Visualizing Hyperbolic Space: Unusual Uses of 4×4 Matrices

Mark Phillips Charlie Gunn

The National Science and Technology Research Center for Computation and Visualization of Geometric Structures (The Geometry Center)

December 15, 1991

Abstract

We briefly discuss hyperbolic geometry, one of the most useful and important kinds of non-Euclidean geometry. Rigid motions of hyperbolic space may be represented by 4×4 homogeneous transformations in exactly the same way as rigid motions of Euclidean space. This is a happy situation for those of us interested in visualizing what life in hyperbolic space might be like, because it means we can use existing graphics hardware and software libraries to animate scenes in hyperbolic space. We present formulas for computing reflections, translations, and rotations in hyperbolic space. These are a bit more complicated than the corresponding formulas for Euclidean geometry, which emphasizes our need for graphics libraries which allow completely arbitrary 4×4 transformations.

The use of 4×4 transformations to represent isometries of hyperbolic space is not new; it has been used since the discovery of non-Euclidean geometry in the 19-century. The new part of our work is the application of this theory to real-time 3D computer graphics technology, which for the first time ever is allowing mathematicians to interactively explore hyperbolic geometry.

The Geometry Center is funded by the National Science Foundation, the Department of Energy, Minnesota Technology, Inc., and the University of Minnesota.

The authors may be reached at: The Geometry Center, 1300 South Second Street, Minneapolis, MN 55407. (612) 626-0888. Email: mbp@geom.umn.edu, gunn@geom.umn.edu.

Introduction

The use of 4×4 matrices to represent affine transformations of Euclidean 3-space is well-known in computer graphics. Most graphics languages include provisions for specifying 4×4 transformations, and most interactive graphics workstations have the ability to multiply 4×4 matrices in hardware. These capabilities were designed with Euclidean geometry in mind, because we think of the space in which we live as Euclidean 3-space.

There are, however, alternate systems of geometry which are of interest in mathematics and physics research and education. One of the most important of these is hyperbolic geometry. Hyperbolic space arises naturally, even more so than Euclidean geometry, in the study and classification of 3-manifolds. It is also frequently taught in introductory geometry courses because it is in some sense the simplest and most elegant type of non-Euclidean geometry. Learning hyperbolic geometry forces one to challenge many assumptions which are usually taken for granted, in the process strengthening one's geometric reasoning skills.

The "space" of hyperbolic geometry consists of the interior of the unit ball in $\mathbf{R}^3$; the boundary of the ball, the unit sphere, is "at infinity". Distance is redefined to approach infinity as we move closer to this sphere. From a hyperbolic point of view, therefore, we can never actually reach the boundary sphere. We can think of hyperbolic space as consisting of points, lines, planes, surfaces, etc, just as in Euclidean space. In hyperbolic space, however, some of the rules of geometry are different. Specifically, Euclid's fifth postulate is not valid: in the hyperbolic plane there are many lines through a given point which do not intersect a given line. Another non-Euclidean property is that the sum of the angles in a planar polygon is always less than 180 degrees. It is possible, for example, to have a "regular right pentagon" (all five sides are equal and all five angles are 90 degrees). Figure 1 shows a tesselation (tiling) of hyper-

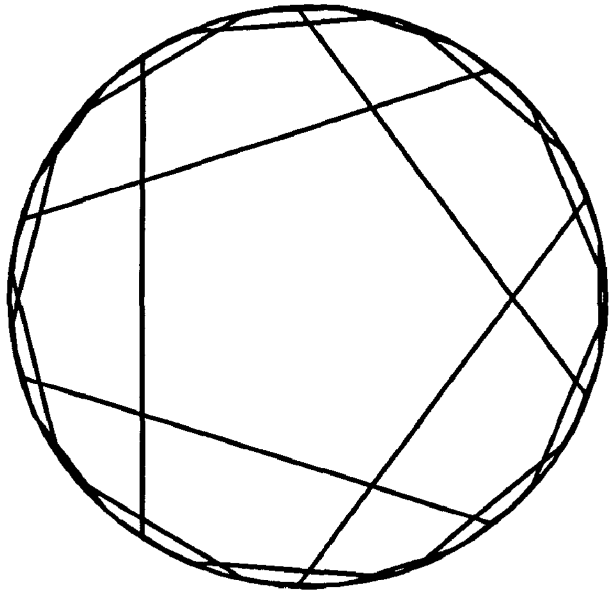

Figure 1: Tiling of the hyperbolic plane by regular right pentagons. All angles in this picture are right angles in the hyperbolic metric, and all pentagons are congruent.

bolic 2-space by such pentagons.

These differences between Euclidean and hyperbolic space mean that the intuition which we have from living in what we perceive as essentially Euclidean 3-space is of little value, and may actually hinder us, in an effort to understand hyperbolic geometry. It would be extremely useful, therefore, for researchers and geometry students alike, to be able to experience some of what life in hyperbolic space might be like.

Fortunately, since the transformations of hyperbolic 3-space can be represented as 4×4 matrices in much the same way as with Euclidean transformations, we can use the matrix capabilities of many graphics languages and hardware systems to create images and to animate motions in hyperbolic space. We must, however, be able to use completely arbitrary 4×4 transformations, because the matrices which arise in hyperbolic geometry are different from those of Euclidean geometry.

Hyperbolic Space

In the following discussion we think of vectors as column vectors; so $\mathbf{a} \in \mathbf{R}^4$ represents the 4×1 matrix $\begin{pmatrix} a_1 \\ a_2 \\ a_3 \\ a_4 \end{pmatrix}$ and its transpose $\mathbf{a}^T$ the 1×4 matrix $(a_1\ a_2\ a_3\ a_4)$. Thus $\mathbf{a}^T\mathbf{b}$ is the usual dot product of $\mathbf{a}$ and $\mathbf{b}$, and $\mathbf{a}\mathbf{b}^T$ is a 4×4 matrix, sometimes called the *outer product* of $\mathbf{a}$ with $\mathbf{b}$.

In computer graphics points in Euclidean 3-space are commonly represented by homogeneous coordinates — i.e. vectors in $\mathbf{R}^4$, where any two vectors which are scalar multiples of each other are considered to represent the same point. The 3-dimensional coordinates (a_1, a_2, a_3) of a point in $\mathbf{R}^3$ are called its *affine coordinates*. We can convert affine coordinates to homogeneous coordinates by appending a 1 as the 4-th coordinate to obtain $(a_1, a_2, a_3, 1)$, and we can convert arbitrary homogeneous coordinates (a_1, a_2, a_3, a_4) to affine coordinates by normalizing to obtain $(a_1/a_4, a_2/a_4, a_3/a_4)$ (assuming $a_4 \neq 0$). The advantage of homogeneous coordinates is that rigid Euclidean motion (isometries), as well as perspective projections, can be represented by multiplication by 4×4 matrices. The isometries of $\mathbf{R}^3$ correspond to the semidirect product of the 3-dimensional orthogonal group $O(3)$ with the 3-dimensional translation group. Recall that an orthogonal matrix $\mathbf{M}$ is one which preserves the inner product of vectors: $\mathbf{Ma} \cdot \mathbf{Mb} = \mathbf{a} \cdot \mathbf{b}$. The inner product in this case is

$$\mathbf{a} \cdot \mathbf{b} = a_1b_1 + a_2b_2 + a_3b_3,$$

where we assume that $\mathbf{a}$ and $\mathbf{b}$ are normalized.

Using other inner products yields non-Euclidean geometries. The inner product

$$\langle \mathbf{a}, \mathbf{b} \rangle_s = a_1b_1 + a_2b_2 + a_3b_3 + a_4b_4.$$

yields spherical geometry, and

$$\langle \mathbf{a}, \mathbf{b} \rangle_h = a_1b_1 + a_2b_2 + a_3b_3 - a_4b_4.$$

yields hyperbolic geometry. Our treatment of hyperbolic geometry is in terms of $\langle \cdot, \cdot \rangle_h$; analogous derivations using $\langle \cdot, \cdot \rangle_s$ instead would yield the corresponding formulas for spherical geometry. Note that the Euclidean inner product, by ignoring the 4-th coordinate, can be seen as a bridge between these two inner products.

$\langle \cdot, \cdot \rangle_h$ is called the *Minkowski inner product*. The Minkowski inner product can also be described as follows. Let

$$\mathbf{I}^{3,1} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix}.$$

Then $\langle \mathbf{a}, \mathbf{b} \rangle_h = \mathbf{a}^T \mathbf{I}^{3,1} \mathbf{b}$. The group of 4×4 matrices which preserve the Minkowski inner product is denoted $O(3,1)$.

Now consider the vectors $V_- = \{\mathbf{a} \in \mathbf{R}^4 \mid\mid \langle \mathbf{a}, \mathbf{a} \rangle_h < 0\}$. The set V_- forms a solid cone along the 4-th axis with vertex at the origin. Hyperbolic 3-space, denoted $\mathbf{H}^3$, is the projectivization of V_-, with the metric induced by the Minkowski inner product; vectors in V_-

correspond to the homogeneous coordinates of points in $\mathbf{H}^3$. Each point in $\mathbf{H}^3$ is represented by a unique vector with 4-th coordinate 1, which can be obtained from any vector in V_- by normalization, just as in the Euclidean case. (The fact that the vector lies in V_- guarantees that the 4-th coordinate is nonzero.) This gives a model of $\mathbf{H}^3$ consisting of those points of V_- with 4-th coordinate 1; this is the same as the interior of the unit ball in 3-space. Hyperbolic space thus consists only of the points inside this ball.

Two-dimensional hyperbolic space, also called the *hyperbolic plane*, consists consists of the interior of the unit disk. Although the discussion below is in terms of hyperbolic 3-space, it extends straightforwardly to any dimension. In particular, the illustrations and examples we give are all in two-dimensions (the 3-rd coordinate is 0) to simplify the computations and the figures.

The geodesics (straight lines) in this model of hyperbolic space are the same as the Euclidean straight lines passing through the unit ball, except that we only consider the part of the line inside the ball. Similarly, the hyperbolic planes in $\mathbf{H}^3$ are the same as the Euclidean planes.

The hyperbolic distance between two points a and b with homogeneous coordinates $\mathbf{a}$ and $\mathbf{b}$ is given by

$$\mathrm{d}^{\mathrm{hyp}}(a,b) = 2\cosh^{-1}\sqrt{\frac{\langle \mathbf{a},\mathbf{b}\rangle_h^2}{\langle \mathbf{a},\mathbf{a}\rangle_h \langle \mathbf{b},\mathbf{b}\rangle_h}}. \qquad (1)$$

A simple calculation shows that this formula is invariant under multiplication of $\mathbf{a}$ and $\mathbf{b}$ by scalars, and hence depends only on a and b. It is also easy to verify that if a remains fixed and we let b approach the boundary of the unit ball, then $\mathrm{d}^{\mathrm{hyp}}(a,b)$ approaches infinity.

The model of hyperbolic space that we are using here is called the *projective* model, or the *Klein* model, after the 19-th century mathematician who popularized it. A more familiar model is the *conformal* model, also known as the *Poincare* model. In the conformal model, geodesics are arcs of circles perpendicular to the boundary sphere (or circle, in two dimensions). Each model of hyperbolic space has its advantages and disadvantages. The projective model seems better suited for visualization and computer graphics, because geodesics appear "straight" and the isometries can be represented by projective linear transformations.

Matrix Formulas

The isometries of $\mathbf{H}^3$ correspond to the matrices in $O(3,1)$, just as the isometries of Euclidean 3-space correspond to the matrices in $O(4)$. We now present formulas for computing the matrices of rigid motions in hyperbolic space.

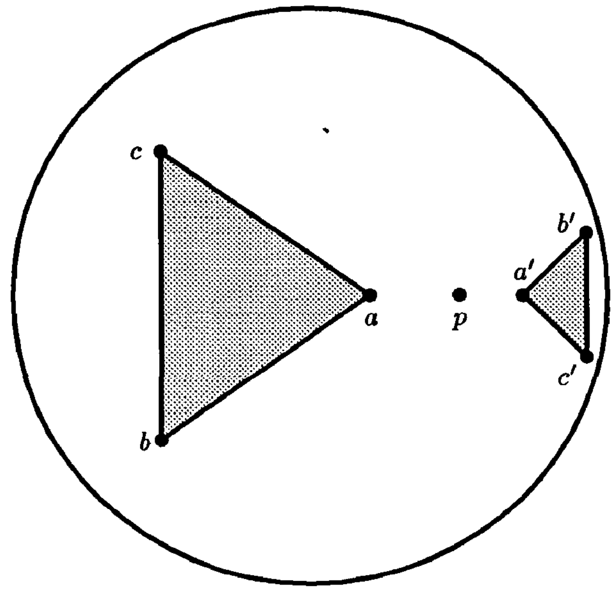

Figure 2: Hyperbolic Reflections. Triangle abc is the reflection of triangle $a'b'c'$ in point p. The two triangles are congruent in hyperbolic space, and hence would appear to be of equal size to an observer inside the space.

Reflections

One of the simplest types of isometries is a reflection. If $\mathbf{p}$ represents the homogeneous coordinates of a point p in $\mathbf{H}^3$, then the 4×4 matrix for the hyperbolic reflection in p is

$$\mathbf{r}_p^{\mathrm{hyp}} = \mathbf{I} - 2\mathbf{p}\mathbf{p}^T\mathbf{I}^{3,1}/\langle \mathbf{p},\mathbf{p}\rangle_h. \qquad (2)$$

This same formula may be used to obtain the matrix for the reflection in a plane as well. In this case, $\mathbf{p}$ represents the homogeneous coordinates of the plane.

Note: (2) can also be used to give the matrix for a Euclidean reflection, by replacing $\mathbf{I}^{3,1}$ with $\mathbf{I}$ and the Minkowski inner product with the dot product.

To use (2) in an example, let $p = (0.5, 0.0, 0)$, and consider the triangle with vertices $a = (0.2, 0.0, 0.0)$, $b = (-0.5, -0.5, 0.0)$, and $c = (-0.5, 0.5, 0.0)$ — see Figure 2. Then we can use the homogeneous coordinates $\mathbf{p} = \begin{pmatrix} 0.5 \\ 0 \\ 0 \\ 1 \end{pmatrix}$ to obtain

$$\mathbf{r}_p^{\mathrm{hyp}} = \begin{pmatrix} 1.666 & 0 & 0 & -1.333 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 1.333 & 0 & 0 & -1.666 \end{pmatrix}$$

To transform a point, say a, by this reflection, we multiply its homogeneous coordinates $\begin{pmatrix} 0.2 \\ 0 \\ 0 \\ 1 \end{pmatrix}$ by this matrix to obtain $\begin{pmatrix} -1 \\ 0 \\ 0 \\ -1.4 \end{pmatrix}$ and then normalize to obtain the point $a' = (0.714, 0, 0)$. Transforming b and c similarly gives $b' = (0.929, 0.214, 0)$, and $c' = (0.929, -0.214, 0)$.

Although the two triangles in 2 look very different from a Euclidean point of view, they are congruent in hyperbolic space. One may verify this by using (1) to compute the hyperbolic lengths of the triangles' edges. For example $\mathrm{d}^{\mathrm{hyp}}(a,b) = \mathrm{d}^{\mathrm{hyp}}(a',b') = 2.074$. (Be sure to use homogeneous coordinates in (1)!)

Translations

We can now define hyperbolic translations in terms of reflections. Just as in Euclidean space, the translation which takes a point a to a point b is the composition of the reflection in a with the reflection in the midpoint m of a and b:

$$\mathbf{T}^{\mathrm{hyp}}_{a,b} = \mathbf{r}^{\mathrm{hyp}}_{m} \cdot \mathbf{r}^{\mathrm{hyp}}_{a}. \tag{3}$$

The homogeneous coordinates $\mathbf{m}$ of the hyperbolic midpoint are given by the formula

$$\mathbf{m} = \mathbf{a}\sqrt{\langle \mathbf{b}, \mathbf{b}\rangle_h \langle \mathbf{a}, \mathbf{b}\rangle_h} + \mathbf{b}\sqrt{\langle \mathbf{a}, \mathbf{a}\rangle_h \langle \mathbf{a}, \mathbf{b}\rangle_h}, \tag{4}$$

where $\mathbf{a}$ and $\mathbf{b}$ are homogeneous coordinates for a and b, respectively.

As an example, consider the triangle from Figure 2 again. And let $b' = (0.3, -0.7, 0)$. We compute the matrix of translation $\mathbf{T}^{\mathrm{hyp}}_{b,b'}$. Using the homogeneous coordinates for b and b' in (4) gives $\mathbf{m} = \begin{pmatrix} -0.1 \\ -0.733 \\ 0 \\ 1.212 \end{pmatrix}$ for the midpoint. Using (2) and (3) then gives

$$\begin{pmatrix} 1.676 & 0.814 & 0 & 1.572 \\ -1.369 & 0.636 & 0 & -1.130 \\ 0 & 0 & 1 & 0 \\ 1.919 & 0.257 & 0 & 2.179 \end{pmatrix} \tag{5}$$

The images of a, b, and c under this transformation are $a' = (0.744, -0.548, 0)$, $b' = (0.3, -0.7, 0)$, and $c' = (0.846, -0.095, 0)$; see Figure 3.

To continue this example, we can translate b' again by (5) and obtain $b'' = (0.585, -0.771, 0)$, which lies on the line containing b and b'. The points b, b', and b'' lie at equally spaced intervals along this line in the hyperbolic metric.

An important fact about hyperbolic translations is that each has a unique axis. This is different from Euclidean translations, where it is only the direction of the axis that matters, not the particular choice of axis.

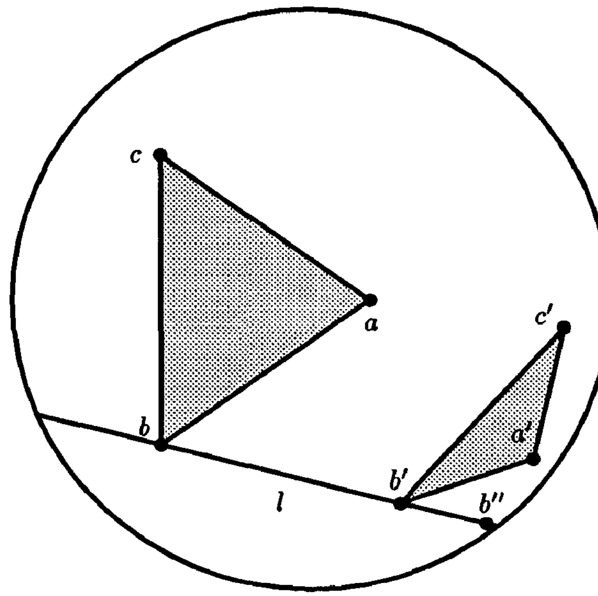

Figure 3: Hyperbolic Translation. Triangle $a'b'c'$ is obtained by translating triangle abc along line l from b to b'; the two triangles are congruent in hyperbolic space.

Rotations

A rotation of $\mathbf{H}^3$ about an axis l through the origin is the same as the Euclidean rotation about the same axis, since this rotation preserves the unit ball. To compute the matrix of rotation about an axis not passing through the origin, we first translate l the origin, do the rotation there, and then translate l back to its original position. The concatenation of these three transformations gives a rotation about the original axis. In order for the angles to work out right, we must translate along the unique line through the origin perpendicular to l. If l_0 is the point of l closest to the origin, this is the translation $\mathbf{T}^{\mathrm{hyp}}_{l_0,0}$.

Specifically, suppose a and b are points in $\mathbf{H}^3$ and we wish to rotate through an angle of θ about the line l through a and b. The point l_0 of l closest to the origin is given by

$$l_0 = \frac{a \cdot (a-b)}{(a-b)\cdot(a-b)} b + \frac{b \cdot (b-a)}{(b-a)\cdot(b-a)} a. \tag{6}$$

Note that in (6) a and b are the *affine* (not homogeneous) coordinates of points in $\mathbf{H}^3$, and $\cdot$ is the usual dot product. The desired hyperbolic rotation is then

$$\mathbf{R}^{\mathrm{hyp}}_{l,\theta} = (\mathbf{T}^{\mathrm{hyp}}_{l_0,0})^{-1} \cdot \mathbf{R}^{\mathrm{euc}}_{u,\theta} \cdot \mathbf{T}^{\mathrm{hyp}}_{l_0,0} \tag{7}$$

where $\mathbf{R}^{\mathrm{euc}}_{u,\theta}$ is the Euclidean rotation of $\mathbf{R}^3$ through an angle of θ about an axis in the direction of u, where

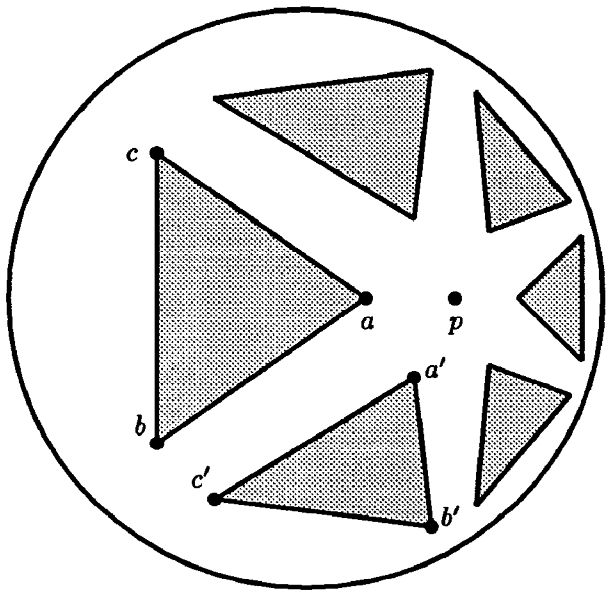

Figure 4: Hyperbolic Rotations. Triangle $a'b'c'$ is obtained by rotating triangle abc about the point p through and angle of $\pi/3$ radians. The other four triangles are obtained by additions rotations through the same angle. All six triangles are congruent in hyperbolic space.

$u = (a - b)/||a - b||$ is a unit vector in the direction of l. $\mathbf{R}^{\mathrm{euc}}_{u,\theta}$ is given by ([3], p. 73)

$$\begin{pmatrix} u_1^2 + c(1-u_1^2) & u_1u_2c_1 - u_3s & u_1u_3c_1 + u_2s & 0 \\ u_1u_2c_1 + u_3s & u_2^2 + c(1-u_2^2) & u_2u_3c_1 - u_1s & 0 \\ u_1u_3c_1 - u_2s & u_2u_3c_1 + u_1s & u_3^2 + c(1-u_3^2) & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

where $c = \cos(\theta)$, $s = \sin(\theta)$, and $c_1 = 1 - \cos(\theta)$.

To give another example using the above triangle, we compute the rotation about the line l through the points $p = (0.5, 0, 0)$ and $q = (0.5, 0, 1)$. This line is perpendicular to the x-y plane (in both the Euclidean and hyperbolic metrics) and hence this rotation preserves the x-y plane.

The point l_0 from (6) is, of course, just p. Using $u = (0, 0, 1)$ in (7) we obtain

$$\begin{pmatrix} 0.333 & -1 & 0 & 0.333 \\ 1. & 0.5 & 0 & -0.5 \\ 0. & 0 & 1 & 0. \\ -0.333 & -0.5 & 0 & 1.167 \end{pmatrix}. \tag{8}$$

The images of a, b, and c by this transformation are then $a' = (0.364, -0.273, 0)$, $b' = (0.421, -0.789, 0)$, and $c' = (-0.308, -0.692, 0)$. Figure 4 shows the resulting triangle, as well as the next five images under the transformation (8).

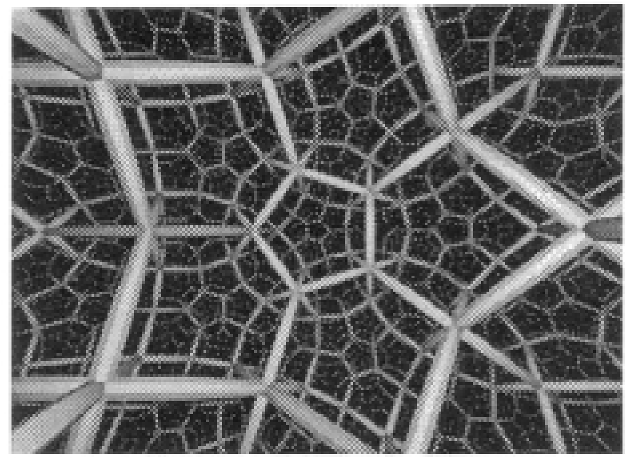

Figure 5: Scene from the video *Not Knot*. This scene shows a tesselation of hyperbolic space by regular right dodecahedra — analogous to a tesselation of Euclidean space by cubes.

Applications

Three recent projects at the Geometry Center have applied these ideas. One is the video *Not Knot* [4]. This video, whose purpose is to illustrate some of the basic concepts of knot theory and the theory of 3-manifolds, includes a fly-through scene of hyperbolic 3-space; see Figure 5. During this fly-through one easily notices that apparent size changes more rapidly in hyperbolic space than in Euclidean space. Angles appear to change as we move closer to them. In fact, however, they are not changing — what changes is our perception of them.

Another project which has used 4×4 matrix technology in this way is a flight simulator for hyperbolic space written by Linus Upson, a Princeton University undergraduate working as a research assistant during the summer of 1991. Patterned after the popular SGI flight simulator, Upson's program allows one to navigate through a scene in hyperbolic space; see Figure /ref-fig:hfly. The program is excellent for conveying a sense of how angles and distances seem to change with motion. The intuition which one gains from this experience is hard to pinpoint but extremely valuable in understanding hyperbolic geometry.

The third Geometry Center project using hyperbolic transformations is a general graphics library which we call the "Object Oriented Graphics Language" (OOGL), begun by Pat Hanrahan in the summer of 1989. This library provides a general framework in which geometric objects and the actions which operate on them may be specified arbitrarily. This makes it easy to define and manipulate objects in hyperbolic space. The interactive viewing program which accompanies OOGL (*MinneView*) has a "hyperbolic mode" in which the translations and rotations controlled by

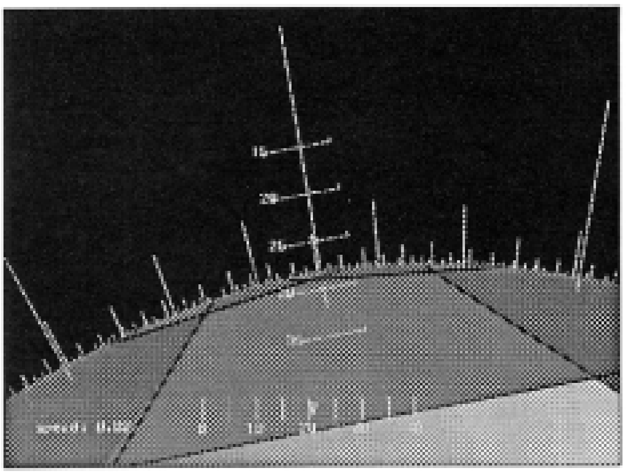

Figure 6: Hyperbolic space flight simulator. This scene shows the view from the cockpit of an airplane flying over a hyperbolic plane in hyperbolic 3-space. The plane is tesselated with regular right pentagons — it is essentially a copy of Figure 1.

mouse motions are hyperbolic rather than Euclidean. A version of this program for SGI IRIS workstations may be obtained on the Internet via anonymous ftp from host `geom.umn.edu` (IP address 128.101.25.31).

Acknowledgements

Figures 1, 2, 3, and 4 were generated by the program *Hypercad*, written Mark Phillips and Robert Miner. This program may be obtained on the Internet via anonymous ftp from host `geom.umn.edu` (IP address 128.101.25.31).

Figure 5 is a frame by Charlie Gunn from [4].

Figure 6 was generated by the hyperbolic flight simulator program written by Linus Upson.

References

[1] Beardon, Alan F. *The Geometry of Discrete Groups*, Springer-Verlag, New York, 1983. Chapter 7 contains many useful formulas for hyperbolic geometry and trigonometry.

[2] Coxeter, H.S.M. *Non-Euclidean Geometry*, University of Toronto Press, 1965. This book provides a comprehensive introduction to hyperbolic geometry, with an emphasis on its definition in terms of projective geometry.

[3] Faux, I.D. and M.J. Pratt. *Computational Geometry for Design and Manufacture*, Wiley and Sons, New York, 1979.

[4] Gunn, Charlie, et. al. "Not Knot" [videotape] Jones and Bartlett. Copies of this video may be ordered by contacting Jones and Bartlett Publishers, Inc, 20 Park Plaza, Suite 1435, Boston, MA 02116-9792.

[5] Thurston, William. *The Geometry and Topology of Three-Manifolds*, volume 1. Princeton University Press, to appear. Chapters 1 and 2 provide a good introduction to hyperbolic geometry.