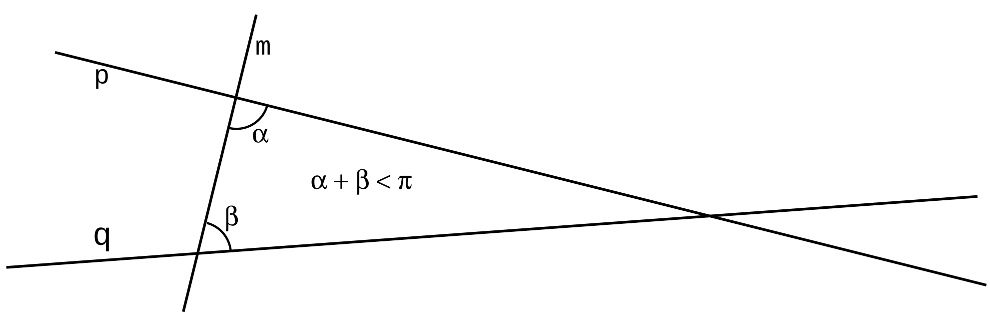
The Hyperbolic Plane and its Immersions into R^3

Andrejs Treibergs
University of Utah

Abstract. The hyperbolic plane is an example of a geometry where the first four of Euclid's Axioms hold but the fifth, the parallel postulate, fails and is replaced by a hyperbolic alternative. We discuss some basic properties of hyperbolic space, including how to coordinatize and measure lengths. We then consider the possibility of isometrically immersing the hyperbolic plane into $\mathbb{R}^3$ in such a way that lengths of curves are preserved. This is possible on small pieces, for example mapping to the pseudosphere. However, Hilbert's Theorem says it is impossible for the whole hyperbolic space.

Euclid's Postulates.

Euclid began with at least an intuitive description of *points, lines, rays, a line segment has two endpoints, straight lines, circles, angles, lengths, triangles* and the other sets and objects which are taken to be understood as usual. He tacitly assumes that points and lines exist, not all points are on the same line, two distinct lines have no more than one point in common, a straight line that contains the vertex B and an interior point of a triangle ABC also contains a point of the segment AC, things which are equal may be made to coincide (for example by a Euclidean motion, congruent figures are equal and conversely,) all sets of objects are finite, a line segment joining the center of a circle to a point outside the circle must contain a point of the circle (continuity,) a point on a line separates the line into two rays, and the existence of an order relation on the line. [M] These assumptions validate straightedge and compass constructions.



Euclid's Fifth Postulate

Euclid's Postulates are the following:

(1) any two points may be joined by a line segment;
(2) any line segment may be extended to form a line;
(3) a circle may be drawn with any given center and distance;
(4) any two right angles are equal;
(5) if a line m intersects two lines p, q such that the sum of the interior angles on the same side of m is less than two right angles, then the lines p and q intersect on the side of m on which the sum of the interior angles is less than the sum of the right angles.

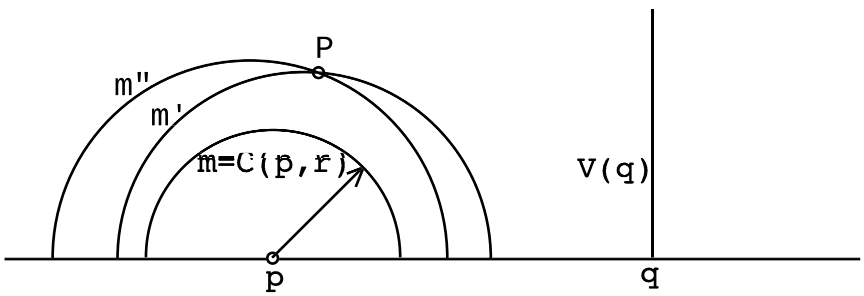
Starting from Euclid, mathematicians felt uneasy about the fifth postulate, and tried to prove it from the other four, or at least replace it with some more self evident statement. Two lines are parallel if they coincide or if they don't intersect. The tacit separation assumption is equivalent to: given three points on a line, one of the points is between the other two. Hence the fifth postulate implies that given a line and a point P not on the line, there exists a line through p parallel to the given line. The familiar equivalent form of the fifth postulate is called Playfair's Axiom:

($5'$) Given any line m and a point p, there is a unique line through p and parallel to m.

From letters found after his death, we now know that Gauß was first to realize in 1816 that a geometry may be constructed in which the Fifth Postulate fails. J. Bolyai and N. Lobachevski independently could prove by 1823 and 1826, resp., and eventually published in 1832 and 1826, resp., such non-Euclidean geometries. They assumed what amounts to Saccheri's Axiom. Saccheri tried to reach a contradiction from this axiom in an effort to prove Euclid's Fifth postulate.

($5''$) Given any line m and a point p not in m, there are at least two lines through p and parallel to m.

This axiom is also known as the hyperbolic axiom. In 1854, Riemann showed a consistent geometry may be constructed assuming instead that no lines are parallel.



m' and m'' are parallels to m through P.

There are various models of hyperbolic geometry. The Klein Model and the Poincaré Model start with the unit disk in the plane, identify certain subsets as lines and give formulas for distances and angles. We shall develop the Upper Halfplane Model. They are all equivalent.

For the space we take the open upper halfplane

$$\mathbb{H}^2 = \{(x, y) \subset \mathbb{R}^2 : y > 0\}.$$

A point can thus be identified with its coordinates (x, y) where $y > 0$. For lines on $\mathbb{H}^2$ we take the collection of all semicircles and vertical lines

$$\mathcal{L} = \{C(p, r) : p \in \mathbb{R}, r > 0\} \cup \{V(q) : q \in \mathbb{R}\}$$

where the set $C(p, r) = \{(x, y) : (x - p)^2 + y^2 = r^2, y > 0\}$ is a semicircle centered on the x-axis, and $V(q) = \{(q, y) : y > 0\}$ is the vertical line at q. The axioms are readily verified. For example, the hyperbolic axiom holds because given a line, say $C(p, r)$, and a point not on the line, say (x, y) with $R^2 = (x-p)^2 + y^2 > r^2$, then $(x, y) \in C(p', r')$ whenever $(x-p')^2 + y^2 = (r')^2$ for any p'. $C(p', r')$ is disjoint from $C(p, r)$ whenever $r' > |p - p'| + r$ which is satisfied for $p' = p$ and p' near p. This is just the triangle inequality: $Q \in C(p', r')$ implies $|Q - (p, 0)| \geq |Q - (p', 0)| - |(p', 0) - (p, 0)| = r' - |p' - p| > r$. The other cases are similar.

Lengths of curves are determined by a Riemannian metric.

$$ds^2 = \frac{dx^2 + dy^2}{y^2}. \tag{1}$$

This means, if we take a parameterized curve $\gamma(t) = (x(t), y(t))$ where x and y are piecewise continuously differentiable for $a \le t \le b$, then the length along the curve is determined by integrating the metric

$$\mathrm{L}(\gamma) = \int_\gamma ds = \int_a^b |\dot\gamma(t)|_{\gamma(t)}\,dt = \int_a^b \frac{\sqrt{\dot x(t)^2 + \dot y(t)^2}}{y(t)}\,dt.$$

Dot means to differentiate with respect to the parameter. Note that this expression is indepedent of choice of parameter. If σ is an increasing function $\dot\sigma > 0$ such that $\sigma(c) = a$ and $\sigma(d) = b$ then by the change of variables formula for integrals (substitution $t = \sigma(\tau)$) we have for a reparameterized curve $\psi(\tau) = \gamma(\sigma(\tau))$,

$$\begin{aligned}\mathrm{L}(\gamma) = \int_{t=a}^b \frac{\sqrt{x_t^2 + y_t^2}}{y}\,dt &= \int_{\tau=c}^d \frac{\sqrt{x_t(\sigma(\tau))^2 + y_t(\sigma(\tau))^2}}{y(\sigma(\tau))}\,\dot\sigma(\tau)\,d\tau = \\ &= \int_c^d \frac{\sqrt{\left(x_t\dot\sigma\right)^2 + \left(y_t\dot\sigma\right)^2}}{y(\sigma(\tau))}\,d\tau = \int_c^d \frac{\sqrt{x_\tau^2 + y_\tau^2}}{y}\,d\tau = \mathrm{L}(\psi).\end{aligned}$$

For this reason, the parameter need not appear in formula (1).

The formula (1) enables us to define an inner product on tangent vectors at a point $(x, y) \in \mathbb{H}^2$. For example, if $V = (v^1(x, y), v^2(x, y))$ and $W = (w^1(x, y), w^2(x, y))$ are the components of the tangent vector at (x, y), then the metric gives an inner product by the formula

$$\langle V, W\rangle_{(x,y)} = \frac{v^1(x, y)w^1(x, y) + v^2(x, y)v^2(x, y)}{y^2} = \frac{V \bullet W}{y^2}.$$

The length of a tangent vector is given by

$$|V|_{(x,y)} = \sqrt{\langle V, V\rangle_{(x,y)}} = \frac{\sqrt{(v^1(x, y))^2 + (v^2(x, y))^2}}{y}.$$

As the inner product is the same as the dot product multiplied by a function y^{-2}, it means that the angle as given by the cosine law is the same in $\mathbb{H}^2$ as it is in the underlying space $\mathbb{R}^2$. Thus if θ is $\angle V, W$ in $\mathbb{H}^2$ and $\theta_{\text{Eucl.}}$ is the usual angle in $\mathbb{R}^2$ then, using the cosine law,

$$\cos\theta = \frac{\langle V, W\rangle_{(x,y)}}{\sqrt{\langle V, V\rangle_{(x,y)}}\sqrt{\langle W, W\rangle_{(x,y)}}} = \frac{V \bullet W}{\sqrt{V \bullet V}\sqrt{W \bullet W}} = \cos\theta_{\text{Eucl.}}.$$

It turns out that the lines are geodesics in this metric. That means that given a pair of points $P, Q \in \mathbb{H}^2$, then the line segment PQ determined by these points has the shortest length among all piecewise continuously differentialble curves that connect P to Q. This is easy to see on a vertical line segment, say $\nu(t) = (p, t)$ where $0 < a \le t \le b$. That is because if say $\gamma(t) = (f(t), t)$ is any other piecewise continuously differentiable curve with$\gamma(a) = \nu(a)$ and $\gamma(b) = \nu(b)$, then $\dot f^2 + 1 \ge 1$ implies

$$\mathrm{L}(\gamma) = \int_a^b \frac{\sqrt{\dot f(t)^2 + 1}}{t}\,dt \ge \int_a^b \frac{dt}{t} = \mathrm{L}(\nu) = \ln\left(\frac{b}{a}\right). \tag{2}$$

If γ moved up and down or dipped below ν then it would have been even longer. To show that this property holds for all line segments, we have to discuss isometries of $\mathbb{H}^2$. Any segment can be carried to a vertical one by such a rigid motion of $\mathbb{H}^2$ so that all curves are carried to curves of the same length.

Properties of the Hyperbolic Plane.

An *isometric mapping* is a mapping that preserves the lengths of curves. For example, if we have two identical sheets of paper, one is flat and the other is rolled up, then the map that identifies corresponding points from the flat page with the rolled up page preserves lengths of curves on the paper, although some points may be mapped to the same point. (In some other contexts, an isometric map might be required to preserve distances between pairs of points, but that is not the notion we are discussing.) If the inverse map F^{-1} is also an isometric map, we call F an *isometry.* Consider a continuously differentiable mapping $F : \mathbb{H}^2 \to \tilde{\mathbb{H}}^2$ which may be regarded as a self map or a map between identical copies. If we denote the coordinates $(\tilde{x}, \tilde{y}) = F(x, y)$ then let us find the equations that F must satisfy to be an isometric map. Suppose that $\gamma : [a, b] \to \mathbb{H}^2$ is a piecewise continuously differentiable curve. Then $F \circ \gamma : [a, b] \to \tilde{\mathbb{H}}^2$ is then the corresponding piecewise continuouly differentiable curve in the target space (the *push-forward* of the curve γ.) We can measure its length in $\tilde{\mathbb{H}}^2$ to get

$$\tilde{\mathrm{L}}(F \circ \gamma) = \int_a^b \frac{\sqrt{\left(\frac{\partial \tilde{x}}{\partial x}\dot{x}(t) + \frac{\partial \tilde{x}}{\partial y}\dot{y}(t)\right)^2 + \left(\frac{\partial \tilde{y}}{\partial x}\dot{x}(t) + \frac{\partial \tilde{y}}{\partial y}\dot{y}(t)\right)^2}}{\tilde{y}}\, dt$$

We say that F is an isometric map if lengths are preserved, $\mathrm{L}(\gamma) = \tilde{\mathrm{L}}(F \circ \gamma)$ for all γ. This implies in the notation of (1) that the metric of $\mathbb{H}^2$ equals the *pulled back* metric of $\tilde{\mathbb{H}}^2$

$$\frac{dx^2 + dy^2}{y^2} = \frac{\left(\frac{\partial \tilde{x}}{\partial x}(x, y)\, dx + \frac{\partial \tilde{x}}{\partial y}(x, y)\, dy\right)^2 + \left(\frac{\partial \tilde{y}}{\partial x}(x, y)\, dx + \frac{\partial \tilde{y}}{\partial y}(x, y)\, dy\right)^2}{\tilde{y}^2(x, y)} = \frac{d\tilde{x}^2 + d\tilde{y}^2}{\tilde{y}^2}. \tag{3}$$

The rigid motions of Euclidean space, translation $(x, y) \mapsto (x + a, y + b)$, reflection $(x, y) \mapsto (-x, y)$ and rotation $(x, y) \mapsto (\cos\alpha\, x - \sin\alpha\, y, \sin\alpha\, x + \cos\alpha\, y)$, where a, b, α are constants give isometries since they preserve lengths of all curves.

Let us describe some basic examples on $\mathbb{H}^2$. These formulas are a little cumbersome in the upper halfplane model and may be easier in some other models. Also using complex arithmetic with $z = x + iy$ may facilitate computation, but we don't do it this way since this may not be familiar. The translation $(\tilde{x}, \tilde{y}) = T_u(x, y) = (x + u, y)$, where $u \in \mathbb{R}$ satisfies $\tilde{y} = y$, $d\tilde{x} = dx$ and $d\tilde{y} = dy$ so (3) holds. The dilation $D_\lambda(x, y) = (\lambda x, \lambda y)$ for $\lambda > 0$ constant satisfies $\tilde{y} = \lambda y$, $d\tilde{x} = \lambda\, dx$ and $d\tilde{y} = \lambda\, dy$ so

$$\frac{d\tilde{x}^2 + d\tilde{y}^2}{\tilde{y}^2} = \frac{\lambda^2\, dx^2 + \lambda^2\, dy^2}{\lambda^2\, y^2} = \frac{dx^2 + dy^2}{y^2},$$

is an isometry of $\mathbb{H}^2$ which is not an isometry of Euclidean space. Similarly the reflection along the y-axis $O(x, y) = (-x, y)$ is an isometry. The most interesting, also without a Euclidean analog, is the inversion with respect to the unit circle at the origin.

$$I(x, y) = \frac{(x, y)}{x^2 + y^2}.$$

We compute

$$\tilde{y} = \frac{y}{x^2 + y^2}, \qquad d\tilde{x} = \frac{(y^2 - x^2)\, dx - 2xy\, dy}{(x^2 + y^2)^2} \qquad d\tilde{y} = \frac{-2xy\, dx + (x^2 - y^2)\, dy}{(x^2 + y^2)^2}$$

so

$$\frac{d\tilde{x}^2 + d\tilde{y}^2}{\tilde{y}^2} = \frac{[(y^2 - x^2)\, dx - 2xy\, dy]^2 + [-2xy\, dx + (x^2 - y^2)\, dy]^2}{(x^2 + y^2)^4 \cdot \left(\frac{y}{x^2 + y^2}\right)^2} = \frac{dx^2 + dy^2}{y^2}.$$

Observe that both I and O reverse orientations. All of these isometries are invertible: $T_u^{-1} = T_{-u}$, $D_\lambda^{-1} = D_{1/\lambda}$, $O^{-1} = O$, $I^{-1} = I$. They generate a group of isometries $\mathcal{G} =< \{T_u, D_\lambda, I, O : u \in \mathbb{R},\ \lambda > 0\} >$ consting of all finite compositions of these operations.

To prove that a subarc $C \subset C(p,r)$ is geodesic, we have to exhibit an isometry $F : \mathbb{H}^2 \to \mathbb{H}^2$ taking $C(p,r)$ to $V(0)$. Then if γ is any other curve spanning the endpoints of C then $F(\gamma)$ is a curve spanning the endpoints of $F(C)$ and we get $\mathrm{L}(\gamma) = \mathrm{L}(F(\gamma)) \ge \mathrm{L}(F(C)) = \mathrm{L}(C)$ because F preserves lengths.

Thus we describe the actions of the generators. The translation evidently maps $T_u(V(p)) = V(p+u)$ and $T_u(C(p,r)) = C(p+u,r)$. Similarly the dilation maps $D_\lambda(V(p)) = V(\lambda p)$ and $D_\lambda(C(p,r)) = C(\lambda p, \lambda r)$ as well as $O(V(p)) = V(-p)$ and $O(C(p,r)) = C(-p,r)$. The inversion is more interesting. We have $I(V(0)) = V(0)$ and if $p \ne 0$ then $I(V(p)) = C\left(\frac{1}{2p}, \frac{1}{2|p|}\right)$ and also $I(C(p,|p|)) = V\left(\frac{1}{2p}\right)$. If $p \notin \{\pm r\}$ then $I(C(p,r)) = C\left(\frac{p}{p^2-r^2}, \frac{r}{|p^2-r^2|}\right)$. Inversion of a semicircle with one end at the origin results in a vertical geodesic. Otherwise inversion maps circles to circles. These identities are tedious algebraic verifications. For example, to check $I(C(p,r))$ we choose $(x,y) \in C(p,r)$ so $(x-p)^2 + y^2 = r^2$ or $x^2+y^2 = r^2 - p^2 + 2xp$ implies $(\tilde{x}, \tilde{y}) = I(x,y)$ satisfies

$$\begin{aligned}
&\left(\tilde{x} - \frac{p}{p^2-r^2}\right)^2 + \tilde{y}^2 = \left(\frac{x}{x^2+y^2} - \frac{p}{p^2-r^2}\right)^2 + \left(\frac{y}{x^2+y^2}\right)^2 = \\
&\frac{\left[x(p^2-r^2) - p(x^2+y^2)\right]^2 + y^2(p^2-r^2)^2}{(x^2+y^2)^2(p^2-r^2)^2} = \frac{(x^2+y^2)(p^2-r^2)^2 - 2xp(x^2+y^2)(p^2-r^2) + p^2(x^2+y^2)^2}{(x^2+y^2)^2(p^2-r^2)^2} \\
&\qquad = \frac{(p^2-r^2)^2 - 2xp(p^2-r^2) + p^2(r^2-p^2+2xp)}{(x^2+y^2)(p^2-r^2)^2} = \frac{p^4 - 2p^2r^2 + r^4 + 2xpr^2 + p^2r^2 - p^4}{(x^2+y^2)(p^2-r^2)^2} \\
&\qquad = \frac{r^2(-p^2+r^2+2xp)}{(x^2+y^2)(p^2-r^2)^2} = \frac{r^2}{(p^2-r^2)^2}.
\end{aligned}$$

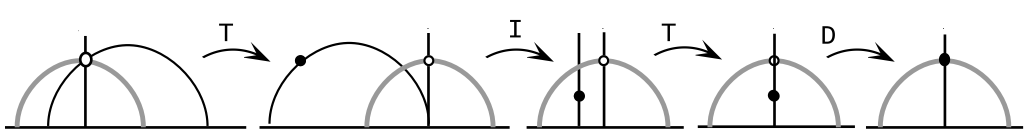
Any point can be taken to any other, for example translating (x,y) to the axis and dilating by $1/y$ moves any point to $(0,1)$ (that is $D_{1/y} \circ T_{-x}(x,y) = (0,1)$) and therefore any point to any point. We can also rotate around any point. To see this, we wish to move $C(p, \sqrt{p^2+1})$ which contains $(0,1)$ to the vertical axis. To do this, we translate by $-p - \sqrt{p^2+1}$ to move the endpint to the origin, invert to map it to $V\left(\frac{-1}{2\sqrt{p^2+1}}\right)$, translate by $\frac{1}{2\sqrt{p^2+1}}$ to put it in $V(0)$, then dilate by $\frac{1}{2\sqrt{p^2+1}(p+\sqrt{p^2+1})}$ to make sure that the $(0,1)$ is fixed. Let us call the inverse map

$$R_p = D_{2\sqrt{p^2+1}(p+\sqrt{p^2+1})} \circ T_{\frac{1}{2\sqrt{p^2+1}}} \circ I \circ T_{p+\sqrt{p^2+1}}.$$

The resulting formula is

$$R_p(x,y) = \frac{p+\sqrt{p^2+1}}{(x-p-\sqrt{p^2+1})^2+y^2}\left(x^2 - 2xp + y^2 - 1, 2y\sqrt{p^2+1}\right). \tag{4}$$

One checks that $R_p(0,1) = (0,1)$, $R_p(V(0)) = C(p, \sqrt{p^2+1})$ and that $O \circ R_p$ rotates the tangent vectors at $(0,1)$ by a (counterclockwise) angle θ where $\cot\theta = p$.



Sequence of isometries whose composition fixes $[0,1]$.

Given any arc $C \subset C(q,r)$ or $C \subset V(a)$ with endpoint (a,b) we may move the endpoint to $(0,1)$ as before via $D_{1/b} \circ T_{-a}$ and rotate around $(0,1)$ with appropriate p so that $F = R_p \circ D_{1/b} \circ T_{-a}$ satisfies $F(C(q,r)) = V(0)$ as desired.

Let us describe the (intrinsic) distance function on $\mathbb{H}^2$. Given any pair of points $P, Q \in \mathbb{H}^2$, the distance

$$\operatorname{dist}(P,Q) = \inf_{\substack{\gamma \text{ is a piecewise} \\ C^1 \text{ curve from } P \text{ to } Q}} \mathrm{L}(\gamma).$$

Since we have already shown that the shortest length is achieved for the geodesics, we can work out the formula by moving the points to $V(0)$ and measuring the length there, or we may simply compute the length of the geodesic connecting the two points. For example, suppose that both points are on the same line $P, Q \in C(p,r)$, say $P = (p + r\cos\theta_1, r\sin\theta_1)$ and $Q = (p + r\cos\theta_2, r\sin\theta_2)$ then the length minimizing curve between them is $\gamma(t) = (p + r\cos t, r\sin t)$ where $\theta_1 \le t \le \theta_2$ so

$$\operatorname{dist}(P,Q) = \int_{\theta_1}^{\theta_2} |\dot\gamma|_\gamma \, dt = \int_{\theta_1}^{\theta_2} \frac{r\,dt}{r\sin t} = \left| \ln\left(\cot\frac{\theta_2}{2}\right) - \ln\left(\cot\frac{\theta_1}{2}\right) \right|,$$

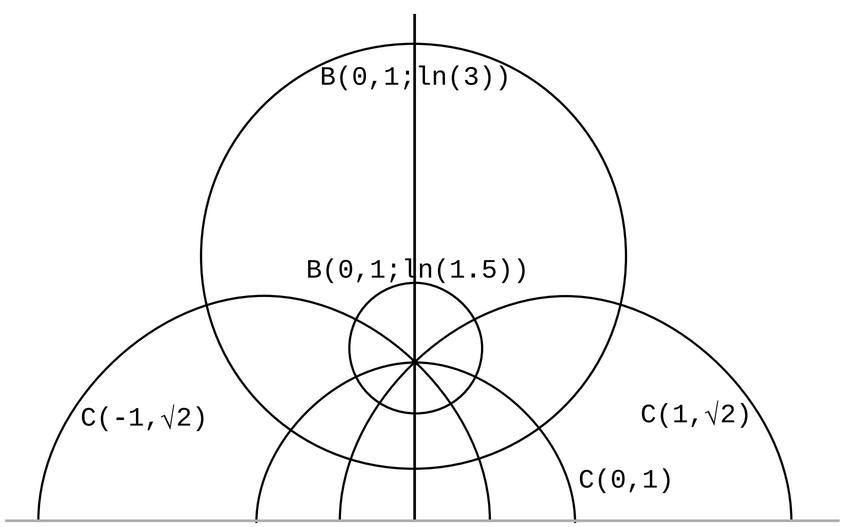
which is independent of r as expected, since dilation is an isometry.

Next we consider the geodesic disks of radius ρ and center $P = (x_0, y_0)$. Let $B(P;\rho) = \{Q : \operatorname{dist}(P,Q) < \rho\}$ be the set of points which are at most a distance ρ from P. It turns out that $B(P,\rho)$ are $\mathbb{R}^2$ circles, but the hyperbolic and Euclidean centers differ. For example, if $P = (0,1)$, and $b = e^\rho$ then

$$B(0,1;\rho) = \left\{ (x,y) : x^2 + \left(y - \frac{1}{2}\left(b + \frac{1}{b}\right)\right)^2 < \frac{1}{4}\left(b - \frac{1}{b}\right)^2 \right\}.$$

This is a circle with center above $(0,1)$. According to (2), the diameter is $2\rho = \ln b^2$. To see that this is the case, we can figure out the trajectory of $(\tilde x, \tilde y) = R_p(0,b)$ as $p \in \mathbb{R}$. From (4), computation shows

$$\frac{(p + \sqrt{p^2+1})^2 (b^2 - 1)^2}{\left((p + \sqrt{p^2+1})^2 + b^2\right)^2} + \left(\frac{2b(p + \sqrt{p^2+1})\sqrt{p^2+1}}{(p + \sqrt{p^2+1})^2 + b^2} - \frac{1}{2}\left(b + \frac{1}{b}\right)\right)^2 = \frac{1}{4}\left(b - \frac{1}{b}\right)^2.$$



Geodesic disks around $[0,1]$.

Let us compute the length of the boundary and area of $B = B(0,1;\rho)$. Parameterizing the boundary curve ∂B

$$x = r\cos t, \qquad y = c + r\sin t, \quad \text{where} \quad c = \frac{1}{2}\left(b + \frac{1}{b}\right) = \cosh\rho \qquad r = \frac{1}{2}\left(b - \frac{1}{b}\right) = \sinh\rho,$$

we find $\dot{x} = -r\sin t$ and $\dot{y} = r\cos t$. Since $c^2 - r^2 = 1$, $c + r = b$, $c - r = \frac{1}{b}$,

$$\begin{aligned}\mathrm{L}(\partial B) &= \int_0^{2\pi} \frac{r\,dt}{c + r\sin t} = \frac{4r}{\sqrt{c^2 - r^2}}\left[\mathrm{Atn}\left(\frac{r+c}{\sqrt{c^2-r^2}}\right) - \mathrm{Atn}\left(\frac{r-c}{\sqrt{c^2-r^2}}\right)\right] \\ &= 2\left(b - \frac{1}{b}\right)\left(\mathrm{Atn}(b) + \frac{\pi}{2} - \mathrm{Atn}(b)\right) = \pi\left(b - \frac{1}{b}\right) = 2\pi\sinh\rho.\end{aligned}$$

Similarly we compute the area. Note that using Green's, theorem

$$\begin{aligned}\mathrm{Area}(B) &= \iint_B \frac{dx\,dy}{y^2} = \oint_\gamma \frac{dx}{y} = \int_0^{2\pi} \frac{-r\sin t\,dt}{c + r\sin t} = \int_0^{2\pi} \frac{(c - c - r\sin t)\,dt}{c + r\sin t} \\ &= \frac{c}{r}\,\mathrm{L}(\partial B) - 2\pi = \pi\left(b + \frac{1}{b}\right) - 2\pi = 2\pi(\cosh\rho - 1).\end{aligned}$$

Hyperbolic space is much larger than Euclidean, and here is one way to see the difference. A ball in Euclidean space has area $\pi\rho^2$, and the length of its boundary circle is $2\pi\rho$. The fact that even infinitessimally, the disk grows faster than the Euclidean disk is a measure of the curvature of the space. We may use the discrepancy from Euclidean to define the curvature at a point.

$$K(P) = \lim_{\rho\to 0} \frac{3\big[2\pi\rho - \mathrm{L}(\partial B(P,\rho))\big]}{\pi\rho^3}. \tag{5}$$

A negative curvature indicates that in the vicinity of a point, little disks grow faster than disks in $\mathbb{R}^2$. Since $\sinh\rho = \rho + \rho^3/6 + \cdots$, we find that the curvature at $P = (0,1)$ on $\mathbb{H}^2$ is

$$K(P) = \lim_{\rho\to 0} \frac{3\big[2\pi\rho - 2\pi\sinh\rho\big]}{\pi\rho^3} = \lim_{\rho\to 0} \frac{6\big[\rho - \left(\rho + \frac{1}{6}\rho^3 + \cdots\right)\big]}{\rho^3} = -1.$$

Since there is an isometry that maps the neighborhood of any point to the neighborhood of $(0,1)$, the lengths of bounding circles are everywhere the same. Thus the curvature of hyperbolic space is identically constant $K \equiv -1$.

Curvature can be defined through derivatives of the metric, in which case the length formula (5) would be a consequence. There is a related formula comparing area growth. See any book discussing intrinsic differential geometry, *e. g.*, [BL], [dC], [S], [W].

Isometric immersions and the Pseudosphere.

To get an intuitive grasp of how this metric behaves, it is natural to wonder if there are surfaces in $\mathbb{R}^3$ which are isometric to $\mathbb{H}^2$. It was Beltrami in 1865 who first suspected that pieces on non-Euclidean geometries can be realized as curved surfaces in $\mathbb{R}^3$. We are looking for a local immersions $X : U \to \mathbb{R}^3$ where X is a continuously differentiable function and $U \subset \mathbb{H}^2$ is an open subset. To be an immersion means that locally there is no degeneration. If $P = (x,y) \in U$ then we require that the derivative vectors $X_x(x,y)$ and $X_y(x,y)$ are linearly independent. By continuity, these derivatives are independent vectors in a neighborhood $V \subset U$ of (x,y). By the implicit function theorem, $X(V)$ is then a two dimensional curved surface in $\mathbb{R}^3$, such that at each point of V, the vectors X_x and X_y are tangent to the surface. It may happen that the image $X(U)$ self intersects, but that does not happen on small pieces V.

Suppose that $(\tilde{x}, \tilde{y}, \tilde{z}) = X$ is an isometry, but this time, the metric of $\mathbb{R}^3$ is assumed to be the Euclidean metric

$$d\tilde{s}^2 = d\tilde{x}^2 + d\tilde{y}^2 + d\tilde{z}^2.$$

Note that an isometric image of a geodesic disk $B \subset \mathbb{H}^2$ will have to scrunch up a lot because it will have to fit inside a Euclidean ball of the same radius, to preserve radial lengths. That is, if X is an isometric map, then

$$X(B(P,\rho)) \subset B_{\text{Eucl.}}(X(P),\rho),$$

where $B_{\text{Eucl.}}(P,\rho) = \{\tilde{Q} \in \mathbb{R}^3 : |\tilde{Q} - X(P)|_{\text{Eucl.}} < \rho\}$ is the usual ρ ball in $\mathbb{R}^3$. The boundary of $X(B(P,\rho))$ has exponentially long length, whereas $B_{\text{Eucl.}}(X(P),\rho)$ grows polynomially for large ρ.

To simplify matters, suppose that we would like to find a surface of revolution which is isometric to a piece of $\mathbb{H}^2$. Suppose there are functions $f(y)$ and $g(y)$ and $\theta(x)$ so that the map takes the form

$$X(x,y) = (f(y)\cos\theta(x), f(y)\sin\theta(x), g(y)).$$

This map is nondegenerate or *regular* if $X_x = (-f\dot{\theta}\sin\theta, f\dot{\theta}\cos\theta, 0)$ and $X_y = (\dot{f}\cos\theta, \dot{f}\sin\theta, \dot{g})$ are independent, which happens when $f \neq 0$, $\dot{\theta} \neq 0$ and $\dot{f}^2 + \dot{g}^2 \neq 0$. We shall try to find f, g, θ defined in a region $U = \{(x,y) \in \mathbb{H}^2 : y > b\}$ where $b > 0$. The translations T_u move points of U just like the rotations around the z-axis are isometries of $\mathbb{R}^3$ which preserve the surface of rotation $X(U)$. Thus the equations for isometric immersion are

$$\begin{aligned}\frac{dx^2+dy^2}{y^2} &= d\tilde{x}^2 + d\tilde{y}^2 + d\tilde{z}^2 = \\ &= (-f\sin(\theta)\dot{\theta}\,dx + \dot{f}\cos(\theta)\,dy)^2 + (f\cos(\theta)\dot{\theta}\,dx + \dot{f}\sin(\theta)\,dy)^2 + \dot{g}^2\,dy^2 = f^2\dot{\theta}^2\,dx^2 + (\dot{f}^2+\dot{g}^2)\,dy^2.\end{aligned}$$

It follows that

$$\frac{1}{y^2} = f(y)^2\,\dot{\theta}(x)^2, \qquad \frac{1}{y^2} = \dot{f}^2(y) + \dot{g}^2(y).$$

By separation of variables, the first equation implies that there is a constant $c_1 > 0$ so that

$$\frac{1}{y^2 f(y)^2} = \dot{\theta}(x)^2 = c_1^2.$$

It follows that $\dot{\theta} = c_1$ so for another constant c_2, $\theta = c_1 x + c_2$. It also follows then that

$$f(y) = \pm\frac{1}{c_1 y}.$$

Note that for $y > 1/c_1$, $0 < f < 1$. Substituting into the second equation yields

$$\dot{g}^2 = \frac{1}{y^2} - \dot{f}^2(y) = \frac{1}{y^2} - \frac{1}{c_1^2 y^4} \qquad \text{or} \qquad \frac{dg}{dy} = \pm\frac{\sqrt{c_1^2y^2-1}}{c_1 y^2}.$$

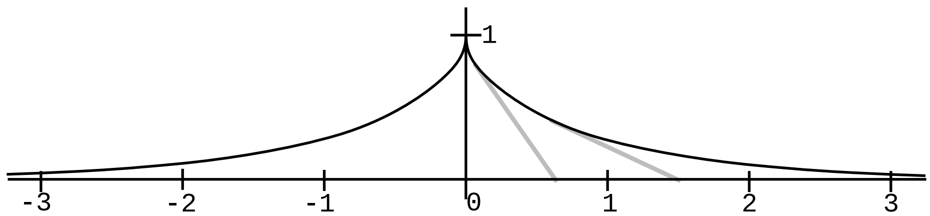
Integrating this equation gives, for some constant

$$\pm g(y) = \ln\left(\sqrt{c_1^2y^2-1} + c_1 y\right) - \frac{\sqrt{c_1^2y^2-1}}{y} + c_3$$

This is only defined for large y so we must have $y > b = \frac{1}{c_1}$. Notice that dilation and translation say that all choices of c_1 and c_2 are equivalent, so we might as well take $c_1 = 1$ and $c_2 = 0$. If we fix vertical translation in $\mathbb{R}^3$ by setting $c_3 = 0$, then the image $X(U)$ us generated by revolving around the $\tilde{z}$-axis the curve

$$(\pm\tilde{z}, \tilde{y}) = \left(\ln\left(\sqrt{y^2-1} + y\right) - \frac{\sqrt{y^2-1}}{y}, \frac{1}{y}\right). \tag{6}$$

The resulting surface of revolution is called the *pseudosphere*. The curve (6) is called the *tractrix*, which is the path followed by a reluctant little dog on a unit length leash, that starts at $(\tilde{z}, \tilde{y}) = (0, 1)$ and whose owner walks along the $\tilde{z}$-axis. One way to see this is to compute the distance along the tangent line from a point on the curve to the $\tilde{z}$-axis.



Tractrix.

Notice that the curve is convex upward whereas the revolution is convex downward. The surface has to be on both sides of its tangent plane near the point because it has to have larger area locally than its tangent plane.

The second fundamental form and a geometric interpretation of curvature.

We describe an extrinsic geometric interpretation of curvature for arbitrary surfaces in Euclidean space. Suppose we're given a parametric surface locally by $X(u^1, u^2)$ near the point P. The tangent plane to $X(M)$ at point $X(u^1, u^2)$ is spanned by the tangent vectors X_1 and X_2. By applying the Gram-Schmidt algorithm to the vector functions, it is possible to find orthonormal vector fields E_1, E_2 that span the tangent space at $X(u^1, u^2)$ and which vary in a C^1 fashion. We can also let $E_3 = E_1 \times E_2$ be the unit vector field normal to the surface. Since the surface is regular, it can be represented as a graph over the tangent plane, so for each P, we may write $X(M)$ as a graph over the tangent plane near P as

$$X(\xi^1, \xi^2) = X(p^1, p^2) + \xi^1 E_1(p^1, p^2) + \xi^2 E_2(p^1, p^2) + f(\xi^1, \xi^2; p^1, p^2) E_3(p^1, p^2).$$

Since E_1 and E_2 are tangent to $X(M)$ at P, $f_1(0, 0; P) = f_2(0, 0; P) = 0$ (at the point $X(P)$) the *second fundamental form* is defined to be the Hessian matrix $h_{ij}(P) = \dfrac{\partial^2 f}{\partial \xi^i \, \partial \xi^j}(0, 0; P)$. The second fundamental form may be regarded as a bilinear form acting on tangent vectors at P, so that of $V = v^1 E_1 + v^2 E_2$ and $W = w^1 E_1 + w^2 E_2$ then

$$II(V, W) = \sum_{i,j=1}^{2} h_{ij} v^i w^j.$$

The mean curvature is half the trace $H = \frac{1}{2}(h_{11} + h_{22}) = \frac{1}{2}(\kappa_1 + \kappa_2)$ and the Gauß curvature is the determinant $K = \det(h_{ij}) = \kappa_1 \kappa_2$, where κ_i are the eigenvalues of h_{ij} at P. These numbers are called the principal curvatures. Because H and K are symmetric functions of eigenvalues, they are defined independently of the choice of the orthonormal basis at P. Thus H and K are invariantly defined quantities of the surface.

It is a basic theorem of Gauß that the Gauß curvature is the same as the curvature defined by (5). If two surfaces are locally isometric, then they have the same Gauß curvature. In other words, one can compute K from the metric alone since curvature depends only on the lengths of curves near the point.

Hilbert's Impossibility Theorem.

David Hilbert, the leading mathematician at the dawn of the twentieth century, proved two fundamental theorems about constant curvature surfaces. The first was for compact surfaces: the only compact, boundaryless surfaces of constant Gauß curvature in $\mathbb{R}^3$ are the spheres. The second is that there are no complete, immersed surfaces of class $\mathcal{C}^4$ with $K \equiv -1$ [H]. We state a version for isometric immersions of the hyperbolic plane. Our discussion is sketchy. For details see [dC], [S], [W].

Theorem. [Hilbert, 1901] *There is no regular smooth isometric immersion* $X : \mathbb{H}^2 \to \mathbb{R}^3$.

Idea of the proof. Let's suppose that there were such an immersion and deduce a contradiction. Suppose the smooth map $X : \mathbb{H}^2 \to \mathbb{R}^3$ gives an isometric immersion. Corresponding to the point $P \in \mathbb{H}^2$ we consider its image $X(P)$ and the tangent plane to the surface at that point. Because the curvature of the surface has to be the same as hyperbolic space $K = -1$, it follows that the second fundamental form is negative definite. It follows that there are two directions in which the second fundamental form vanishes, called *asymptotic directions.* Thus we can choose two linearly independent unit tangent vectors V_1 and V_2 at P and their corresponding vectors $\tilde{V}_i = dX(V_i)$ at $X(P)$ so that at the point in question,

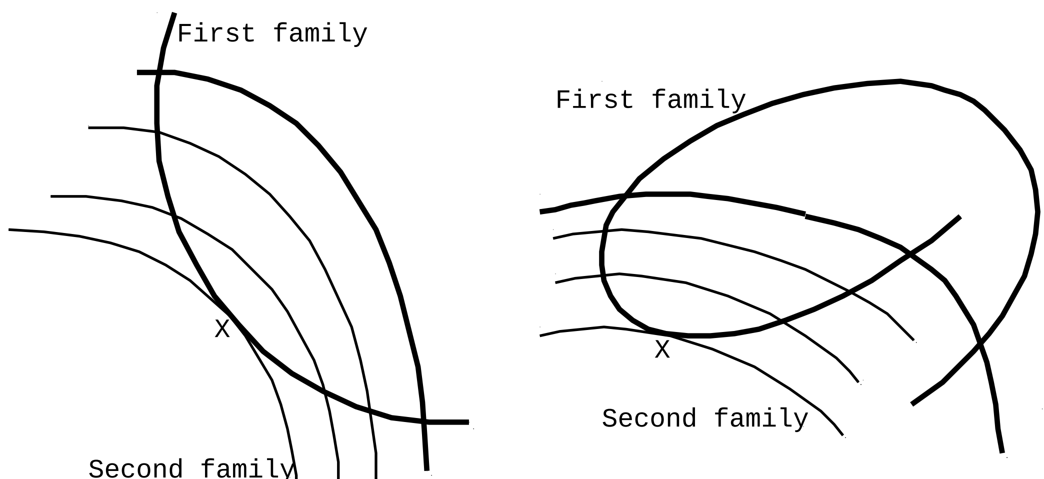
$$II(\tilde{V}_1, \tilde{V}_1) = II(\tilde{V}_2, \tilde{V}_2) = 0.$$

Note that $-V_i$ is also an asymptotic direction. By making sure that the $\pm$ is chosen consistently, we can arrange that the vector fields vary smoothly from point to point near P. In fact, we can find formulas for V_i in terms of h_{ij}. Since X was assumed to be an isometric immersion, it is a covering map and these vector fields continue to globally defined vector fields on all of $\mathbb{H}^2$. The fact that the curvature never vanishes implies that neither vector field has a singularity.

We now use the vector fields to find a new coordinate system for $\mathbb{H}^2$. Through each point of $\mathbb{H}^2$ we can find integral curves of the vector fields. Namely, for each $P \in \mathbb{H}^2$ there is a function $\sigma_i(s, P) : \mathbb{R} \times \mathbb{H}^2 \to \mathbb{H}^2$ so that

$$\begin{aligned} \frac{d\sigma_i}{ds}(s, P) =& V_i(\sigma_i(s, P)), \qquad && \text{for all } s \in \mathbb{R} \text{ and } P \in \mathbb{H}^2, \\ \sigma_i(0, P) =& P && \text{for all } P \in \mathbb{H}^2. \end{aligned}$$

The curve $s \mapsto \sigma_i(s, P)$ simply follows the vector field V_i. Call the curves which follow V_1 the first family of asymptotic curves and those that follow V_2 the second family.



Picture proof that curves from different families cross at most once.

These curves have some nice properties. First of all, as they are integral curves they can never cross themselves. For for a fixed P, the curve $s \mapsto \sigma_i(s, P)$ is unbounded in both directions. This follows from the Poincaré-Bendixon Theorem from ODE's, for, otherwise, if a trajectory stays in a bounded set, it would have to converge to a limiting periodic (cyclic) trajectory and the vector field would have to vanish somewhere inside the cycle. It is also true that two integral curves from different families may cross at most once. If this were not the case, then there would have to be some point (X in the figure) where some other pair of trajectories from different families would have to be tangent. However, this would imply that V_1 is parallel to V_2 there, which never happens.

Thus, if we choose some point $O \in \mathbb{H}^2$, then $s \mapsto \sigma_1(s, O)$ is a unit speed curve through O that gives one new coordinate axis. Then following a curve from the other family gives the other coordinate. Hence there is a map $\Psi : (s,t) \mapsto \sigma_2(t, \sigma_1(s, O))$ from $\Psi : \mathbb{R}^2 \to \mathbb{H}^2$. Because each pair of coordinate curves can cross at most once, this map Ψ is injective. It is a fact that it is also surjective, thus we have a new global coordinates for $\mathbb{H}^2$. Such coordinates are called *Chebychev* coordinates. They have some more amazing properties. One is that it didn't matter which family is followed first, $\sigma_2(t, \sigma_1(s, O)) = \sigma_1(s, \sigma_2(t, O))$ for all (s,t). This is because all rectangles bounded by pairs of asymptotic curves from both families have the property that opposite edges have the same length. This fact is used in proving the surjecivity of Ψ.

Let $\varphi(s,t) = \angle(V_1(s,t), V_2(s,t))$ denote the angle between the two asymptotic directions at the point (s,t). Because the vectors V_1 and V_2 are linearly independent, this angle can be chosen to satisfy

$$0 < \varphi(s,t) < \pi \qquad \text{for all } (s,t) \in \mathbb{H}^2. \tag{7}$$

φ is an extrinsic quantity, because it depends on how the surface sits in $\mathbb{R}^3$ and is computed in terms of h_{ij}. It turns out that the Riemannian Metric for the surface in Chebychev coordinates takes the form

$$ds^2 + 2\cos(\varphi)\, ds\, dt + dt^2. \tag{8}$$

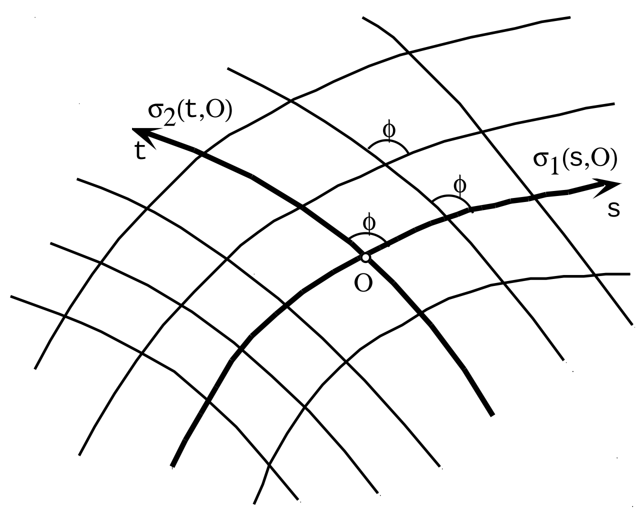
This metric is sometimes called the weavers metric, because the coordinate rectangles have opposite sides of equal length, but may distort by skewing, just as the threads in cloth.

The Gauß curvature may be computed from this metric, giving

$$\frac{\partial^2 \varphi}{\partial s \partial t} = -K \sin(\varphi). \tag{9}$$

This equation was discovered by Hazzidakis in 1880. It is a *Sine-Gordon* type equation and occurs also in the theory of nonlinear waves. The formulas (8) and (9) and the other facts require some differential geometric compuation. Finally, the area form may be computed from the fact that both coordinate curves s and t are unit speed but make an angle of φ, thus

$$d\,\text{Area} = \sin(\varphi)\, ds\, dt.$$



Chebychev Coodrinates.

Let us compute the area of the coordinate rectangle $\mathcal{R}(a,b) = \{(s,t) : |s| < a, |t| < b\}$ for the isometric immersion $X(\mathbb{H}^2)$ using the fact that $K \equiv -1$. Substituting (9), integrating by parts, and using (7) yields

$$\begin{aligned}
\text{(10)} \qquad \text{Area}(\mathcal{R}(a,b)) &= \iint_{\mathcal{R}} d\,\text{Area} = \int_{-a}^{a}\int_{-b}^{b} -K\,\sin(\varphi)\,dt\,ds \\
&= \int_{-a}^{a}\int_{-b}^{b} \frac{\partial^2\varphi}{\partial s\partial t}\,dt\,ds = \int_{-a}^{a}\left\{\frac{\partial\varphi}{\partial s}(s,b) - \frac{\partial\varphi}{\partial s}(s,-b)\right\} ds \\
&= \varphi(a,b) - \varphi(-a,b) - \varphi(a,-b) + \varphi(-a,-b) \le 2\pi.
\end{aligned}$$

Since we have a global coordinate system, any distance disk $B(O,\rho) \subset \mathcal{R}(a,b)$ provided that a, b are large enough. Inequality (10) implies

$$2\pi(\cosh\rho - 1) = \text{Area}(B(O,\rho)) \le 2\pi,$$

but this is a contradiction for any ρ large enough so $\cosh\rho > 2$. □

Further Developments.

Efimov made a great generalization of Hilbert's nonexistence theorem for complete surfaces of nonpositive curvature. [E1],[K].

Theorem. [N. Efimov (1961)] *There is no C^2 isometric immersion of a complete, two dimensional, Riemannian manifold $M \subset \mathbb{R}^3$ whose curvature satisfies $K \le -1$.*

Efimov's further claim to generalize the result to surfaces whose curvature decays to zero slowly (like inverse square of distance) is still controversial [E2].

The movie only hinted at John Nash's important positive results for isometric embedding of any Riemannian manifold. If one decreases the regularity, then by crinkling the surface sufficiently, one can isometrically embed the hyperbolic plane into $\mathbb{R}^3$.

Theorem. [N. Kuiper(1955) & J. Nash (1956)] *If any Riemannian manifold (M^n, g) admits a C^1 immersion into $\mathbb{R}^q$, with $q \ge n+1$, then it admits a C^1 isometric immersion into $\mathbb{R}^q$.*

Also, if you are willing to embed to a Euclidean space of high enough dimension, you can do it for *any* Riemannian manifold. For example, this theorem allows you to find C^3 isometric embeddings $\mathbb{H}^2$ into $\mathbb{R}^{99}$.

Theorem. [J. Nash (1956)] *For $k \ge 3$, any C^k manifold (M^n, g) can be C^k-isomtrically immersed into $\mathbf{R}^q$ where $q \ge \frac{3}{2}n(n+1)(n+9)$.*

Thus one can embed hyperbolic space into high dimensional Euclidean Space. Nash invented the "hard" implicit function theorem of PDE's to prove this result. It has since been proved by M. Günther(1989) using "off the shelf" elliptic methods. He holds the world record for the embedding dimension $q \ge \frac{1}{2}n(n+3) + 5$.

For the special case of hyperbolic space, there is a special explicit formula that works in a lower dimension.

Theorem. [D. Blanuša (1955)] *There is a explicit C^∞ proper isometric embedding of the hyperbolic plane $\mathbb{H}^2$ into $\mathbb{R}^6$.*

Remarks. These notes were inspired by my talk "The hyperbolic plane is too big for $\mathbf{R}^3$" given in the Undergraduate Colloquium at the University of Utah on March 25, 2003. They are offered as an instructional module, which may be useful for the course on Curves and Surfaces, geometric methods in education, beginning analysis, advanced calculus or for an introduction to proofs. do Carmo's undergraduate text [dC] on Curves and Surfaces presents a proof of Hilbert's Theorem. So do the books of Blaschke & Leichtweiß [BL], Stoker[S] and Willmore [W]. Struik dicsusses history [St].

Embedding questions for negatively curved surfaces are discussed by Poznyak [P] and Rozendorn [R]. Efimov's theorem [E] is detailed by T. Klotz [K]. For metrics closer to Euclidean, other geometric conditions must be imposed to prove impossibility of isometric immersion, it e.g., [CT]. An advanced reference for isometric immersions is Gromov [G].

References

[BL] W. Blaschke & K. Leichtweiß, *Elementare Differentialgeometrie 5th. ed.*, Springer, Berlin, 1973.

[CT] H. Chan & A. Treibergs, *Nonpositively curved surfaces in* $\mathbb{R}^3$, J. Differential Geometry **57** (2001), 389–407.

[dC] M. do Carmo, *Differential Geometry of Curves and Surfaces*, Prentice Hall, Englewood Cliffs, 1976.

[E1] N. V. Efimov, *Impossibility of a complete regular surface in Euclidean 3-Space whose Gaussian curvature has a negative upper bound*, Doklady Akad. Nauk. SSSR **150** (1963), 1206–1209 (Russian); Engl. transl. in Sov. Math. (Doklady) **4** (1963), 843–846.

[E2] ———, *Diffrential criteria for homeomorphism of certain mappings with application to the theory of surfaces*, Mat. Sb., Nov. Ser. **76** (1968) (Russian); Engl. transl. in Math. USSR (Sbornik) **5** (1968), 475–488.

[G] M. Gromov, *Partial Differential Relations*, Ergebnisse der Mathematik und ihrer Grenzgebiete; 3. Folge, Bd. 9, Springer, Berlin, 1986.

[H] D. Hilbert, *Über Flächen von constanter Gaußscher Krümmung*, Trans. Amer. Math. Soc. **2** (1901), 87–99.

[FK] F. Klein, *Elementary Mathematics from an Advanced Standpoint: Geometry*, Dover, New York, 1948, Originally published by Macmillan Co., New York, 1940; Engl. trans. of *Elementarmathematik vom hoheren Standpunkte, Band 2: Geometrie* 3rd ed., Springer, Berlin, 1925.

[K] T. Klotz Milnor, *Efimov's theorem about complete immersed surfaces of negative curvature*, Advances in Math. **8** (1972), 474–543.

[M] B. Meserve, *Fundamental concepts of geomertry*, Dover, Mineola, 1983., Originally published by Addison-Wesley, Reading, 1959.

[R] E. R. Rozendorn, *Surfaces of negative curvature*, Geometry III: Theory of surfaces (J. D. Burago & V. A. Zalgaller, eds.), v. 48, VINITI, Moscow, 1989 (Russian); Engl. transl., *Encyclopedia of Mathematical Sciences*, Springer, Berlin, 1992, pp. 87–178.

[S] J. Stoker, *Differenial Geometry*, Wiley-Interscience, New York, 1969.

[St] D. Struik, *A Concise History of Mathematics, 3rd. ed.*, Dover, New York, 1967.

[W] T. Willmore, *An Introduction to Differential geometry*, Oxford University press, london, 1961.

University of Utah,
Department of Mathematics
155 South 1400 East, Rm 233
Salt Lake City, UTAH 84112-0090
*E-mail address:*treiberg@math.utah.edu