $f^*(ds^2) = ds^2$).