The Hyperbolic Plane

Rich Schwartz

November 19, 2014

1 Euclid's Postulates

Hyperbolic geometry arose out of an attempt to understand Euclid's fifth postulate. So, first I am going to discuss Euclid's postulates. Here they are:

1. Given any two distinct points in the plane, there is a line through them.
2. Any line segment may be extended to a line.
3. Given any point and any radius, there is a circle with that radius centered at that point.
4. All right angles are equal to one another.
5. Suppose that you have two lines L_1 and L_2 and a third line M. Suppose that M intersects L_1 and L_2 at two interior angles whose sum is less than the sum of two right angles on one side. Then L_1 and L_2 meet on that side. (See Figure 1.)

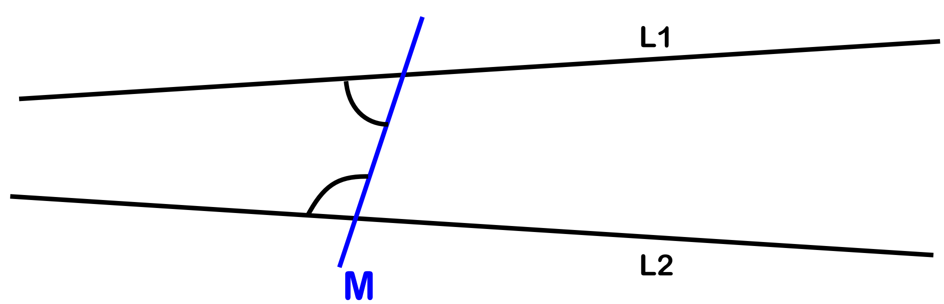

Figure 1: Euclid's fifth postulate.

Euclid's fifth postulate is often reformulated like this: For any line L and any point p not on L, there is a unique line L' through p such that L and L' do not intersect – i.e., are parallel. This is why Euclid's fifth postulate is often called the parallel postulate.

Euclid's postulates have a quaint and vague sound to modern ears. They have a number of problems. In particular, they do not uniquely pin down the Euclidean plane.

A modern approach is to make a concrete model for the Euclidean plane, and then observe that Euclid's postulates hold in that model. So, from a modern point of view, the Euclidean plane is $\boldsymbol{R}^2$, namely the set of ordered pairs of real numbers. The elements of $\boldsymbol{R}^2$ are called *points*. A *line* in $\boldsymbol{R}^2$ is defined to be the solution set to a linear equation of the form

$$\{(x,y)|\ Ax+By+C=0\},$$

where $A,B,C \in \boldsymbol{R}$. The *distance* between two points (x_1,y_1) and (x_2,y_2) is defined to be

$$\sqrt{(x_2-x_1)^2+(y_2-y_1)^2},$$

and this distance makes $\boldsymbol{R}^2$ into a metric space. Suppose that two lines intersect at a point p. Let C be a small circle centered at p. The two lines divide C into 4 arcs, having length A and B. The angles between the two lines are $\pi A/(A+B)$ and $\pi B/(A+B)$. These numbers don't depend on the radius of the circle chosen.

Euclid's postulates hold in the above model of the Euclidean plane. However, they also hold when we make the same definitions, except that we use $\overline{\boldsymbol{Q}} \cap \boldsymbol{R}$ in place of $\boldsymbol{R}$. here $\overline{\boldsymbol{Q}} \cap \boldsymbol{R}$ is the set of real number solutions to polynomials with integer coefficients.

For about 2000 years, mathematicians worked with Euclid's postulates, but many felt uneasy about the parallel postulate. It seemed more complicated than the others. So, a major unsolved problem arose: Deduce the fifth postulate from the other four. This problem was unsolved for about 2000 years!

2 The Hyperbolic Plane

In the early part of the 1800s, Three mathematicians, Janos Bolyai, Nicolai Lobachevski, and Karl Gauss, independently discovered that the parallel postulate did not follow from the other of Euclid's postulates. These three mathematicians found models in which the first four postulates held but the fifth did not. Their models, all isomorphic to each other, are now known as the *hyperbolic plane*.

The way I have stated the history of this thing is a vast simplification of a complicated story. There is a priority dispute. Gauss never published his own account, because he felt that the world was not yet ready for hyperbolic geometry. Knowing Gauss, who was possibly the smartest being ever to walk the planet and also a complete hard-ass, his claim seem very plausible. Also, Gauss knew/invented so much other mathematics (e.g. hypergeometric functions) which fit very naturally into the hyperbolic world.

So, here is a model for a hyperbolic plane: As a set, it consists of complex numbers $x + iy$ with $y > 0$. Geometrically, the hyperbolic plane is the open upper half plane – everything above the real axis. This set is denoted $\boldsymbol{H}^2$.

Geodesics: A *geodesic* in $\boldsymbol{H}^2$ is either a semicircle meeting the real axis at right angles, or a vertical ray emanating from a point on the real axis. See Figure 2. (Since the word *line* is already taken, we're using a different word for the objects which play the role of lines in the hyperbolic plane.)

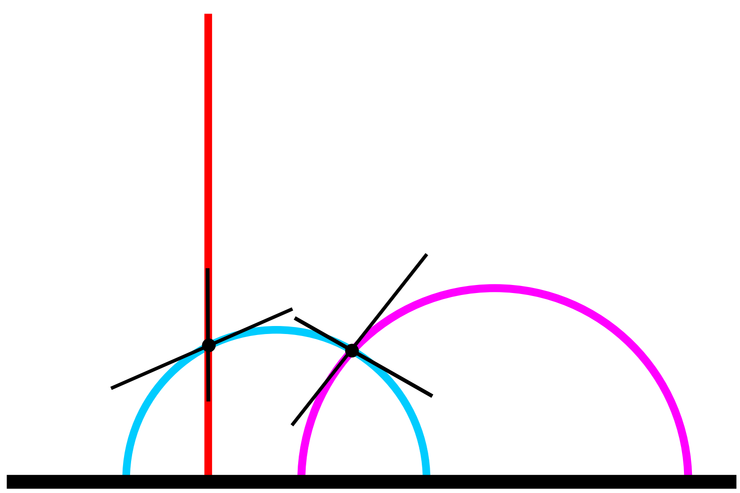

Figure 2: geodesics and angles in the hyperbolic plane

Angle: The *angle* between two geodesics is defined to be the angle between the tangents to the geodesics at their intersection point. Figure 2 shows two examples.

Distance: Let p and q be two points in $\boldsymbol{H}^2$. Let o and r denote the points where the geodesic meets the real axis, as in Figure 3.

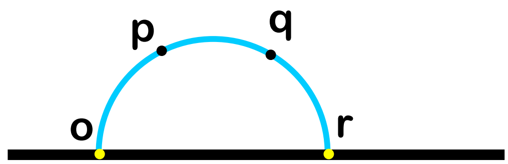

Figure 3

We define

$$\text{distance}(p,q) = \left| \ln \frac{|o-q||p-r|}{|o-p||q-r|} \right|. \tag{1}$$

Here $|o-p|$ computes the Euclidean distance from o to p. Likewise for the other terms.

When p and q lie on the same vertical segment, we interpret $r=\infty$. In this case, the second two factors cancel out, and Equation 1 is interpreted as

$$\text{distance}(p,q) = \left| \ln \frac{|o-q|}{|o-p|} \right|. \tag{2}$$

To see that Equation 2 is really a limiting case of Equation 1, consider the case when p and q are on a geodesic which is a very large Euclidean circle. In this case, r is an enormous real number, and $|r-p| \approx |r-q|$. So, the ratio of these two terms is almost 1. In the limit, these terms cancel out exactly.

It is pretty clear that our notion of distance satisfies the first two metric space axioms. That is,

1. $d(p,q) \geq 0$ with $d(p,q)=0$ if and only if $p=q$.

2. $d(p,q) = d(q,p)$ for all p and q.

The distance also satisfies the triangle inequality, for reasons I will explain below.

3 Symmetries

A symmetry of the metric space (X,d) is a bijection $f: X \to X$ which preserves the distance d. That is,

$$d(p,q) = d(f(p), f(q)),$$

for all $p, q \in X$. For example, the map $f(x, y) = (x, y) + (1, 0)$ is a symmetry of the Euclidean plane. Euclid's fourth postulate is really saying that the Euclidean plane is totally symmetric. In other words, if A_1, A_2 are a pair of perpendular lines, and B_1, B_2 are another pair of perpendicular lines, then there is a symmetry f such that $f(A_1) = B_1$ and $f(A_2) = B_2$. In other words, you can find symmetries which move any point to any other point, and which rotate by any angle. This is sometimes summarized by saying that the Euclidean plane is *homogeneous* and *isotropic*.

The hyperbolic plane also is homogeneous and isotropic. In order to explain this, I have to say what are the symmetries of the hyperbolic plane.

Lemma 3.1 *Let t be any real number. The map $f(z) = z + t$ is a symmetry of the hyperbolic plane.*

Proof: To see this, we will use Equation 1. If we replace p by $p' = p + t$ and q by $q' = q + t$ then o is replaced by $o' = o + t$ and r is replaced by $r' = r + t$. But then $|o' - p'| = |o - p|$, etc, so we get $d(p', q') = d(p, q)$. ♠

Lemma 3.2 *Let t be a positive real number. The map $f(z) = tz$ is a symmetry of the hyperbolic plane.*

Proof: Here o, p, q, r are replaced by to, tp, tq, tr, and we get the same answer in Equation 1 after making the replacements. ♠

Our last two results show that $\boldsymbol{H}^2$ is homogeneous, because we can map any point to any other point using a combination of the maps in the above two lemmas.

It turns out that $\boldsymbol{H}^2$ is also isotropic, but this takes a little more work to prove. The next series of lemmas will establish this fact. These lemmas are all about the map $f(z) = -1/z$.

Lemma 3.3 *The map $f(z) = -1/z$ is a bijection of the hyperbolic plane.*

Proof: If $p = x + iy$ with $y > 0$, then

$$f(p) = -1/p = \frac{-x + iy}{\sqrt{x^2 + y^2}}.$$

The imaginary part of $f(p)$ is positive. So, f maps $\boldsymbol{H}^2$ to itself. Note that $f(f(p)) = p$, so that f must be a bijection. ♠

Here is the key lemma, the result which makes hyperbolic geometry work.

Lemma 3.4 *The map $f(z) = -1/z$ maps hyperbolic geodesics to hyperbolic geodesics.*

Proof: We will consider the case when the geodesic is a circle which is not centered at 0 and does not contain 0. The other cases are limiting cases, and the result for them follows by taking limits. If S is such a circle and a is any real constant, $f(aS) = (-1/a)f(S)$. Hence, the result is true for S if and only if it is true for aS. So, it suffices to consider the case when S is centered at the point 1. The equation for S is $|z-1| = r$, where $r \neq 1$.

Let $w = f(z) = -1/z$. The endpoints of S are $1+r$ and $1-r$, so the endpoints of $f(S)$ are $-1/(1+r)$ and $-1/(1-r)$. We guess that the center of $f(S)$ is the average of these two points:

$$c = \frac{1}{2}\Big(\frac{-1}{1+r} + \frac{-1}{1-r}\Big).$$

Half the distance between the endpoints of $f(S)$ is

$$R = \frac{1}{2}\Big(\frac{-1}{1+r} - \frac{-1}{1-r}\Big).$$

So, to finish the proof, we just have to verify that $|w - c| = R$ for all $z \in S$. What amounts to the same thing is showing that

$$(w-c)(\overline{w}-c) - R^2 = 0.$$

The left hand side expands out to

$$\frac{|z-1|^2 - r^2}{(r^2-1)|z|^2},$$

which is 0. This completes the proof. ♠

Lemma 3.5 *$f(z) = -1/z$ is a symmetry of the hyperbolic plane.*

Proof: We have already seen that f is a bijection of $\boldsymbol{H}^2$ and maps geodesics to geodesics. Therefore, if we replace p by $f(p)$ and q by $f(q)$, then we replace o by $f(o)$ and r by $f(r)$. We compute

$$|(-1/o) - (-1/p)| = \frac{|o-p|}{|o||p|},$$

and similarly for the other terms. But then

$$\frac{|f(o)-f(q)||f(p)-f(r)|}{f(o)-f(p)||f(q)-f(r)|} = \frac{|o-q||p-r|}{|o-p||q-r|} \times \frac{|o||q||p||r|}{|o||q||p||r|} = \frac{|o-q||p-r|}{|o-p||q-r|}.$$

But this means that Equation 1 does not change when we make the substitutions. ♠

Lemma 3.6 *Let $f(z) = -1/z$. Let C be any circle in the plane not containing the origin. Then $f(C)$ is another circle.*

Proof: Note that $f(aC) = (-1/a)f(C)$. So, we can rotate the picture so that C is centered on the real axis. But then C is the double of a hyperbolic geodesic. That is, $C = C_+ \cup C_-$, where C_+ is a hyperbolic geodesic and C_- is the reflection of C_+ in the real axis. But then $f(C_+)$ is another hyperbolic geodesic and by symmetry $f(C_-)$ is the reflection of $f(C_+)$. Therefore $f(C)$ is the union of two semi-circles – i.e. a circle. ♠

Lemma 3.7 *Let $f(z) = -1/z$. Then f preserves angles between geodesics in the hyperbolic plane.*

Proof: We have to consider the matrix of partial derivatives of f. Let's call this matrix *df*. The map f preserves angles between curves provided that *df* is the product of a rotation and a dilation. If this was not the case, then f would map tiny circles to curves which were very close to tiny non-circular ellipses. Since f maps circles to circles, *df* must be the product of a dilation and a rotation. ♠

Lemma 3.8 *$g(z) = -\overline{z}$ is a symmetry of the hyperbolic plane.*

Proof: We have $g(x+iy) = -x+iy$. Geometrically, g is just a reflection in the imaginary axis. The rest of the proof in this case is just like what we've done several times already. ♠

Lemma 3.9 *$h(z) = 1/\overline{z}$ is a symmetry of the hyperbolic plane.*

Proof: The map h is a composition of the map $f(z) = -1/z$ and $g(z) = -\overline{z}$, and the composition of symmetries is again a symmetry. ♠

Lemma 3.10 *Let γ be any geodesic in $\boldsymbol{H}^2$. There is a symmetry r such that $r(p) = p$ for all $p \in \gamma$ and $r \circ r$ is the identity. That is, there is a mirror reflection symmetry which fixes γ.*

Proof: Consider first the case when γ is a vertical ray. There is a symmetry of the form $f(z) = z + t$ so that $f(\gamma)$ is the vertical ray through the origin. Then $f \circ g \circ f^{-1}$ is the desired symmetry. Here $g(z) = -\overline{z}$ is the symmetry considered above.

Now consider the case when γ is a semicircle. There is a symmetry of the form $f(z) = az + b$ so that $f(\gamma)$ is the semicircle contained in the unit circle $|z| = 1$. Then $f \circ h \circ f^{-1}$ is the desired symmetry. Here $h(z) = 1/\overline{z}$ is the symmetry considered above. ♠

Lemma 3.11 *All the symmetries so far considered preserve angles between geodesics.*

Proof: The maps $f(z) = z+t$ and $f(z) = az$ and $f(z) = -\overline{z}$ are either symmetries of the Euclidean plane or dilations of the Euclidean plane. Therefore, they all preserve angles between curves. Also, we proved this explicitly for the map $f(z) = -1/z$. The other maps considered are compositions of the ones we've just talked about. ♠

Lemma 3.12 *Let p be any point in the hyperbolic plane, and let θ be any angle. Then there is a hyperbolic symmetry which fixes p and rotates by θ about p.*

Proof: We can find geodesics C_1 and C_2 which intersect at p. Let R_1 and R_2 be the reflections through C_1 and C_2. Both R_1 and R_2 fix the point p, since p lies on both C_1 and C_2. The composition $R = R_1R_2$ rotates around p by 2α, where α is the smaller of the two angles between C_1 and C_2. So. if we take $\alpha = \theta/2$, we get the desired result. ♠

Now we know that $\boldsymbol{H}^2$ is isotropic: We can find a symmetry which fixes any point and rotates by that angle around the point.

4 Satisfying the Postulates

Now let's go through Euclid's postulates for the hyperbolic plane.
First Postulate: Suppose that p and q are two points in the hyperbolic plane. If p and q lie on the same vertical ray, then this ray is is the geodesic connecting them. Otherwise, one can construct the geodesic through p and q as follows: Draw the segment $\overline{pq}$, then take the perpendicular bisector to this segment. The perpendicular bisector intersects the real axis at the center of the desired geodesic. See Figure 4.

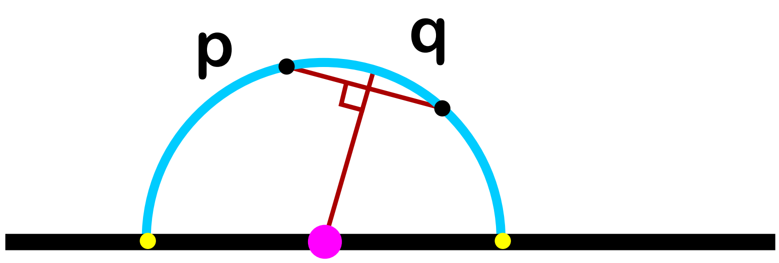

Figure 4: Constructing the geodesic containing p and q

Second Postulate: Let A be an arc of a geodesic. If A is a vertical segment, we just take the vertical ray containing A. If A is an arc of a circle, we just extend A to be the whole semicircle.

Third Postulate: The distance formula in Equation 1 shows that the real axis is infinitely far from every point in $\boldsymbol{H}^2$. So, suppose we start at a point p. Then, in each direction, we can move outward by r units along a geodesic going through p in that direction. This produces for us the circle of radius r about p. Amazingly, the circles in $\boldsymbol{H}^2$ are actually Euclidean circles as well. I'll explain this in the next section.

Fourth Postulate: This follows from the fact that $\boldsymbol{H}^2$ is homogeneous and isotropic. Suppose that A_1, A_2 are a pair of perpendicular geodesics and so are B_1, B_2. We can find a symmetry which maps $A_1 \cap A_2$ to $B_1 \cap B_2$. Call this point p. Let θ be the angle between A_1 and B_1. We can then find a symmetry which rotates about p by θ. The composition of our two symmetries fixes p and maps A_1 to B_1. But, all these maps preserve angles, so our symmetry must map A_2 to B_2 as well.

Figure 5 shows that the fifth postulate fails. There are many lines through the point p which do not intersect the line L.

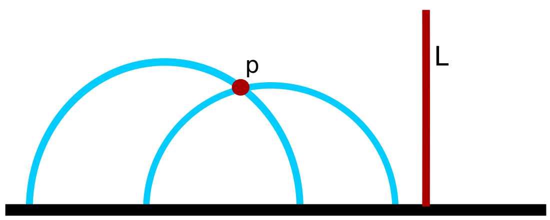

Figure 5: Parallel postulate fails.

5 The Shapes of Circles

In this section I will prove that the cirles in $\boldsymbol{H}^2$ are round circles. Let $p \in \boldsymbol{H}^2$ be some point. Let L_1 and L_2 be any geodesics through p.

Lemma 5.1 *For any $r > 0$ there is a Euclidean circle C which meets L_1 and L_2 at right angles and has a point which is hyperbolic distance r from p.*

Proof: Using symmetries, we can assume that $p = i$ and L_1 is the vertical ray through p. We consider all the circles which intersect L_1 at the point ie^r, a point on L_1 which is r away from i. We also insist that L_1 is a diameter of C, so that C intersects L_1 at right angles. We now adjust the radius of C until it intersects L_2 in a right angle at one of the intersection points. By symmetry, C intersects L_2 at right angles at both intesection points. ♠

Lemma 5.2 *C is also a hyperbolic circle.*

Proof: Let R_1 and R_2 be the reflections in L_1 and L_2. Both R_1 and R_2 map Euclidean circles to Euclidean circles and also preserve angles. Note that $R_1(C)$ is another circle which meets L_1 at the same two points, and is perpendicular to L_1 at those points. But this forces $R_1(C) = C$. Similarly $R_2(C) = C$. This means that $R_1R_2(C) = C$. Suppose that the angle between L_1 and L_2 is an irrational multiple of π. This means that $(R_1R_2)^n$, for various values of n, rotates through a dense set of possible angles. But then there is a dense set of points on C which are all the same hyperbolic distance from p. But then all points on C are the same hyperbolic distance from p. This means that C is a hyperbolic circle around p. ♠

Now we know that any point on L_1 is contained in a hyperbolic circle that is also a Euclidean circle. This accounts for all the hyperbolic circles. Therefore, every hyperbolic circle is a Euclidean circle.

6 The Triangle Inequality

Now I want to take care of one piece of unfinished business. I haven't yet shown that the hyperbolic distance satisfies the triangle inequality. That is, suppose we have 3 points p, q, r. We want to show that

$$d(p,q) + d(q,r) \geq d(p,r).$$

The proof relies on the fact that hyperbolic circles are also Euclidean circles.

We can move the picture by a symmetry so that

$$p = i, \qquad q = \alpha + ai, \qquad r = bi.$$

Let $Q = ai$ be the point on the y axis which is on the same horizontal line as q. See Figure 6. Here a and b and α are constants which depend on the distances involved. We have

$$d(p,Q) = \ln(a), \qquad d(Q,r) = \ln(b/a), \qquad d(p,r) = \ln(b).$$

In particular

$$d(p,Q) + d(Q,r) = \ln(a) + \ln(b/a) = \ln(b) = d(p,r).$$

So p, Q, r satisfy the triangle inequality. However, we want to prove this for q rather than Q.

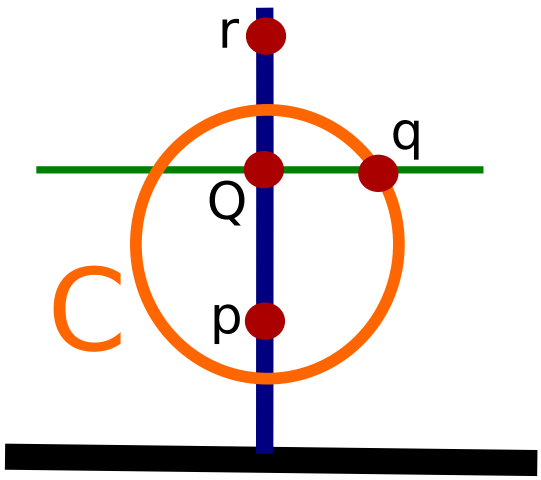

Figure 6: Triangle inequality proof.

Let C be the circle centered at p (in the hyperbolic sense) which contains q. By symmetry, the y-axis is a diameter of C. From this picture, the point Q is contained in the disk bounded by C. Therefore

$$d(p,q) \geq d(p,Q).$$

Similarly,

$$d(r,q) \geq d(r,Q).$$

But then

$$d(p,q) + d(q,r) \geq d(p,Q) + d(Q,r) = d(p,q).$$

This proves the triangle inequality.