NOTES ON HYPERBOLIC GEOMETRY

ROBERT L. BRYANT

Abstract. These are some notes to help you follow the lectures in class on the construction of a non-Euclidean model of the Hilbert plane, i.e., a model in which Hilbert's Euclidean Parallel Postulate does not hold. As it happens, our model will also satisfy Dedekind's Axiom, so, in particular, all of the various continuity axioms will be satisfied for our model as well. This particular model is known as the *hyperbolic plane*, and it has a number of different (but isomorphic) realizations.

Contents

Date: November 2024.
Key words and phrases. hyperbolic geometry .
This is Draft Version 2.0.

1. Introduction

In these notes, I will provide the details for the construction of the *hyperbolic plane*, a model for our incidence, betweenness, and congruence axioms for which the Parallel Postulate does not hold. It will also satisfy Dedekind's Axiom and, hence, all of the continuity axioms that we have discussed.

It turns out that it is useful to have more than one way to represent the hyperbolic plane, though all of them are equivalent. (We'll show that they are equivalent, i.e., isomorphic.)

2. The Klein Disk Model

The first model we'll discuss is the *Klein model* K. Although it would be possible to verify all of Hilbert axioms in this model directly (once we have fully defined the model), it turns out that this is somewhat complicated algebraically, so we'll rely on the other models to help us avoid this algebraic mess.[1]

2.1. **Points and Lines.** The *points* of K are the points $P = (x, y) \in \mathbb{R}^2$ that lie inside the unit circle, i.e.,

$$x^2 + y^2 < 1.$$

We'll call these 'K-points', when there is any danger of confusion. The points on the unit circle are sometimes called *ideal points*, but I emphasize that they are *not* points in the model.

A *line* of K (i.e., a 'K-line') is a chord of the unit circle, i.e., it is described by a *linear equation*

$$ax + by - c = 0,$$

where $a^2 + b^2 > c^2$ is necessary in order that this Euclidean line pass through some K-point. (Any K-line has a K-point on it, by Incidence Axiom 2, and the closest point on the line $ax+by-c = 0$ to the origin is $(x, y) = \big(ca/(a^2+b^2), cb/(a^2+b^2)\big)$.)

We define *incidence* and *betweenness* in this model to be the same as incidence and betweenness in the coordinate plane $\mathbb{R}^2$.

Remark 1 (Basic axioms of the Klein model). In earlier discussions, in class and homework, we have already seen that Hilbert's incidence and betweenness axioms hold for the points and lines of K.

We also know that Dedekind's Axiom holds (because Dedekind's Axiom holds for the coordinate plane and this easily implies Dedekind's Axiom for chords of the unit circle, i.e., lines in the model K).

We have also seen, through simple examples, that the Parallel Postulate does *not* hold.

[1] Still, the algebra isn't so complicated except for the final SAS Postulate (C-6), so it's instructive to see how this works.

Exercise 2.1. Show that two distinct K-lines

$$a_1x + b_1y - c_1 = 0 \qquad \text{and} \qquad a_2x + b_2y - c_2 = 0$$

intersect in a K-point (i.e., in a point inside the unit circle) if and only if

$$(c_1c_2 - a_1a_2 - b_1b_2)^2 < ({a_1}^2 + {b_1}^2 - {c_1}^2)({a_2}^2 + {b_2}^2 - {c_2}^2).$$

(Hint: Use Cramer's Rule to find the point of intersection (x, y) of the two lines and check that $x^2 + y^2 < 1$ is equivalent to the above inequality.)

Remark 2 (Normalized equations of K-lines). Since a K-line $ax{+}by{-}c = 0$ satisfies $a^2{+}b^2{-}c^2 > 0$, we can let λ satisfy $\lambda^2 = a^2{+}b^2{-}c^2$ and note that, if $(\bar{a}, \bar{b}, \bar{c}) = (\frac{a}{\lambda}, \frac{b}{\lambda}, \frac{c}{\lambda})$, then

$$\bar{a}x + \bar{b}y - \bar{c} = 0$$

defines the same line as $ax{+}by{-}c = 0$, but it satisfies $\bar{a}^2 + \bar{b}^2 - \bar{c}^2 = 1$. We say that an equation $ax{+}by{-}c = 0$ for a K-line ℓ is *normalized* if $a^2{+}b^2{-}c^2 = 1$. Any K-line ℓ is defined by exactly two normalized equations:

$$ax{+}by{-}c = 0 \qquad \text{and} \qquad (-a)x{+}(-b)y{-}(-c) = 0.$$

when $a^2 + b^2 - c^2 = 1$.

2.2. **Segment Length and Congruence.** The first thing to do in the model K is to define congruence of segments. Here is how this can be done:

Let A and B be distinct K-points. Since they are distinct, they lie on a unique K-line ℓ, which meets the unit circle in two ideal points P and Q with

$$Q * A * B \qquad\qquad \text{and} \qquad\qquad A * B * P.$$

(Note that, if we were to switch A and B, we would have to switch P and Q in order to maintain the required betweenness relations.)

Since all of the vectors A, B, P, and Q are distinct and lie on a line, all of the difference vectors

$$P{-}A,\ P{-}B,\ A{-}Q,\ B{-}Q$$

are positive multiples of the vector $P{-}Q$. Thus, their ratios are well-defined as real numbers. In fact, we clearly have

$$\frac{P{-}A}{P{-}B} > 1 \qquad \text{and} \qquad \frac{Q{-}B}{Q{-}A} > 1.$$

Definition 1 (K-length and segment congruence). The *K-length* of the K-segment AB is the (positive) real number[2]

$$\overline{AB} = \frac{1}{2}\log\left(\frac{P{-}A}{P{-}B}\cdot\frac{Q{-}B}{Q{-}A}\right) > 0. \tag{2.1}$$

Moreover, two K-segments AB and CD are *congruent*, i.e., $AB \cong CD$, if and only if $\overline{AB} = \overline{CD}$.

[2]Here and elsewhere, the base of the logarithm is assumed to be Euler's constant e, i.e., the *natural logarithm, denoted in some texts by* ln. *Since we will not have any need to use the so-called* common logarithm, i.e., $\log_{10}$, this should not cause confusion.

Remark 3. Note that if we switch A and B (and therefore P and Q), the formula shows that $\overline{AB} = \overline{BA}$, so the K-length really does depend only on the pair of points, not their order.

The choice of the coefficient $\frac{1}{2}$ in the formula for K-length simplifies many of the later formulae. We could have used any (positive) coefficient and obtained a workable definition of K-length (corresponding to the choice of unit length segment in Theorem 4.3 of Greenberg), but, again, it turns out that this particular choice simplifies many later formulae. We'll return to this point later.

Exercise 2.2. Let P and Q be any two distinct ideal points. Show that a K-point X lies on the line through Q and P if and only if there is a real number x such that

$$X = \frac{1}{1+e^{2x}}\,Q + \frac{e^{2x}}{1+e^{2x}}\,P.$$

Next, let $a < b$ be any two real numbers. Show that the points

$$A = \frac{1}{1+e^{2a}}\,Q + \frac{e^{2a}}{1+e^{2a}}\,P \qquad \text{and} \qquad B = \frac{1}{1+e^{2b}}\,Q + \frac{e^{2b}}{1+e^{2b}}\,P$$

are K-points (i.e., lie inside the ideal circle) and satisfy $Q * A * B$ and $A * B * P$. Show also that $A * X * B$ if and only if $a < x < b$.

Using the formula (2.1), show that

$$\overline{AB} = b - a.$$

(Hint: First, observe that $X = (1-t)Q + tP$ for some t satisfying $0 < t < 1$ and then solve for x. Next, observe that $X = (1-s)\,A + s\,B$ for some s and that this s satisfies $0 < s < 1$ if and only if $a < x < b$.)

Finally, use the above observations to show that, if $A * X * B$ then

$$\overline{AB} = \overline{AX} + \overline{XB}.$$

Exercise 2.3. Explain why the previous exercises imply that, with this definition of K-congruence of segments, the model K satisfies congruence axioms C-1, C-2, and C-3.

2.3. **Angle Congruence.** It turns out that defining congruence of angles is not so easy in the K-model. However, it can be done as follows:

First, note that

$$ax + by - c = 0$$

is the equation of a K-line ℓ, so $a^2 + b^2 - c^2 > 0$. Then the two half-planes bounded by ℓ are described by the inequalities

$$ax + by - c > 0 \qquad \text{or} \qquad ax + by - c < 0.$$

By replacing (a, b, c) by $(-a, -b, -c)$, we can switch the two half planes.

Now suppose that $A = (x_A, y_A)$, $B = (x_B, y_B)$, and $C = (x_C, y_C)$ are distinct K-points that are not collinear. Let ℓ be the line through A and B, with normalized equation

$$a_1x + b_1y - c_1 = 0$$

(i.e., ${a_1}^2 + {b_1}^2 - {c_1}^2 = 1$), and let m be the line through A and C with normalized equation

$$a_2x + b_2y - c_2 = 0$$

(i.e., ${a_2}^2 + {b_2}^2 - {c_2}^2 = 1$).

After replacing (a_i, b_i, c_i) with $(-a_i, -b_i, -c_i)$ if necessary, we can assume that

$$a_1x_C + b_1y_C - c_1 > 0 \qquad \text{and} \qquad a_2x_B + b_2y_B - c_2 > 0.$$

Thus, C is on the 'positive' side of ℓ and B is on the 'positive' side of m, according to the equations we use to define them. The point is that the two linear functions $a_1x + b_1y - c_1$ and $a_2x + b_2y - c_2$ are positive on the interior of the angle $\angle BAC$.

Because ℓ and m pass through the K-point A, we know from Exercise 2.1 that

$$|c_1c_2 - a_1a_2 - b_1b_2| < \sqrt{a_1{}^2 + b_1{}^2 - c_1{}^2}\sqrt{a_2{}^2 + b_2{}^2 - c_2{}^2} = 1.$$

Definition 2 (K-angle measure and angle congruence). The *radian K-measure* of $\angle BAC$ is

$$(\angle BAC)^r = \cos^{-1}\big(c_1c_2 - a_1a_2 - b_1b_2\big). \tag{2.2}$$

Also K-angles $\angle ABC$ and $\angle DEF$ are *congruent*, i.e., $\angle ABC \cong \angle DEF$, if and only if $(\angle ABC)^r = (\angle DEF)^r$.

Remark 4. The above inequality ensures that the quantity inside the parenthesis is strictly between $-1 = \cos\pi$ and $1 = \cos 0$, so this does define $(\angle BAC)^r$ as a number between 0 and π. (Recall that the cosine function is strictly decreasing on the interval $[0, \pi]$, so it does have a well-defined inverse function when restricted to this interval.)

Also, note the use of radian measure rather than the more 'traditional' degree measure. Of course, we could convert radians to degrees, but it turns out that formulae work out simpler when we use radians.

Exercise 2.4. Explain why, if $\angle BAC$ is an angle and $\overrightarrow{AB'} = \overrightarrow{AB}$ while $\overrightarrow{AC'} = \overrightarrow{AC}$, then

$$(\angle B'AC')^r = (\angle BAC)^r.$$

Thus, the K-radian measure depends only on the two rays that make up the angle, not on the particular points.

Also, if D is a K-point that satisfies $D * A * C$ (so that $\overrightarrow{AD}$ is the opposite ray to $\overrightarrow{AC}$), explain why

$$(\angle BAD)^r = \pi - (\angle BAC)^r.$$

(Hint: $\cos(\pi - x) = -\cos(x)$.) Hence, by definition, a K-angle is congruent to its supplement if and only if its radian K-measure is $\frac{\pi}{2}$.

Exercise 2.5. Let $A = (0, 0)$, $B = (r_1 \cos\theta_1, r_1 \sin\theta_1)$, and $C = (r_2 \cos\theta_2, r_2 \sin\theta_2)$ where $0 < \theta_2 - \theta_1 < \pi$ and $r_1, r_2 > 0$. Using the definitions, show that

$$(\angle ABC)^r = \theta_2 - \theta_1\,,$$

i.e., when $A = (0, 0)$, the radian K-measure of angles agrees with Euclidean angle measure (in radians). Show, however, that, when $A = (u, v)$, $B = (u+\epsilon, v)$, and $C = (u, v+\epsilon)$, are K-points with $\epsilon > 0$ some small number, we have

$$\cos\big((\angle ABC)^r\big) = \frac{uv}{\sqrt{(1-u^2)(1-v^2)}},$$

so angles that 'look right' in the Euclidean sense need not be right in the K-model. Use this formula to construct a Lambert quadrilateral in the model K and show that its fourth angle is acute.

Exercise 2.6. Show that, for any K-ray $\overrightarrow{AB}$, any x satisfying $0 < x < \pi$, and on any given side of the line ℓ passing through A and B, there is a unique ray $\overrightarrow{AC}$ satisfying

$$(\angle BAC)^r = x.$$

Explain why this shows that, with this definition of angle congruence, the model K satisfies Congruence Axioms C-4 and C-5.

(Hint for the first part: Because of the properties of the cosine function, the hypothesis $0 < x < \pi$ is equivalent to the hypothesis $-1 < \cos x < 1$, so you are being asked to show that there is a unique ray $\overrightarrow{AC}$ defined by a normalized equation $a_2\, x + b_2\, y - c_2 = 0$, where

$$a_2\, x_A + b_2\, y_A - c_2 = 0 \qquad \text{and} \qquad c_1c_2 - a_1a_2 - b_1b_2 = \cos x,$$

where the line through A and B satisfies a normalized equation $a_1\, x + b_1\, y - c_1 = 0$ with $a_1\, x_C + b_1\, y_C - c_1 > 0$ and $a_2\, x_B + b_2\, y_B - c_2 > 0$. This is three linear equations for (a_2, b_2, c_2) (don't forget that the equation is normalized, so ${a_2}^2 + {b_2}^2 - {c_2}^2 = 1$) plus some inequalities. How does that help?)

2.4. **Side-Angle-Side.** Assuming all the exercises up to this point, we have now verified that the model K satisfies all of the axioms for a Hilbert plane *except* C-6, i.e., the infamous SAS Postulate. It is now possible (though the algebra is tedious), to *prove* by algebraic calculations that C-6 is satisfied as well. Right now, though, I'll just state this result as a theorem, and leave the proof to the next section, after we have developed further insight using a different (but isomorphic) model.

Theorem 1. *The model K, with incidence, betweenness, and congruence of segments and angles as defined above, satisfies the axioms of a Hilbert plane (in particular, C-6) and Dedekind's Axiom, but it does not satisfy the Parallel Postulate.*

Remark 5 (The hyperbolic laws of cosines and sines). In order not to leave the reader mystified, I'll just indicate how one could prove C-6 directly by algebra.

Consider a K-triangle $\triangle ABC$. To simplify notation, let $\overline{BC} = a$, $\overline{AC} = b$, and $\overline{AB} = c$ and let $\alpha = (\angle BAC)^r$, $\beta = (\angle CBA)^r$, $\gamma = (\angle ACB)^r$. Then, using the above definitions, one can prove by algebra that the following identities hold

$$\cosh c = \cosh a\, \cosh b - \sinh a\, \sinh b\, \cos\gamma\,, \tag{2.3}$$

$$\frac{\sin\alpha}{\sinh a} = \frac{\sin\beta}{\sinh b} = \frac{\sin\gamma}{\sinh c}\,, \tag{2.4}$$

and

$$\cos\gamma = -\cos\alpha\, \cos\beta + \sin\alpha\, \sin\beta\, \cosh c\,. \tag{2.5}$$

The identity (2.3) is known as the *first hyperbolic law of cosines*, the identity (2.4) is known as the *hyperbolic law of sines*, and the identity (2.5) is known as the *second hyperbolic law of cosines*.

The first hyperbolic law of cosines shows that if we know a and b and γ (i.e., the length of two sides and the measure of the included angle), we can compute c (the remaining side length) (since the function cosh is strictly increasing on the interval $(0, \infty)$. Now, knowing a, b, and c, we can solve for α and β by using the permuted identities (since $\sinh x > 0$ whenever $x > 0$)

$$\cosh a = \cosh b\, \cosh c - \sinh b\, \sinh c\, \cos\alpha$$

and

$$\cosh b = \cosh a\, \cosh c - \sinh a\, \sinh c\, \cos\beta.$$

Thus, knowing the length of two sides of a triangle and the measure of the included angle determines the length of the other side and the measures of the other angles. But this is exactly the content of C-6.

Unfortunately, the proof of (2.3) just using the formulae for a, b, c, and γ that come straight from the definitions is very complicated. In the model that we will develop in the next section, though, we can bring to bear some insight from 3-dimensional linear algebra, and this simplifies the proof of (2.3) dramatically.

3. The Hyperbolic Sheet Model

This model, the *hyperbolic sheet model* S turns out to be good because it makes the closest connection to linear algebra, and linear algebra is easier than complicated polynomial algebra.

You have probably already noticed that expressions of the form $a^2 + b^2 - c^2$ kept coming up in our discussion of K-geometry, but it wasn't obvious why this was happening. Our model in 3-space is going to make this clear, because we are going to use the *Minkowskian* (aka *Lorentzian*) *inner product* on $\mathbb{R}^3$ instead of the standard (Euclidean) one.

3.1. **Minkowskian linear algebra in dimension 3.** We need to modify the familiar definitions of the (Euclidean) inner product (also known as 'dot product') and cross product. In this subsection, we'll collect the facts about this that we need.

Definition 3. The *Minkowsian inner product* of two vectors (a, b, c) and (u, v, w) in $\mathbb{R}^3$ is defined to be the real number

$$(a, b, c) \cdot (u, v, w) = au + bv - cw. \tag{3.1}$$

Definition 4. The Minkowskian *cross product* of two vectors, is defined to be the expression

$$(a, b, c) \times (u, v, w) = (bw - cv,\ cu - aw,\ bu - av). \tag{3.2}$$

Remark 6. Note the minus sign in the third term in the formula for the inner of two vectors in $\mathbb{R}^3$. Also, if you look carefully at the formula for the cross product, you'll note that the sign of the third term there is also reversed from the Euclidean case. These are essential for everything that we will do in this model. Despite these minus signs, the inner product will have most of the properties of the more familiar Euclidean inner (aka 'dot') product.

Remark 7 (Linearity). For example, we have the familiar linearity in each variable separately, and symmetry for the dot product and anti-symmetry for the cross product:

$$\begin{aligned} X \cdot Y &= Y \cdot X \\ X \cdot (a\,Y + b\,Z) &= a\,(X \cdot Y) + b\,(X \cdot Z) \\ X \times Y &= -Y \times X \\ X \times (a\,Y + b\,Z) &= a\,(X \times Y) + b\,(X \times Z) \end{aligned} \tag{3.3}$$

Remark 8 (Relations between the two products). We also have that $X \times Y = 0$ only when X and Y are linearly dependent and that

$$X \cdot (X \times Y) = Y \cdot (X \times Y) = 0 \tag{3.4}$$

for all vectors X and Y in $\mathbb{R}^3$. Thus, when X and Y are linearly independent, the vector $X \times Y$ is orthogonal (in the Minkowski sense) to the plane spanned by X and Y.

There is also the important formula

$$(X \times Y) \cdot (X \times Y) = (X \cdot Y)^2 - (X \cdot X)(Y \cdot Y), \tag{3.5}$$

in which, you should note, the right hand side is the *negative* of the right hand side you find in the Euclidean inner product formula.

Finally, there is the *triple product determinant formula*

$$X \cdot (Y \times Z) = \det \begin{pmatrix} X \\ Y \\ Z \end{pmatrix}. \tag{3.6}$$

(Remember that we are writing our vectors as row vectors.) As a consequence, we still have the 'cyclic permutation identity' for the triple product:

$$X \cdot (Y \times Z) = Y \cdot (Z \times X) = Z \cdot (X \times Y). \tag{3.7}$$

Exercise 3.1. Verify the identity (3.5), then show that the following generalization holds:

$$(X \times Y) \cdot (X \times Z) = (X \cdot Y)(X \cdot Z) - (X \cdot X)(Y \cdot Z). \tag{3.8}$$

(The original case is when $Z = Y$.) (Hint: Once you have verified that (3.5) holds for all X and Y in $\mathbb{R}^3$, substitute $Y + Z$ for Y in that formula and expand both sides, using the fact that you already know that the formula holds for Y and for $Y = Z$.)

3.2. **Points and Lines.** With that bit of linear algebra out of the way, the first order of business is to define the points and lines of the model and the incidence relation.

3.2.1. *Points.* The *points* of S will be the vectors $A = (u, v, w) \in \mathbb{R}^3$ that lie on the *upper sheet* S of the hyperboloid of 2-sheets defined by $u^2 + v^2 - w^2 = -1$, i.e., the ones that satisfy $w > 0$. (In fact, because $w^2 = 1 + u^2 + v^2 \geq 1$, the hypothesis $w > 0$ implies that $w \geq 1$.)

Note that an S-point $A \in \mathbb{R}^3$ satisfies $A \cdot A = -1$, but this is not quite enough, because we also have to check that A lies on the upper sheet, i.e., that its third coordinate is positive.

3.2.2. *Correspondences between K and S.* If $(u, v, w) \in S$ lies in ℓ, then $w \neq 0$, so there is a well-defined point

$$(x, y) = \left(\frac{u}{w}, \frac{v}{w}\right) = f(u, v, w)$$

that satisfies

$$x^2 + y^2 = (u/w)^2 + (v/w)^2 = 1 + (u^2 + v^2 - w^2)/w^2 = 1 - 1/w^2 < 1,$$

so (x, y) is a K-point. Conversely, if (x, y) is a K-point, then

$$(u, v, w) = \left(\frac{x}{\sqrt{1 - x^2 - y^2}}, \frac{y}{\sqrt{1 - x^2 - y^2}}, \frac{1}{\sqrt{1 - x^2 - y^2}} \right) = g(x, y)$$

is a point on S, and we have

$$g\big(f(u, v, w)\big) = (u, v, w) \qquad \text{and} \qquad f\big(g(x, y)\big) = (x, y),$$

so this establishes a one-to-one correspondence between K-points and S-points.

3.2.3. *Lines.* An S-*line* ℓ is a plane in $\mathbb{R}^3$ that passes through the origin $(0, 0, 0)$ and that has nonempty intersection with S.

By definition, the point A is *incident* with ℓ if and only if A lies in the plane ℓ.

A plane ℓ through the origin has an equation of the form

$$a\,u + b\,v - c\,w = 0$$

for some constants (a, b, c) (unique up to replacement by $(\lambda a, \lambda b, \lambda c)$ for any nonzero real number λ).

A plane ℓ with the above equation meets S in a point (u, v, w) if and only if

$$a\,\frac{u}{w} + b\,\frac{v}{w} - c = a\,x + b\,y - c = 0.$$

From what we already know about the model K, we know that this can happen for some $(u, v, w) \in S$ if and only $a^2 + b^2 - c^2 > 0$.

Exercise 3.2. Show that the mappings $g : K \to S$ and $f : S \to K$ (which are inverse to each other), make the lines in the two models correspond as well. In other words, show that the two mappings f and g preserve collinearity, i.e., carry collinear triples to collinear triples.

Thus, we can always assume, if necessary, that an equation $a\,u + b\,v - c\,w = 0$ for an S-line ℓ is *normalized* so that $a^2 + b^2 - c^2 = 1$, and this determines (a, b, c) up to replacing it by $(-a, -b, -c)$. In other words, the S-lines correspond 2-to-1 to the points of the *hyperboloid of one sheet*

$$\Lambda = \big\{ (a, b, c) \in \mathbb{R}^3 \,\big|\, a^2 + b^2 - c^2 = 1 \ \big\} \subset \mathbb{R}^3.$$

Remark 9 (Choosing sides). This 'double counting' is actually an advantage, because we have two choices for the normalized equation of a line ℓ, and those two choices correspond to deciding which side of the line ℓ is going to be regarded as the positive side. What I mean by this is that, when we write the equation of ℓ as $(a, b, c) \cdot (u, v, w) = 0$ where $P = (a, b, c) \in \Lambda$ satisfies $P \cdot P = 1$, then the two sides of the line in S are the two sets

$$H_+ = \{ A \in S \mid P \cdot A > 0 \ \} \quad \text{and} \quad H_- = \{ A \in S \mid P \cdot A < 0 \ \}.$$

Replacing P by $-P$ switches these two sides. Thus, a choice of a normalized $P \in \Lambda$ to define a line ℓ corresponds to choosing not only the line ℓ, but also which of the two sides we want to call 'positive'.

3.3. **Segment congruence.** We now want to see how to compute length of segments in the S-model. The following argument will show that, for K-points A and B, the all-important formula

$$g(A)\cdot g(B) = -\cosh\bigl(\,\overline{AB}\,\bigr) = -\frac{e^{\overline{AB}}+e^{-\overline{AB}}}{2} \tag{3.9}$$

holds, so that the formula for distance in the S-model is expressed simply in terms of the inner product (and the hyperbolic cosine function).

Let $P = (\cos\alpha, \sin\alpha)$ and $Q = (\cos\beta, \sin\beta)$ be two distinct ideal points of the K-model, and consider corresponding vectors $\hat{P} = (r\cos\alpha, r\sin\alpha, r)$ $\hat{Q} = (r\cos\beta, r\sin\beta, r)$ in $\mathbb{R}^3$, where r is a positive real number Then $\hat{P}$ and $\hat{Q}$ are linearly independent, and they satisfy

$$\hat{P}\cdot\hat{P} = \hat{Q}\cdot\hat{Q} = 0$$

and

$$\hat{P}\cdot\hat{Q} = -r^2\bigl(1-\cos(\alpha{-}\beta)\bigr) < 0$$

(after all, $\cos(\alpha-\beta) < 1$ because P and Q are distinct).

Choose $r > 0$ so that $r^2\bigl(1-\cos(\alpha{-}\beta)\bigr) = \frac{1}{2}$. Then $\hat{P}\cdot\hat{Q} = -\frac{1}{2}$.

Now, using the formulae for these inner products, we have that, for each $a\in\mathbb{R}$, the point

$$\hat{A} = e^a\hat{P} + e^{-a}\hat{Q}$$

lies in S and the plane ℓ spanned by $\hat{P}$ and $\hat{Q}$. Moreover, $\hat{A} = g(A)$ where A is the K-point

$$A = \left(\frac{1}{1+e^{2a}}\right)Q + \left(\frac{e^{2a}}{1+e^{2a}}\right)P$$

which satisfies $Q*A*P$. Conversely, every K-point A that lies between the ideal points P and Q is $A = f(\hat{A})$ where $\hat{A}$ is of the above form.

Similarly, if $b\neq a$ is another real number, then, letting

$$\hat{B} = e^b\hat{P} + e^{-b}\hat{Q}$$

we get $\hat{B} = g(B)$ where $Q*B*P$ and that

$$B = \left(\frac{1}{1+e^{2b}}\right)Q + \left(\frac{e^{2b}}{1+e^{2b}}\right)P$$

Now, the above formulae for the inner products of $\hat{P}$ and $\hat{Q}$ show that $\hat{A}\cdot\hat{B} = -\cosh(a-b) = -\cosh(b-a)$.

Meanwhile, using the formula we have already verified for K-points, we have

$$\overline{AB} = |a-b|.$$

Thus, under the mapping $g : K\to S$, we have the desired formula

$$g(A)\cdot g(B) = -\cosh\bigl(\,\overline{AB}\,\bigr) = -\frac{e^{\overline{AB}}+e^{-\overline{AB}}}{2}.$$

Remark 10. The above exercise shows that the peculiar notion of K-length that we defined in the K-model actually has a simpler, more natural formula in the S-model in terms of the inner product. This motivates the following definition:

Definition 5. The *S-length* of an S-segment AB is

$$\overline{AB} = \cosh^{-1}\bigl(-A \cdot B\bigr) > 0.$$

Two S-segements AB and CD are *congruent* if they have the same S-length. Equivalently, $AB \cong CD$ if and only if $A \cdot B = C \cdot D$.

Remark 11. Any linear transformation $L : \mathbb{R}^3 \to \mathbb{R}^3$ that preserves the Minkowskian inner product must preserve the hyperboloid of two sheets and must therefore either preserve the upper sheet or exchange it with the lower sheet.

The key consequence of the above definition is that if such an L does preserve S (equivalently, preserves the upper cone), then it carries the points and lines of S to the points and lines of S and preserves the S-length of all segments.

3.4. **Angle Congruence.** Now, we want to concentrate on defining the S-angle measure for a general angle $\angle BAC$, but, ultimately, I am going to be working with the triangle $\triangle ABC$, where A, B, and C are non-collinear S-points.

Thus, I'm going to introduce notation so that $a > 0$ will be the length of the side opposite $\angle A$, i.e., BC; $b > 0$ will be the length of the side opposite $\angle B$, i.e., AC; and $c > 0$ will be the length of the side opposite $\angle C$, i.e., AB. Thus, by the results of the previous section on length, we have

$$A \cdot A = B \cdot B = C \cdot C = -1$$

and

$$B \cdot C = -\cosh a, \qquad C \cdot A = -\cosh b, \qquad A \cdot B = -\cosh c\,.$$

It's important to note that, because A, B, and C are assumed to form a triangle, i.e., that they are *not* linearly dependent, the triple product $A \cdot (B \times C)$ is not zero. By switching B and C if necessary, we can assume, without loss of generality, that

$$A \cdot (B \times C) > 0.$$

Now, A and B are distinct S-points, which are, therefore, linearly independent in $\mathbb{R}^3$. Using the above dot products, we see that

$$V = \frac{B - \cosh c\, A}{\sinh c}$$

is well-defined (since $\sinh c > 0$), and we can compute from the above inner product formulae that

$$A \cdot V = 0 \qquad \text{and} \qquad V \cdot V = 1.$$

Now, consider the vector $M = A \times V = (A \times B)/(\sinh c)$ (since $A \times A = 0$). Then M must satisfy $M \cdot A = M \cdot B = 0$, and we have

$$M \cdot M = (A \times V) \cdot (A \times V) = -(A \cdot A)(V \cdot V) = 1,$$

so $M \cdot X = 0$ is a normalized equation for the line ℓ passing through A and B. Moreover,

$$M \cdot C = C \cdot M = \frac{C \cdot (A \times B)}{\sinh c} = \frac{A \cdot (B \times C)}{\sinh c} > 0,$$

so the equation $M \cdot X = 0$ has C on the positive side of the line through A and B.

Next, we consider the line m through A and C. Define the vector

$$W = \frac{C - \cosh b\, A}{\sinh b}$$

which, using our formulae for dot products of A, B, and C, satisfies

$$A \cdot W = 0 \qquad \text{and} \qquad W \cdot W = 1.$$

Now, consider the vector $N = -A \times W = -(A \times C)/(\sinh b)$ (since $A \times A = 0$). Then N must satisfy $N \cdot A = N \cdot C = 0$ and we get

$$N \cdot N = (A \times W) \cdot (A \times W) = -(A \cdot A)(W \cdot W) = 1,$$

so $N \cdot X = 0$ is a normalized equation for the line m passing through A and C. Moreover,

$$N \cdot B = B \cdot N = -\frac{B \cdot (A \times C)}{\sinh b} = \frac{A \cdot (B \times C)}{\sinh b} > 0,$$

so the equation $N \cdot X = 0$ has B on the positive side of the line through A and C.

Because the plane m is spanned by A and W and is different from the plane ℓ spanned by A and V, it follows that $M = (a_2, b_2, c_2)$ and $N = (a_1, b_1, c_1)$ are linearly independent. Since A lies on both m and ℓ, we have

$$-1 < c_1c_2 - a_1a_2 - b_1b_2 < 1.$$

Definition 6. The S-*measure* (in radians) of $\angle BAC$ is

$$(\angle BAC)^r = \cos^{-1}\bigl(c_1c_2 - a_1a_2 - b_1b_2\bigr) = \cos^{-1}(-M \cdot N).$$

Two S-angles are *congruent* if and only if they have the same S-measure.

Exercise 3.3. Trace through the algebra to verify that for any three non-collinear K-points A, B, C, we have

$$(\angle BAC)^r = \bigl(\angle g(B)g(A)g(C)\bigr)^r.$$

This shows that the above definition of S-measure in radians corresponds under the isomorphism $g : K \to S$ of the K-measure of angles in radians defined in the K-model.

Finally, we can now use our formulae to compute $M \cdot N = -\cos\bigl((\angle BAC)^r\bigr)$.

$$\begin{aligned} M \cdot N &= -(A \times V) \cdot (A \times W) = (A \cdot A)(V \cdot W) \\ &= -\left(\frac{B - \cosh c\, A}{\sinh c}\right) \cdot \left(\frac{C - \cosh b\, A}{\sinh b}\right) \\ &= \frac{\cosh a - \cosh b\, \cosh c}{\sinh b\, \sinh c}. \end{aligned} \tag{3.10}$$

This rearranges to give the famous *First Hyperbolic Law of Cosines*:

$$\cosh a = \cosh b\, \cosh c - \sinh b\, \sinh c\, \cos\alpha, \tag{3.11}$$

where $\alpha = (\angle BAC)^r$.

Exercise 3.4. Using the fact that, when x is small, we have $\sinh x \approx x$ and $\cosh x \approx 1 + \frac{1}{2}x^2$, explain why, when a, b, and c are all very small, we have

$$\cos\bigl((\angle BAC)^r\bigr) \approx \frac{b^2 + c^2 - a^2}{2bc},$$

so, for very small triangles, the *Euclidean* Law of Cosines is *approximately* true.

Exercise 3.5. For an equilateral hyperbolic triangle, with $b = c = a > 0$, show that

$$\cos\big((\angle BAC)^r\big) = \frac{\cosh^2 a - \cosh a}{\sinh^2 a} = \frac{\cosh a}{1 + \cosh a} > \frac{1}{2}\,,$$

so $(\angle BAC)^r < \pi/3$. Thus, the angle defect of a hyperbolic equilateral triangle is always positive! (In fact, as the side length grows without bound, the angle measure at each vertex decreases to zero, so, for any α satisfying $0 < \alpha < \pi/3$, there is an equilateral hyperbolic triangle whose angles all have measure α.)

Exercise 3.6 (The hyperbolic Law of Sines). Using the notation $\alpha = (\angle BAC)^r$, $\beta = (\angle CBA)^r$, $\gamma = (\angle ACB)^r$, prove that any triangle $\triangle ABC$ satisfies

$$\frac{\sin\alpha}{\sinh a} = \frac{\sin\beta}{\sinh b} = \frac{\sin\gamma}{\sinh c}\,. \tag{3.12}$$

(Hint: The three ratios are positive numbers, so it suffices to show that their squares are equal. Let

$$R = \left(\frac{\sin\alpha}{\sinh a}\right)^2 = \frac{(1 - \cos^2\alpha)}{(\cosh^2 a - 1)}\,.$$

Note that, by the first hyperbolic law of cosines, we have

$$\cos^2\alpha = \frac{(\cosh a - \cosh b\,\cosh c)^2}{(\cosh^2 b - 1)\,(\cosh^2 c - 1)}.$$

Using this, express R in terms of $\cosh a$, $\cosh b$, and $\cosh c$, being sure to simplify as much as possible. How does this help?).

Theorem 2 (C-6 holds)**.** *If, for two S-triangles $\triangle ABC$ and $\triangle DEF$, we have the congruences*

$$AB \cong DE, \qquad AC \cong DF, \qquad \textit{and} \qquad \angle BAC \cong \angle EDF,$$

then we also have

$$BC \cong EF, \qquad \angle ABC \cong \angle DEF \qquad \textit{and} \qquad \angle BCA \cong \angle EFD.$$

In other words, the two triangles are congruent.

Proof. By the hyperbolic Law of Cosines, we have (using the notation above for the triangle $\triangle ABC$)

$$\cosh a = \cosh b\,\cosh c - \sinh b\,\sinh c\,\cos\big((\angle BAC)^r\big),$$

so knowing $b = \overline{AC}$, $c = \overline{AB}$ and $(\angle BAC)^r$ determines $a = \overline{BC}$. The above hypotheses then imply that, since we know $\overline{DE} = c$, $\overline{DF} = b$ and $(\angle EDF)^r = (\angle BAC)^r$, we then also know $\overline{EF} = a$. Now we know that all three corresponding sides of the two triangles are congruent, and the hyperbolic Law of Cosines then gives us formulae for the S-radian measure of the other two corresponding angles. □

Exercise 3.7 (The second hyperbolic Law of Cosines). Using the above established notation, give a proof of the relation (2.5). Note that this relation (and its two corresponding relations that come from permuting the vertices) implies the hyperbolic AAA theorem, namely, two triangles are congruent if their corresponding angles are congruent. This theorem has no analog in Euclidean geometry because of the existence of similar triangles that are not congruent.

3.5. **Some Motions of the S-plane.** In this part, I will construct some explicit linear maps of $\mathbb{R}^3$ to itself that preserve the inner product. First, though, it's useful to observe that

$$X \cdot Y = \tfrac{1}{4}\big((X+Y) \cdot (X+Y) - (X-Y) \cdot (X-Y) \big),$$

which shows that a linear map $L : \mathbb{R}^3 \to \mathbb{R}^3$ preserves the inner product as long as it preserves the (simpler) quadratic form $Q(Z) = Z \cdot Z = u^2 + v^2 - w^2$ (when $Z = (u, v, w)$).

Here are three crucial examples:

- Define $F : \mathbb{R}^3 \to \mathbb{R}^3$ by

$$F(u, v, w) = (u, -v, w).$$

Since $u^2 + (-v)^2 - w^2 = u^2 + v^2 - w^2$, the linear map F preserves the quadratic form and hence the inner product. Note that F carries S to itself (since it doesn't change the sign of w) and preserves all of the incidence, betweenness, and congruence relations.

- For any angle θ, define $R_\theta : \mathbb{R}^3 \to \mathbb{R}^3$ by

$$R_\theta(u, v, w) = \big(\cos\theta\, u + \sin\theta\, v,\ -\sin\theta\, u + \cos\theta\, v,\ w\big).$$

Since $\cos^2\theta + \sin^2\theta = 1$, we have

$$(\cos\theta\, u + \sin\theta\, v)^2 + (-\sin\theta\, u + \cos\theta\, v)^2 - w^2 = u^2 + v^2 - w^2,$$

so R_θ (which is just ordinary rotation by an angle of θ around the w-axis) does preserve the inner product and doesn't change the sign of w. Thus, it carries S to itself and preserves incidence, betweenness, and congruence, so it is a motion. You can also verify that

$$R_\theta \circ R_\psi = R_{\theta+\psi}\,.$$

- For any real number t, define $B_t : \mathbb{R}^3 \to \mathbb{R}^3$ by

$$B_t(u, v, w) = \big(u,\ \cosh t\, v + \sinh t\, w,\ \sinh t\, v + \cosh t\, w\big).$$

Since $\cosh^2 t - \sinh^2 t = 1$, we have

$$u^2 + (\cosh t\, v + \sinh t\, w)^2 - (\sinh t\, v + \cosh t\, w)^2 = u^2 + v^2 - w^2,$$

so B_t (which is what in physics is called the Lorentz *boost* of magnitude t) does preserve the inner product. It slightly less obvious, but it's also true that B_t doesn't change the sign of the last entry w, but it doesn't (why?). You can also verify that

$$B_t \circ B_s = B_{t+s}\,.$$

Now the way is clear for the following crucial fact:

Lemma 1. *For any pair of S-points A and B with $\overline{AB} = x > 0$, there is a motion that is a rotation R_θ followed by a boost B_r (with $r \geq 0$) followed by another rotation R_ψ that carries A to $(0, 0, 1)$ and B to $(\sinh x, 0, \cosh x)$ with $x > 0$.*

Proof: Suppose $A = (u, v, w)$ then we can write $A = (r\sin\theta, -r\cos\theta, w)$ for some $r \geq 0$ and some θ, and we see that

$$R_\theta(A) = (0, -r, w).$$

Since $r^2 - w^2 = -1$ with $r \geq 0$ and $w > 0$, it follows that there is a $t \geq 0$ so that $w = \cosh t$ and $r = \sinh t$. Then we have

$$B_t\big(R_\theta(A)\big) = B_t(0, -r, w) = B_t(0, -\sinh t, \cosh t) = (0, 0, 1).$$

Thus, the motion $B_t \circ R_\theta$ takes A to $(0, 0, 1)$.

Now let $B' = B_t\big(R_\theta(B)\big) = (\bar{u}, \bar{v}, \bar{w})$. We know that $(\bar{u}, \bar{v}) \neq (0, 0)$ since B' cannot be $B_t\big(R_\theta(A)\big) = (0, 0, 1)$ because A and B were distinct. Thus

$$(\bar{u}, \bar{v}) = (s\cos\psi, -s\sin\psi)$$

for some $s > 0$ and some angle ψ. Then $s = \sinh x$ for some $x > 0$ and we have $-1 = \bar{u}^2 + \bar{v}^2 - \bar{w}^2 = \sinh^2 x - \bar{w}^2$, so $w = \cosh x$. Thus,

$$R_\psi(B') = R_\psi(\sinh x\cos\psi, -\sinh x\sin\psi, \cosh x) = (\sinh x, 0, \cosh x),$$

while $R_\psi(0, 0, 1) = (0, 0, 1)$.

Thus, the motion $R_\psi \circ B_t \circ R_\theta$ of S has the desired property: It takes A to $(0, 0, 1)$ and B to $(\sinh x, 0, \cosh x)$. $\square$

This applies immediately to give the following proposition about triangles:

Proposition 1. *Let A, B, and C be any three non-collinear S-points. Suppose that $\overline{AB} = x > 0$, $\overline{AC} = y > 0$, and $(\angle BAC)^r = \phi \in (0, \pi)$. Then there exists a motion of S that carries A to $A' = (0, 0, 1)$, B to $B' = (\sinh x, 0, \cosh x)$, and C to $C' = (\sinh y\,\cos\phi,\ \sinh y\,\sin\phi,\ \cosh y)$.*

Proof: We already know that there is a motion M that carries A to $A' = (0, 0, 1)$ and B to $B' = (\sinh x, 0, \cosh x)$. Let C' be the point to which C is carried by this motion M. Then $C' = (u, v, w)$ with $w > 0$ and $u^2 + v^2 - w^2 = -1$.

Since $-w = A' \cdot C' = A \cdot C = -\cosh y$, it follows that $w = \cosh y$. Since $u^2 + v^2 = w^2 - 1 = \cosh^2 y - 1 = \sinh^2 y$, it follows that there must be an angle ψ so that

$$C' = (\sinh y\,\cos\psi,\ \sinh y\,\sin\psi,\ \cosh y).$$

Now, $\sin\psi$ cannot be zero, because, if it were then C' would lie in the linear span of A' and B', which would make A', B', and C' collinear. But they cannot be collinear because A, B, and C are not collinear. Thus, $v = \sinh y \sin\psi$ is not zero. Hence, by applying the motion $F(u, v, w) = (u, -v, w)$ if necessary, we can get a new motion that still leaves A' and B' alone, but ensures that $C' = (u, v, w)$ has $v > 0$. Thus, we have a motion such that

$$A' = (0, 0, 1), \quad B' = (\sinh x, 0, \cosh x), \quad C' = (\sinh y\,\cos\psi,\ \sinh y\,\sin\psi,\ \cosh y)$$

with $\sin\psi > 0$, i.e., with $0 < \psi < \pi$. However, if we now compute the measure of the angle $\angle B'A'C'$ using the easily obtained results that $P' = (1, 0, 0)$ and $Q' = (\cos\psi, \sin\psi, 0)$, we get

$$\cos\psi = P' \cdot Q' = P \cdot Q = \cos\big((\angle BAC)^r\big) = \cos\phi,$$

so $\psi = \phi$, and we are done. $\square$

Finally, we can give another proof of SAS:

Theorem 3. *Let $\triangle ABC$ and $\triangle DEF$ be S-triangles that satisfy $AB \cong DE$, $AC \cong DF$ and $\angle BAC \cong \angle EDF$. Then there is a motion of the S-plane that carries A to D, B to E, and C to F. In particular, all the six corresponding parts of the triangles are congruent, so the triangles are congruent.*

Proof: By definition, the hypotheses say that $\overline{AB} = \overline{DE}$, $\overline{AC} = \overline{DF}$ and $(\angle BAC)^r = (\angle EDF)^r$. By the above Proposition, this implies that there is a motion of the S-plane that carries A to D, B to E, and C to F. But this determines all of the inner products in the set $\{D, E, F\}$ to be the same as the inner products in the set $\{A, B, C\}$, and all of the angle measures and segment measures are expressed in terms of the inner products, so all of the corresponding parts are congruent. $\square$

3.6. **Sub-models without Dedekind's Axiom.** One of the main theorems in our book (though it's not proved there) is this:

Theorem 4. *Any Hilbert plane that satisfies Dedekind's Axiom is isomorphic to either the Euclidean plane or the hyperbolic plane.*

Thus, we have found all of the models of Hilbert's axioms that satisfy Dedekind's Axiom. However, just as it's useful to know Euclidean models that don't necessarily satisfy Dedekind's Axiom because they can be used to show that certain constructions can't be carried out by ruler and compass, it's also useful to have models of Hilbert's Axioms that don't satisfy Dedekind's Axiom or the Parallel Postulate in order to show that certain constructions cannot be made in neutral geometry.

3.6.1. *Some sub-fields of the real numbers.* A subset $\mathbb{F}$ of the real numbers $\mathbb{R}$ is said to be a *sub-field* of $\mathbb{R}$ if it contains 0 and 1 and is closed under addition, subtraction, multiplication, and division (by nonzero numbers, of course).

The smallest sub-field of $\mathbb{R}$ is $\mathbb{Q}$, the field of rational numbers, i.e., the real numbers of the form p/q where p and $q \neq 0$ are integers. Every sub-field of $\mathbb{R}$ contains $\mathbb{Q}$.

An example of another subfield is $\mathbb{Q}(\sqrt{2})$, the set of numbers of the form $a + b\sqrt{2}$ where a and b are rational numbers. It's clear that this set is closed under addition, subtraction, and multiplication, but it's not so clear that it's closed under division, but the following calculation with the number $a + b\sqrt{2} \neq 0$

$$\frac{1}{a+b\sqrt{2}} = \frac{(a-b\sqrt{2})}{(a+b\sqrt{2})(a-b\sqrt{2})} = \frac{(a-b\sqrt{2})}{a^2-2b^2} = \frac{a}{a^2-2b^2} - \frac{b}{a^2-2b^2}\sqrt{2},$$

shows that $\mathbb{Q}(\sqrt{2})$ is closed under taking multiplicative inverses of nonzero numbers, so it's closed under division as well. [3]

Note the important fact that $\mathbb{Q}(\sqrt{2})$ is, itself, a vector space over $\mathbb{Q}$ of dimension 2, with basis 1 and $\sqrt{2}$.

Similarly, one can prove that $\mathbb{Q}(\sqrt{2}, \sqrt{3})$, i.e., the set of numbers of the form

$$x = a + b\sqrt{2} + c\sqrt{3} + d\sqrt{6} = a + b\sqrt{2} + c\sqrt{3} + d\sqrt{2}\sqrt{3}$$

[3]Note that $a^2 - 2b^2 \neq 0$, since, otherwise, we'd have $2 = (a/b)^2$, but 2 is not the square of a rational number.

where a, b, c, and d belong to $\mathbb{Q}$, is a sub-field of $\mathbb{R}$ and has dimension 4 over $\mathbb{Q}$. (The trick is to notice that

$$(a+b\sqrt{2}+c\sqrt{3}+d\sqrt{6})(a+b\sqrt{2}-c\sqrt{3}-d\sqrt{6}) = (a+b\sqrt{2})^2-3(c+d\sqrt{2})^2 \in \mathbb{Q}(\sqrt{2}),$$

since this shows that

$$\frac{1}{(a+b\sqrt{2}+c\sqrt{3}+d\sqrt{6})} = \frac{(a+b\sqrt{2}-c\sqrt{3}-d\sqrt{6})}{(a+b\sqrt{2})^2-3(c+d\sqrt{2})^2},$$

and we know that $(a+b\sqrt{2})^2-3(c+d\sqrt{2})^2$, which lies in $\mathbb{Q}(\sqrt{2})$ has an inverse in $\mathbb{Q}(\sqrt{2})$.

Note that $\mathbb{Q}(\sqrt{2},\sqrt{3})$, as a vector space over $\mathbb{Q}$, has dimension 4. (What is a basis?)

A sub-field $\mathbb{F} \subset \mathbb{R}$ that has finite dimension as a vector space over $\mathbb{Q}$ is said to be a *real number field*. One of the basic theorems of abstract algebra is that, if $\mathbb{F}$ is a number field that has dimension n over $\mathbb{Q}$, then there is an element $x \in \mathbb{F}$ that satisfies a polynomial of degree n

$$p(x) = x^n + a_1x^{n-1} + a_2x^{n-2} + \cdots + a_n = 0,$$

where the a_i are rational numbers. Also, the set $\{1, x, x^2, \ldots, x^{n-1}\}$ turns out to be a $\mathbb{Q}$-basis for $\mathbb{F}$, and the polynomial $p(t) = t^n + a_1t^{n-1} + a_2t^{n-2} + \cdots + t_n$ is *irreducible*, i.e., it is not the product of two polynomials with rational coefficients with degrees lower than n.

Conversely, every irreducible polynomial $p(t)$ with rational coefficients that has a real root $x \in \mathbb{R}$ gives rise to a real number field $\mathbb{F}$, which is the set of all numbers of the form

$$q = r_0 + r_1\,x + r_2\,x^2 + \cdots + r_{n-1}\,x^{n-1},$$

where the r_i are rational numbers.

Another field that turns out to be very useful is the sub-field $\mathbb{K} \subset \mathbb{R}$ that consists of all the possible (real) numbers one can make starting with $\mathbb{Q}$ and using the operations of addition, subtraction, multiplication, division by nonzero elements, and *taking square roots of positive numbers*. It takes a little work to show that $\mathbb{K}$ actually *is* a field (and that it is not all of $\mathbb{R}$).

The field $\mathbb{K}$ is *not* a number field because it has infinite dimension over $\mathbb{Q}$, but it has a very interesting property: *Any subfield $\mathbb{F} \subset \mathbb{K}$ that does have finite dimension n over $\mathbb{Q}$ has to satisfy $n = 2^k$ for some integer k.*

In particular, $\mathbb{K}$ does not contain $2^{1/3}$, the cube root of 2, because $\mathbb{Q}(2^{1/3})$ has a basis $\{1, 2^{1/3}, 2^{2/3}\}$, so it has dimension 3 over $\mathbb{Q}$, and 3 is not a power of 2. In fact, for the same reason, it doesn't contain any $x \in \mathbb{R}$ that satisfies an irreducible rational polynomial of degree 3.

3.6.2. *On ruler-and-compass constructions.* Now let $S(\mathbb{K}) \subset S$ denote the set of S-points (u, v, w) such that u, v, and w belong to $\mathbb{K}$. Also, take the lines in $S(\mathbb{K})$ to be the S-lines $au + bv - cw = 0$ for which a, b, and c belong to $\mathbb{K}$. Then it is not hard to show that $S(\mathbb{K})$ is a Hilbert plane (where we use the same notions of congruence of segments and angles as in the whole of S).

Moreover, $S(\mathbb{K})$ satisfies Archimedes' Axiom and the Circle-Circle Continuity Principle. *This implies, in particular, that any ruler-and-compass construction can be carried out in $S(\mathbb{K})$.*

However, $S(\mathbb{K})$ doesn't satisfy either Dedekind's Axiom or the Parallel Postulate. Here is an example of how this fact can be used:

Proposition 2. *There is no ruler-and-compass construction in neutral geometry that allows us to trisect a general segment.*

Proof. Consider the points $A = (0, 0, 1)$ and $B = (\sqrt{3}, 0, 2)$ in $S(\mathbb{K})$. Then $\overline{AB} = \cosh^{-1}(-A \cdot B) = \cosh^{-1}(2)$. If $C = (u, v, w) \in S$ is the point on AB such that $\overline{AC} = \frac{1}{3}\overline{AB}$, then $\overline{AC} = \frac{1}{3}\cosh^{-1}(2)$. So

$$w = A \cdot C = -\cosh\bigl(\tfrac{1}{3}\cosh^{-1}(2)\bigr).$$

Obviously, if C lies in $S(\mathbb{K})$, then w lies in $\mathbb{K}$. But now I am going to show that w does not lie in $\mathbb{K}$. The reason is the *triple argument formula* for hyperbolic cosine:

$$\cosh(3r) = 4\cosh(r)^3 - 3\cosh(r).$$

Taking $r = \frac{1}{3}\cosh^{-1}(2)$, we get

$$2 = \cosh(3r) = -4w^3 + 3w,$$

so w satisfies the polynomial equation $4w^3 - 3w + 2 = 0$, but the polynomial $p(t) = 4t^3 - 3t + 2$ does not factor with rational coefficients (why not?), so, by the above discussion, w does not lie in $\mathbb{K}$, so C does not lie in $S(\mathbb{K})$.

Now, as mentioned above, every ruler-and-compass construction can be carried out in $S(\mathbb{K})$, so if there were a ruler-and-compass construction that worked in neutral geometry to trisect segments, then it could be applied in this case to the A and B given above and would produce C as an element of $S(\mathbb{K})$, which as we have seen, cannot be done. □

4. The Poincaré Disk Model

The two models we have found so far have different advantages. In the Klein model K, it's easy to see incidence, betweenness, parallelism, etc., but it's hard to see (or even compute) congruence of segments or angles. In the upper half-sheet model S, it's easy to compute distances and angles, and motions are just linear transformations that preserve the Minkowski inner product, but it's three dimensional, and the Minkowski inner product is a little non-intuitive at first.

There is another disk model, the *Poincaré disk*, that has the advantage that it is 2-dimensional and incidence and betweenness are easy to see (not quite as easy as the Klein model), but the big advantage is that angle congruence is very easy to understand.

4.1. **The points of the model.** The points of the Poincaré disk P are again going to be the points *inside* the unit circle, i.e., the points of the form (s, t) where $s^2 + t^2 < 1$. (Just as in the K-model, the unit circle $s^2 + t^2 = 1$ will be referred to as the *ideal circle*, and its points will be referred to as *ideal points*. But, just to be clear: Ideal points are not P-points.)

I am going to define a mapping from P-points to K-points as follows. For a P-point (s, t), set

$$h(s, t) = \left(\frac{2s}{(1+s^2+t^2)}, \frac{2t}{(1+s^2+t^2)}\right) = (x, y) \in K.$$

Thus, $h : P \to K$. Note that, using this formula

$$1 - x^2 - y^2 = 1 - \frac{4s^2 + 4t^2}{(1{+}s^2{+}t^2)^2} = \frac{(1 - s^2 - t^2)^2}{(1 + s^2 + t^2)^2} > 0,$$

so (x, y) is, indeed, a point in K.

Exercise 4.1. Show that h is one-to-one by showing that the inverse mapping is given by

$$h^{-1}(x, y) = \left(\frac{x}{1 + \sqrt{1{-}x^2{-}y^2}}, \frac{y}{1 + \sqrt{1{-}x^2{-}y^2}} \right) = (s, t) \in P.$$

Hint: solve the above equations for s and t, using the relation derived above:

$$\sqrt{1{-}x^2{-}y^2} = \frac{(1 - s^2 - t^2)}{(1 + s^2 + t^2)}.$$

Exercise 4.2. Show that the formula

$$p(u, v, w) = \left(\frac{u}{1 + w}, \frac{v}{1 + w} \right) = (s, t)$$

defines a mapping $p : S \to P$ from the hyperboloid upper sheet model to the Poincaré disk that is one-to-one and onto and that it satisfies

$$p\big(g(x, y)\big) = h^{-1}(x, y) = (s, t),$$

so all three models are related by these mappings.

Show also that the inverse mapping $p^{-1} : P \to S$ is given by

$$p^{-1}(s, t) = \left(\frac{2s}{(1 - s^2 - t^2)}, \frac{2t}{(1 - s^2 - t^2)}, \frac{(1 + s^2 + t^2)}{(1 - s^2 - t^2)} \right).$$

(This fact will come in handy below.)

4.2. **Lines in P.** We might as well define the P-lines so that the mappings $h : P \to K$ and $h^{-1} : K \to P$ carry collinear points to collinear points. This will, of course, mean that a *line* in P will be defined by an equation of the form

$$0 = a\,x + b\,y - c = a\left(\frac{2s}{(1{+}s^2{+}t^2)} \right) + b\left(\frac{2t}{(1{+}s^2{+}t^2)} \right) - c$$

where $a^2 + b^2 - c^2 > 0$. Clearing fractions, this is

$$c\,(1{+}s^2{+}t^2) - 2a\,s - 2b\,t = 0,$$

which is the equation of an ordinary line through $(0, 0)$ when $c = 0$ and the equation of a circle

$$(s - a/c)^2 + (t - b/c)^2 = (a/c)^2 + (b/c)^2 - 1 > 0$$

when $c \neq 0$.

Exercise 4.3. Explain why the circle above has its center $C = (a/c, b/c)$, *outside* the circle $s^2 + t^2 = 1$. The (Euclidean) radius of the circle is $r = \sqrt{(a/c)^2 + (b/c)^2 - 1}$. Explain why this circle meets $s^2 + t^2 = 1$ in two points, say P and Q , and explain why the triangles APC and AQC (where $A = (0, 0)$ is the (Euclidean) center of the ideal circle $s^2 + t^2 = 1$) have right angles at P and Q.

Conversely, show that, if a circle $(s - p)^2 + (t - q)^2 = r^2$ meets the ideal circle $s^2 + t^2 = 1$ in two points at right angles to the ideal circle, then $p^2 + q^2 > 1$ and $r^2 = p^2 + q^2 - 1$, so we can write $p = a/c$, $q = b/c$ where $a^2 + b^2 - c^2 > 0$.

This motivates the following:

Definition 7. A *P-line* is either a diameter of the ideal circle $s^2 + t^2 = 1$ or the arc inside the ideal circle of a circle that intersects the ideal circle at right angles.

4.3. **Segment congruence.** We'll just transfer the notion of segment congruence directly from the easy one for the S-model: If $A = (s_1, t_1)$ and $B = (s_2, t_2)$, then we define the P-length of the segment AB so that

$$\cosh \overline{AB} = \frac{(1 + {s_1}^2 + {t_1}^2)(1 + {s_2}^2 + {t_2}^2) - 4s_1s_2 - 4t_1t_2}{(1 - {s_1}^2 - {t_1}^2)(1 - {s_2}^2 - {t_2}^2)}.$$

You might want to check, using the Exercises above, that this agrees with our previous definition, in the sense that

$$-A \cdot B = \cosh \overline{AB} = \cosh \overline{p(A)p(B)}$$

for any two distinct S-points A and B.

Exercise 4.4. Explain why the P-circle with center $C = (s_1, t_1) \in P$ and radius $r > 0$ is described by an equation in the st-plane of the form

$$(1 + {s_1}^2 + {t_1}^2)(1 + s^2 + t^2) - 4s_1s - 4t_1t = (\cosh r)(1 - {s_1}^2 - {t_1}^2)(1 - s^2 - t^2),$$

which can be rearranged and written in the form

$$s^2 + t^2 - 2p\, s - 2q\, t - v = 0$$

for some real numbers p, q, and v. In particular, this is the equation of a *Euclidean* circle in the st-plane! Thus, P-circles (as sets) are Euclidean circles!

Is is necessarily true that the P-center of a P-circle is the Euclidean center of the circle? Can you find three non-collinear P-points such that there is no P-circle that passes through them? (Remember that we did not prove, in neutral geometry, that every triangle can be *circumscribed* by a circle. This example explains why we didn't try to prove this.)

4.4. **Angle congruence.** The really nice thing about the Poincaré model is that angle congruence is the same as Euclidean angle congruence! In fact, using a little bit of (tedious, but not difficult) high school algebra, it can be checked that two P-rays (i.e., Euclidean circles or lines through the origin), $\overrightarrow{XY}$ and $\overrightarrow{XZ}$ defined by the respective equations

$$c_1\,(s^2 + t^2 + 1) - 2a_1\, s - 2b_1\, t = 0,$$
$$c_2\,(s^2 + t^2 + 1) - 2a_2\, s - 2b_2\, t = 0,$$

such that the left hand side of the first (respectively, second) equation is positive on Z (respectively, Y) form a Euclidean angle θ if and only if

$$\cos\theta = \frac{c_1c_2 - a_1a_2 - b_1b_2}{\sqrt{{a_1}^2 + {b_1}^2 - {c_1}^2}\,\sqrt{{a_2}^2 + {b_2}^2 - {c_2}^2}},$$

(As we have seen, the corresponding pair of K-lines in K meet if and only if the right hand side of this equation is between -1 and 1.) In other words, *the P-radian angle measure between P-lines is the* same *as the Euclidean angle measure.*

5. The Poincaré Upper Half-Plane Model

Our final model is also due to Henri Poincaré, and is known as the *upper half-plane model U*. As usual, I will start by defining the U-points and U-lines.

5.1. **The points.** A *point* in the model U is a point $(a, b) \in \mathbb{R}^2$ that satisfies $b > 0$. Thus, the U-points are the ones that lie in the upper half-plane of $\mathbb{R}^2$.

5.2. **The lines.** The *lines* in the model U are the vertical coordinate lines together with the Euclidean circles in the ab plane with centers on the a-axis. Thus, a line is defined by an equation of the form

$$p\,(a^2 + b^2) - 2q\,a + r = 0$$

where p, q, and r are (real) constants satisfying $q^2 - pr > 0$. (Why do we need this inequality?) Note that replacing (p, q, r) by $(\lambda p, \lambda q, \lambda r)$ merely multiplies the equation by a constant, so it doesn't change the curve that it defines.

A U-line is incident with a U-point if and only if the curve (circle or line) passes through the point in the Euclidean sense. (Note that $p = 0$ if and only if the U-line is an Euclidean line.)

Exercise 5.1. Show that the (unique) U-line though the U-points $P = (a_1, b_1)$ and $Q = (a_2, b_2)$ has its equation of the above form, where

$$\begin{aligned} p &= 2(a_2 - a_1), \\ q &= {a_2}^2{+}{b_2}^2{-}{a_1}^2{-}{b_1}^2, \\ r &= 2a_1({a_2}^2{+}{b_2}^2) - 2a_2({a_1}^2{+}{b_1}^2). \end{aligned}$$

Explain why the U-model satisfies the incidence and betweenness axioms. (Be sure to explain how you understand betweenness.)

5.3. **An identification between P and U.** There is a mapping $\phi : P \to U$, defined by

$$(a, b) = \phi(s, t) = \left(\frac{-2t}{(s-1)^2 + t^2}, \frac{1 - s^2 - t^2}{(s-1)^2 + t^2}\right)$$

that is,

$$a = \frac{-2t}{(s-1)^2 + t^2}, \qquad b = \frac{1 - s^2 - t^2}{(s-1)^2 + t^2}.$$

Then ϕ is one-to-one and onto, and its inverse $\phi^{-1} : U \to P$, is defined by

$$s = \frac{a^2 + b^2 - 1}{a^2 + (b+1)^2}, \qquad t = \frac{-2a}{a^2 + (b+1)^2}.$$

Exercise 5.2. Show that the correspondence ϕ carries P-lines to U-lines and conversely. Thus, it preserves incidence and betweenness. (Hint: Use the above substitutions in the equations defining P-lines or U-lines and verify that this correspondence makes them correspond.)

Remark 12. The map ϕ may seem a little strange at first. It comes from complex analysis, and has a rather natural formula there: If we write $z = s + i\,t$ and use this to identify the complex plane $\mathbb{C}$ with $\mathbb{R}^2$, and, particularly, the Poincaré disk ($s^2+t^2 < 1$) with the unit complex disk defined by $|z|^2 = z\bar{z} < 1$, then the mapping ϕ can be written in the *linear fractional form*

$$a + i\,b = w = \frac{i(1+z)}{(1-z)},$$

because, when one writes out the latter expression, one obtains

$$\frac{i(1+z)}{(1-z)} = \frac{-t + i\,(1+s)}{(1-s) - i\,t} = \frac{-2t + i\,(1-s^2-t^2)}{(s-1)^2 + t^2}\,.$$

You might want to check that solving for z in terms of w, i.e.,

$$s + i\,t = z = \frac{iw+1}{iw-1} = \frac{i(a+i\,b)+1}{i(a+i\,b)-1}$$

gives the form of the inverse mapping of ϕ.

5.4. **Segment and Angle Congruence.** Since we know that $\phi : P \to U$ preserves incidence and betweenness, we can use ϕ to transfer the notions of segment and angle measure from P to U, and this will complete our definition of the U-model.

Exercise 5.3. Show (using a previous exercise from Section 4), that, if we define the U-distance $\overline{AB}$ between U-points $A = (a_1, b_1)$ and $B = (a_2, b_2)$ by the rule

$$\cosh \overline{AB} = \frac{(a_1 - a_2)^2 + {b_1}^2 + {b_2}^2}{2b_1 b_2}\,, \tag{5.1}$$

then the map $\phi : P \to U$ preserves distances. (Hint: This is just algebra, but, in order to make the calculation manageable by hand, you want to first see how the various parts simplify, such as how $1 - {s_1}^2 - {t_1}^2$ simplifies when you make the substitution that expresses s_1 and t_1 in terms of a_1 and b_1, etc.)

Remark 13. The fact that the above formula for distance is much simpler than the formula for distance in the P-model (or the K-model for that matter) is what makes the U-model particularly attractive.

Exercise 5.4. Use the above formula for distance to explain why a U-circle (i.e., the set of U-points X that have a fixed U-distance $r > 0$ from a given U-center $C = (a, b)$) is an Euclidean circle.

Exercise 5.5. Show that two U-rays $\overrightarrow{PQ}$ and $\overrightarrow{PR}$ defined by a pair of equations

$$p_1\,(a^2 + b^2) - 2q_1\,a + r_1 = 0,$$
$$p_2\,(a^2 + b^2) - 2q_2\,a + r_2 = 0,$$

such that each equation for a given ray is positive on the other ray (except at P, of course), meet at an Euclidean angle of θ if and only if

$$\cos\theta = -\frac{\bigl(q_1 q_2 - \tfrac{1}{2}(p_1 r_2 + p_2 r_1)\bigr)}{\sqrt{{q_1}^2 - p_1 r_1}\,\sqrt{{q_2}^2 - p_2 r_2}}.$$

If the number on the right hand side is not between -1 and 1, then the two U-lines do not meet. (Hint: By scaling, one can assume that ${q_1}^2 - p_1 r_1 = 1$ and ${q_2}^2 - p_2 r_2 = 1$, i.e., that the equations for the lines are *normalized*).

Remark 14. One can now show (though, even assuming several of the above exercises, it is somewhat tedious) that the Euclidean radian measure of angles between U-rays is the same as the radian U-measure of U-angles. The easiest way to do it using algebra is to compare with the S-model, but there is a more conceptual way to do it, which is to use the fact (proved in every complex analysis course) that complex linear fractional transformations (such as the map $w = i(1+z)/(1-z)$ described above) always preserve angles and then quote the result from the Poincaré disk.

5.5. **Motions.** Another great advantage of the U-model is that it is easy to write down motions that preserve U-distance between points.

For example, the *scaling transformation*

$$S_\lambda(a, b) = (\lambda a, \lambda b)$$

when $\lambda > 0$ carries the points of U into themselves, carries U-lines to U-lines and preserves U-distance (and hence angles).

As another example, the *translation transformation*

$$T_p(a, b) = (a + p, b)$$

for all $p \in \mathbb{R}$ carries the points of U into themselves, carries U-lines to U-lines and preserves U-distance (and hence angles).

Finally, as a third (less immediate) example, the *inversion transformation*

$$I(a, b) = \left(\frac{a}{a^2 + b^2}, \frac{b}{a^2 + b^2} \right).$$

This transformation fixes points on the Euclidean unit circle $a^2 + b^2 = 1$ and carries points inside the Euclidean unit circle to points outside and *vice versa*. The map $I : U \to U$ preserves U-distance (see the Exercise below). (It also satisfies $I(I(P)) = P$ for all U-points P. Since the Euclidean unit circle is a U-line, this map is just reflection across this line.)

Exercise 5.6. Verify that S_λ, T_p, and I do, in fact, preserve U-distance. In other words, if P and Q are (distinct) U-points then

$$\overline{S_\lambda(P) S_\lambda(Q)} = \overline{T_p(P) T_p(Q)} = \overline{I(P) I(Q)} = \overline{PQ}.$$

(Hint: Because $\cosh : [0, \infty) \to [1, \infty)$ is one-to-one and onto, it suffices to prove that the hyperbolic cosines of all these numbers are the same. Now let $P = (a_1, b_1)$ and $Q = (a_2, b_2)$ and compute. The hardest case (still not very hard) is for the inversion I.)

Since any point of U can be carried to any other point of U by a combination of scaling and translation, it follows that we can understand angles and distances by looking at them for a single U-point, say $A = (0, 1)$ $(= \phi(0, 0)$ when $(0, 0) \in P)$.

Exercise 5.7. Show that, if a U-circle γ has center $C = (a, b)$ and passes through the points (a, b_+) and (a, b_-) where $0 < b_- < b < b_+$, then $b^2 = b_+b_-$. Since the *Euclidean* center of γ is $\bigl(a, \frac{1}{2}(b_+ + b_-)\bigr)$, the Euclidean center is always *higher* than the U-center. (Hint: First, treat the case where $C = (0, 1)$, let the U-radius of the circle be $r > 0$ and determine b_+ and b_-, verifying the formula in this case. Now, use Exercise 5.6 to show that, since this formula holds for $C = (0, 1)$, the formula must hold for any possible center.)

6. Curves and Calculus

This final section is about a somewhat more advanced topic: Curves in the various models and the notion of their *length*.

6.1. **The standard formulae.** You may remember from your calculus course how we computed the length of a curve: When it's presented as a graph $y = f(x)$ in $\mathbb{R}^2$ with $a \le x \le b$, the textbooks tell you that the length of the curve is

$$L = \int_a^b \sqrt{1 + f'(x)^2}\,\mathrm{d}x.$$

The usual justification is that this integral is approximated by the sum

$$S = \sum_{k=1}^{n} \sqrt{(x_k - x_{k-1})^2 + \bigl(f(x_k) - f(x_{k-1})\bigr)^2}$$

where $a = x_0 < x_1 < \cdots < x_{n-1} < x_n = b$ and where the k-th term in the sum is the (Euclidean) distance between the two points $A_{k-1} = \bigl(x_{k-1}, f(x_{k-1})\bigr)$ and $A_k = \bigl(x_k, f(x_k)\bigr)$ on the curve. In other words, S is the length of the polygonal path got by taking the sum of the length of the segments $A_{k-1}A_k$ from $k = 1$ to $k = n$. In calculus, we show that, if the function f is 'reasonable' (continuously differentiable is more than enough), then, as the maximum length of the individual segments $A_{k-1}A_k$ decreases, this will converge to the length of the graph.

More generally, given a curve $\gamma(\tau) = \bigl(x(\tau), y(\tau)\bigr)$ with $a \le \tau \le b$, the (Euclidean) length of $\gamma : [a, b] \to \mathbb{R}^2$ is

$$L = \int_a^b \sqrt{x'(\tau)^2 + y'(\tau)^2}\,\mathrm{d}\tau = \int_a^b \sqrt{\gamma'(\tau)\cdot\gamma'(\tau)}\,\mathrm{d}\tau.$$

For example, for the circle of radius r, we can take $\gamma(\tau) = (r\cos\tau, r\sin\tau)$ with $0 \le \tau \le 2\pi$, and we get the expected result

$$L = \int_0^{2\pi} \sqrt{\gamma'(\tau)\cdot\gamma'(\tau)}\,\mathrm{d}\tau = \int_0^{2\pi} r\,\mathrm{d}\tau = 2\pi r.$$

In the general case of a Hilbert plane M satisfying Dedekind's Axiom, we can use our measurement theorem to *define* a notion of length of a curve. (Of course, we need to know what we mean by a 'curve', but I'm going to take a very naïve definition of this: I'll require that, when we establish (any) coordinates in our Hilbert model, then a (differentiable) curve will be a mapping $\gamma : [a, b] \to M$ that is differentiable in coordinates.

In any case, our definition of the length of γ will be that it will be the least upper bound of all the numbers S of the form

$$S = \sum_{k=1}^{n} \overline{\gamma(\tau_{k-1})\gamma(\tau_k)},$$

where $a = \tau_0 < \tau_1 < \tau_2 < \cdots < \tau_{n-1} < \tau_n = b$, i.e., the sum of the lengths of the segments in any polygonal path approximation to the curve γ.

Now, just as in the case of Euclidean geometry, as n goes to infinity, these sums increase and go to a limit (when the curve is continuously differentiable), which is called the length of the curve in the Hilbert plane M.

6.2. **Length of curves in S.** It turns out that the easiest case to work in is the case of the hyperbolic sheet model $S \subset \mathbb{R}^3$. There, it turns out that, when $\gamma : [a, b] \to S$ is a differentiable curve, i.e.,

$$\gamma(\tau) = \bigl(u(\tau), v(\tau), w(\tau)\bigr)$$

where u, v, and w are differentiable functions satisfying $u(\tau)^2 + v(\tau)^2 - w(\tau)^2 = -1$, then the formula for length is

$$L = \int_a^b \sqrt{\gamma'(\tau) \cdot \gamma'(\tau)}\, \mathrm{d}\tau = \int_a^b \sqrt{u'(\tau)^2 + v'(\tau)^2 - w'(\tau)^2}\, \mathrm{d}\tau,$$

exactly as in the Euclidean case, but using the Minkowskian inner product!

Here is how this is justified: For starters, consider the case of a line, say $\gamma(\tau) = (\sinh\tau, 0, \cosh\tau)$. Then $\gamma'(\tau) = (\cosh\tau, 0, \sinh\tau)$, so $\gamma'(\tau) \cdot \gamma'(\tau) = 1$, so

$$L = \int_a^b 1\, \mathrm{d}\tau = b - a,$$

which agrees with the fact that the distance between $\gamma(a)$ and $\gamma(b)$ is

$$\cosh^{-1}\bigl(-\gamma(a){\cdot}\gamma(b)\bigr) = \cosh^{-1}\bigl(\cosh(a - b)\bigr) = |a - b| = b - a$$

(when $a < b$). Thus, the above formula gives the right answer for all the line segments on that particular line. However, because the length formula uses only the inner product and any two line segments of the same length can be matched up using a rigid motion (which does not change the inner product), this means that the above formula assigns the right length to all line segments.

Since the formula works for all line segments and since length of a curve is *defined* as the supremum of all lengths of polygonal paths got by subdividing the curve, it follows that this formula defines a length for curves in the hyperbolic plane.

What does this give for the circle of radius r? Well, by use of motions, we can assume that the center of the circle is at $A = (0, 0, 1)$, then $A \cdot X = -\cosh r$ (the equation for a circle of radius r centered at A), implies that

$$X = \bigl(\sinh r\, \cos\tau,\ \sinh r\, \sin\tau,\ \cosh r\bigr) = \gamma(\tau)$$

for some $0 \le \tau \le 2\pi$. Now we can compute that $\gamma'(\tau) \cdot \gamma'(\tau) = \sinh^2 r$, so we get

$$L = \int_0^{2\pi} \sinh r\, \mathrm{d}t = 2\pi \sinh r.$$

Thus, the circumference of the circle of radius r grows *very* rapidly with r since, for large r, we have $\sinh r = (e^r - e^{-r})/2 \approx e^r/2$, so $L \approx \pi e^r$ for sufficiently large r!

Exercise 6.1. Check this answer as follows: Let γ be a circle of radius r in the hyperbolic plane with center O and let $R_0, R_1, \ldots, R_n = R_0$ be $n \ge 3$ points on γ that are 'equally spaced' around γ in the sense that the ray from O to R_i is between the rays from O to R_{i-1} and R_{i+1} and each of the angles $\angle R_{i-1}OR_i$ has radian measure $2\pi/n$. Use the hyperbolic Law of Cosines to show that

$$\begin{aligned}\cosh(\,\overline{R_{i-1}R_i}\,) &= \cosh^2 r - \sinh^2 r\cos(2\pi/n)\\ &= 1+\sinh^2 r\bigl(1-\cos(2\pi/n)\bigr) = 1+2\sinh^2 r\sin^2(\pi/n).\end{aligned}$$

Then, using the identity $\cosh x - 1 = 2\sinh^2(x/2)$, conclude that

$$\overline{R_{i-1}R_i} = 2\sinh^{-1}\bigl(\sin(\pi/n)\sinh r\bigr).$$

Thus, the total perimeter of the n-gon $R_0R_1R_2\cdots R_{n-1}R_nR_0$ is

$$S_n = 2n\sinh^{-1}\bigl(\sin(\pi/n)\sinh r\bigr),$$

and the limit as n tends to infinity is

$$L = \lim_{n\to\infty} S_n = \lim_{n\to\infty} 2n\sinh^{-1}\bigl(\sin(\pi/n)\sinh r\bigr) = \lim_{t\to 0^+}\frac{2\sinh^{-1}\bigl(\sin(\pi t)\sinh r\bigr)}{t}.$$

Use calculus (specifically, L'Hôpital's Rule and the Chain Rule) to show that his limit is $L = 2\pi\sinh r$. (Hint: To compute the derivative of $\sinh^{-1}(u)$ at $u=0$, use the fact that the derivative of $\sinh(v)$ at $v=0$ is $\cosh(v) = \cosh 0 = 1$ and the usual formula for inverse funtions.)

Exercise 6.2. Show, that, if $\gamma : [a,b]\to S$ is described in 'hyperbolic polar coordinates' in the form

$$\gamma(\tau) = \bigl(\sinh r(\tau)\,\cos\phi(\tau),\ \sinh r(\tau)\,\sin\phi(\tau),\ \cosh r(\tau)\bigr)$$

for two functions $r(\tau)$ and $\phi(\tau)$ while $a\le\tau\le b$, then

$$L = \int_a^b \sqrt{r'(\tau)^2 + \bigl(\sinh r(\tau)\bigr)^2\,\phi'(\tau)^2}\,\mathrm{d}\tau.$$

(Hint: This is just calculus and the definition: You just need to compute $\gamma'(\tau)\cdot\gamma'(\tau)$ for γ in this form.)

6.3. **Length of curves in P.** The formula for hyperbolic length of curves in the Poincaré disk turns out to be this: Given a curve $\bigl(s(\rho),t(\rho)\bigr)$ within P, the P-length of the curve traced out in P as ρ goes from a to b can be shown to be

$$L = 2\int_a^b \frac{\sqrt{s'(\rho)^2+t'(\rho)^2}}{1-s(\rho)^2-t(\rho)^2}\,\mathrm{d}\rho.$$

Exercise 6.3. Use the formula for $p : S\to P$ given by

$$p(u,v,w) = \left(\frac{u}{1+w},\frac{v}{1+w}\right)$$

and that fact that p preserves distances between points to show that if $\gamma : [a,b]\to S$ is a differentiable curve in S of the form $\gamma(\rho) = \bigl(u(\rho),v(\rho),w(\rho)\bigr)$, then the above formula assigns the same length L to the S-curve γ and the P-curve $p\circ\gamma$ given by

$$\bigl(s(\rho),t(\rho)\bigr) = p\circ\gamma(\rho) = \left(\frac{u}{1+w},\frac{v}{1+w}\right).$$

(Hint: You will need to use the facts that $\gamma\cdot\gamma = -1$ and $\gamma\cdot\gamma' = 0$ when working out the formula for the P-curve.)

6.4. **Length of curves in U.** Perhaps the simplest of all is the formula for length of curves in the model U. It turns out that, for a differentiable curve $\gamma : [\tau_0, \tau_1] \to U$ of the form $\gamma(\tau) = \big(a(\tau), b(\tau)\big)$, we have

$$L = \int_{\tau_0}^{\tau_1} \frac{\sqrt{a'(\tau)^2 + b'(\tau)^2}}{b(\tau)}\, \mathrm{d}\tau.$$

One can prove this, given the formula for the length of curves in the model P, by using the transformation rule relating P to U that we have used before, i.e., $\phi : P \to U$ given by

$$(a, b) = \phi(s, t) = \left(\frac{-2t}{(s-1)^2 + t^2}, \frac{1 - s^2 - t^2}{(s-1)^2 + t^2}\right),$$

and calculus. It's a little bit messy, so here is a simple example for you to try:

Exercise 6.4. Check that this integral formula gives us the right formula for distance between two points $(0, b_1)$ and $(0, b_2)$ for $b_2 > b_1 > 0$ by using the parametrization of the segment $\gamma(\tau) = (0, \tau)$ with $b_1 \leq \tau \leq b_1$ and comparing it with the given definition of the distance in Exercise 5.3.

6.5. **Length of curves in K.** Finally, a little calculus using the coordinate changes between P and K shows that the formula for the length of curves in K is as follows: For a differentiable curve $\gamma : [\tau_0, \tau_1] \to K$ of the form $\gamma(\tau) = \big(x(\tau), y(\tau)\big)$, we have

$$L = \int_{\tau_0}^{\tau_1} \frac{\sqrt{\big(1-y(\tau)^2\big)\, x'(\tau)^2 + 2x(\tau)y(\tau)x'(\tau)y'(\tau) + \big(1-x(\tau)^2\big)\, y'(\tau)^2}}{\big(1-x(\tau)^2-y(\tau)^2\big)}\, \mathrm{d}\tau$$

(The relative complexity of this formula indicates why it's more difficult to calculate distances and angles in K than in P, H, or U.)

References

[1] William F. Reynolds, *Hyperbolic Geometry on a Hyperboloid*, The American Mathematical Monthly,**100** (1993), 442–455. http://www.jstor.org/stable/2324297

Duke University Mathematics Department, P.O. Box 90320, Durham, NC 27708-0320
Email address: `bryant@math.duke.edu`
URL: `http://www.math.duke.edu/~bryant`