A ROUNDABOUT INTRODUCTION TO HYPERBOLIC AREA

SARAH ZHANG

ABSTRACT. The Gauss-Bonnet theorem gives a simple formula for the area of any "reasonable" hyperbolic polygon based on its internal angle measures. To approach this result, we give an abbreviated overview of Möbius transformations, two models of hyperbolic space, convexity in the hyperbolic plane, and related formulas for hyperbolic area.

CONTENTS

1. INTRODUCTION

When it comes to two dimensional spaces, the usual go-to is the Euclidean plane, where lines are straight, shapes are flat, and the curvature of the space is uniformly zero. Another is the surface of a sphere, where the shortest paths between points are portions of great circles (potentially nonunique if the points are at polar opposite locations) and the curvature of the space is uniformly positive. Another still is the hyperbolic plane, a space with constant negative curvature often called "wrinkly" for (among other crimes) "having too much area everywhere." We contend with this one. (For a hands-on visualization, see [2] or look up some of the online tutorials for crocheting/tiling/etc. the hyperbolic plane.)

The report is organized as follows: Section 2 explores the properties of Möbius transformations on the Riemann sphere. Section 3 defines the hyperbolic metric and introduces two models of the hyperbolic plane. Section 4 explores convexity in the hyperbolic plane. Section 5 defines hyperbolic area and works towards a proof of the generalized Gauss-Bonnet theorem, which states that a reasonable hyperbolic n-gon has area $(n-2)\pi$ minus the sum of its internal angles.

2. MÖBIUS TRANSFORMATIONS

Let $\overline{\mathbb{C}} = \mathbb{C} \cup \{\infty\}$ be the one-point compactification of the complex plane, also known as the **Riemann sphere**. Circles in $\mathbb{C}$ are also circles in $\overline{\mathbb{C}}$, and the inclusion

Date: August 31, 2019.

of ∞ allows for another type of circle: the union of a straight line and ∞. This ends up still looking like a circle, since compactification wraps the complex plane into a continuous surface. Traveling an infinite distance in any direction from the origin (or any other point in $\mathbb{C}$) sends you to the point ∞.

From a more quantitative perspective, any "circle" in $\overline{\mathbb{C}}$ can be described as the solution set to the equation

$$\alpha z\overline{z} + \beta z + \overline{\beta}\overline{z} + \gamma = 0,$$

where α, $\gamma \in \mathbb{R}$ and $\beta \in \mathbb{C}$.

Remark 2.1. The solution set is a circle in $\mathbb{C}$ if and only if $\alpha \neq 0$.

We can show that certain types of functions preserve circles.

Proposition 2.2. *A homeomorphism $f : \overline{\mathbb{C}} \to \overline{\mathbb{C}}$ defined by $f(z) = az + b$ and $f(\infty) = \infty$ with $a, b \in \mathbb{C}$ and $a \neq 0$ maps circles to circles.*

Proof. The image of ∞ is already defined, so we can just consider two cases in $\mathbb{C}$: lines and circles.

Let L be a line in $\mathbb{C}$. Recalling Remark 2.1, it is the solution set of

$$\beta z + \overline{\beta}\overline{z} + \gamma = 0, \tag{2.3}$$

where $\beta \in \mathbb{C}$ and $\gamma \in \mathbb{R}$.

Suppose x is a solution to (2.3). Let y be the image of x under f. Then $x = \frac{1}{a}(y - b)$. Substituting into (2.3), we get

$$\begin{aligned} &\beta\frac{1}{a}(y-b) + \overline{\beta}\ \overline{\frac{1}{a}(y-b)} + \gamma \\ =&\frac{\beta}{a}y + \overline{\frac{\beta}{a}}\overline{y} - \frac{\beta}{a}b - \overline{\frac{\beta}{a}b} + \gamma = 0. \end{aligned} \tag{2.4}$$

Since $-\frac{\beta}{a}b$ and $-\overline{\frac{\beta}{a}b}$ are complex conjugates, their sum with γ is real. Thus, y also satisfies the equation of a line in $\mathbb{C}$, and f maps lines to lines.

Now let O be a circle in $\mathbb{C}$. It is the solution set of

$$\alpha z\overline{z} + \beta z + \overline{\beta}\overline{z} + \gamma = 0, \tag{2.5}$$

where $\beta \in \mathbb{C}$; $\alpha, \gamma \in \mathbb{R}$; and $\alpha \neq 0$.

Suppose w is a solution to (2.5). Let u be the image of w under f. Then with a similar substitution as in the linear case, we get

$$\begin{aligned} &\alpha(\frac{1}{a}(u-b))\overline{(\frac{1}{a}(u-b))} + \beta\frac{1}{a}(u-b) + \overline{\beta}\ \overline{\frac{1}{a}(u-b)} + \gamma \\ =\ &\frac{\alpha}{|a|^2}(u-b)(\overline{u}-\overline{b}) + \frac{\beta}{a}u + \overline{\frac{\beta}{a}}\overline{u} + \text{(real number)} \\ =\ &\frac{\alpha}{|a|^2}(u\overline{u} - b\overline{u} - \overline{b}u) + \frac{\beta}{a}u + \overline{\frac{\beta}{a}}\overline{u} + \text{(real number)} \\ =\ &\frac{\alpha}{|a|^2}u\overline{u} + (\frac{\beta}{a} - \overline{\frac{\alpha}{|a|^2}}\overline{b})u + \overline{(\frac{\beta}{a} - \overline{\frac{\alpha}{|a|^2}}\overline{b})}\overline{u} + \text{(real number)} = 0. \end{aligned} \tag{2.6}$$

We see that $\frac{\alpha}{|a|^2}$ is real and the coefficients of u and $\overline{u}$ are complex conjugates, so u also satisfies the equation of a circle in $\mathbb{C}$. Thus, f maps circles to circles and, in general, maps "circles" to "circles." □

Proposition 2.7. *The homeomorphism $J : \overline{\mathbb{C}} \to \overline{\mathbb{C}}$ defined by $J(z) = \frac{1}{z}$ for $\overline{\mathbb{C}} \setminus \{0, \infty\}$, $f(\infty) = 0$, and $f(0) = \infty$ maps circles to circles.*

Proof. Let w be a solution to $\alpha z\overline{z} + \beta z + \overline{\beta}\overline{z} + \gamma = 0$, where $\beta \in \mathbb{C}$; $\alpha, \gamma \in \mathbb{R}$; and $\alpha \neq 0$.

Let $u = J(w) = \frac{1}{w}$, so that $w = \frac{1}{u}$. Then

$$\alpha \frac{1}{u}\overline{\frac{1}{u}} + \beta\frac{1}{u} + \overline{\beta}\overline{\frac{1}{u}} + \gamma = 0. \tag{2.8}$$

Multiplying both sides of (2.8) by $u\overline{u}$ gives

$$\alpha + \beta\overline{u} + \overline{\beta}u + \gamma u\overline{u}, \tag{2.9}$$

which is the equation of another circle. Thus, J maps circles to circles. □

Compositions of the functions seen in the previous two propositions define a larger set of homeomorphisms.

Definition 2.10. A **Möbius transformation** is a function $m : \overline{\mathbb{C}} \to \overline{\mathbb{C}}$ of the form

$$m(z) = \frac{az + b}{cz + d},$$

where $\{a, b, c, d\} \subset \mathbb{C}$ and $ad - bc \neq 0$.

Remark 2.11. The explicit inverse of a Möbius transformation with the above form is given by

$$m^{-1}(z) = \frac{dz - b}{-cz + a}.$$

Proposition 2.12. *Möbius transformations map circles to circles.*

Proof. By definition, Möbius transformations are compositions of functions that map circles to circles; thus, they also map circles to circles. □

The existence of inverses (you can check that the formula in Remark 2.11 works if you want) and the fact that compositions of circle-preserving transformations will also preserve circles demonstrate that the set of all Möbius transformations, denoted $\mathbf{M\ddot{o}b}^+$, is a group under composition.

Remark 2.13. For any $a \neq 0$, we let $\frac{a}{0} = \infty$ in $\overline{\mathbb{C}}$ as a result of taking the limit

$$\lim_{z \to 0} \frac{a}{z}.$$

For any $m(z) = \frac{az+b}{cz+d}$, we calculate $m(\infty)$ by taking the limit

$$\lim_{z \to \infty} \frac{az + b}{cz + d} = \lim_{z \to \infty} \frac{a + \frac{b}{z}}{c + \frac{d}{z}} = \frac{a}{c}.$$

From this, we see that $m(\infty) = \infty$ if and only if $c = 0$.

Definition 2.14. A **fixed point** of a nontrivial (non-identity) Möbius transformation m is a point $z \in \overline{\mathbb{C}}$ satisfying $m(z) = z$.

Theorem 2.15. *The only element of Möb$^+$ with more than two fixed points is the identity element $e(z) = z$.*

Proof. Let $m(z) = \frac{az+b}{cz+d}$ not be the identity transformation. We consider the cases $c = 0$ and $c \neq 0$.

If $c = 0$, then $m(z) = \frac{a}{d}z + \frac{b}{d}$. From Remark 2.12, we know that ∞ is a fixed point of m. How many fixed points in $\mathbb{C}$ are there? Since $m \neq e$, we can't have both $\frac{a}{d} = 1$ and $b = 0$. But if $\frac{a}{d} = 1$ and $b \neq 0$, then there are no solutions. Otherwise, if $\frac{a}{d} \neq 1$, then $d - a \neq 0$ and we can rearrange $z = \frac{a}{d}z + \frac{b}{d}$ to get

$$z = \frac{b}{d-a}.$$

If $c \neq 0$, then ∞ isn't a fixed point of m. Any fixed points will be complex numbers satisfying $z = \frac{az+b}{cz+d}$, which can be rearranged into

$$cz^2 + (d-a)z - b = 0.$$

Solving gives $z = \frac{a-d \pm \sqrt{(d-a)^2+4bc}}{2c}$, which produces one or two solutions.

Thus, any nonidentity Möbius transformation has either one or two fixed points. $\square$

Definition 2.16. Let G be a group and let X be a set. G **acts on** X if there exists a homomorphism from G to the group of bijections of X.

G acts **transitively** on X if for all $x, y \in X$, there exists $g \in G$ such that $g(x) = y$.

G acts **uniquely transitively** on X if for all $x, y \in X$, there exists exactly one $g \in G$ such that $g(x) = y$.

Lemma 2.17. *Let a be an element of a set X. If G acts on X and for every $x \in X$, there exists $g \in G$ such that $g(x) = a$, then G acts transitively on X.*

Proof. Fix $x, y \in X$. By assumption, there exist $g_x, g_y \in G$ such that $g_x(x) = a$ and $g_y(y) = a$. Then $(g_y^{-1} \circ g_x)(x) = y$. Since $g_y^{-1} \circ g_x \in G$, G acts transitively on X. $\square$

Theorem 2.18. *Given two triples of distinct points $(x_1, x_2, x_3), (y_1, y_2, y_3) \in \overline{\mathbb{C}}^3$, there exists exactly one Möbius transformation m such that*

$$m(x_1) = y_1 \qquad m(x_2) = y_2 \qquad m(x_3) = y_3$$

Proof. Suppose $x_1, x_2, x_3 \in \mathbb{C}$. Consider the Möbius transformation

$$t(z) = \frac{(x_2 - x_3)z - (x_2 - x_3)x_1}{(x_2 - x_1)z - (x_2 - x_1)x_3}. \tag{2.19}$$

Then $t(x_1) = 0$, $t(x_2) = 1$, and $t(x_3) = \infty$.

Suppose not all x_i are strictly complex numbers. Without loss of generality, let $x_1 = \infty$ and $x_2, x_3 \in \mathbb{C}$. Consider the transformation

$$t(z) = \frac{x_3 - x_2}{-z + x_3}. \tag{2.20}$$

Then $t(\infty) = 0$, $t(x_2) = 1$, and $t(x_3) = \infty$.

There similarly exists a Möbius transformation s that maps

$$s(y_1) = 0 \qquad s(y_2) = 1 \qquad s(y_3) = \infty$$

which has an inverse that maps

$$s^{-1}(0) = y_1 \qquad s^{-1}(1) = y_2 \qquad s^{-1}(\infty) = y_3$$

Then the composition $m = t \circ s^{-1}$ is a Möbius transformation such that $m(x_1) = y_1, m(x_2) = y_2$, and $m(x_3) = y_3$.

Suppose m is not unique. Let n be a Möbius transformation such that $n(x_1) = y_1, n(x_2) = y_2$, and $n(x_3) = y_3$. Then $m^{-1} \circ n$ has three fixed points:

$$(m^{-1} \circ n)(0) = 0 \qquad (m^{-1} \circ n)(1) = 1 \qquad (m^{-1} \circ n)(\infty) = \infty$$

By Theorem 2.15, $m^{-1} \circ n$ must be the identity transformation, implying that $m = n$. Thus, m is unique and Möb$^+$ acts uniquely transitively on the set of distinct triples of points in $\overline{\mathbb{C}}$ (or **uniquely triply transitively** on $\overline{\mathbb{C}}$). □

Proposition 2.21. *Möb$^+$ acts transitively on the set of circles in $\overline{\mathbb{C}}$ and on the set of discs in $\overline{\mathbb{C}}$.*

Proof. The proof for circles results from taking two circles A and B in $\overline{\mathbb{C}}$, and applying Theorem 2.18 to find $m \in$ Möb$^+$ mapping three distinct points in A to three distinct points in B. Since Möbius transformations map circles to circles and three distinct points in $\overline{\mathbb{C}}$ uniquely determine a circle (or line), $m(A)$ must equal B.

The proof for discs is similar. Circles in $\overline{\mathbb{C}}$ determine two discs, however, so given discs A and B and the transformation m that maps the circle C_A determining A to the circle C_B determining B, $m(A)$ might produce B or the disc opposite to B. If the latter is true, we can construct another Möbius transformation $n^{-1} \circ J \circ n$, with $J(z) = \frac{1}{z}$ (which maps $\overline{\mathbb{R}}$ to itself but interchanges the two discs determined by $\overline{\mathbb{R}}$) and n mapping C_B to $\overline{\mathbb{R}}$. Then $(n^{-1} \circ J \circ n \circ m)(A) = B$. □

Möbius transformations can be further classified into three types based on the number of fixed points they have and other properties.

Definition 2.22. Two Möbius transformations m and n are **conjugate** if there exists a Möbius transformation p such that

$$m = p \circ n \circ p^{-1}.$$

This introduces a concept of similarity between transformations.

Lemma 2.23. *Conjugate maps have the same number of fixed points.*

Proof. Let $m, n \in$ Möb$^+$ be conjugate by p so that $m = p \circ n \circ p^{-1}$.

Suppose x is a fixed point of m. Then $(p \circ n \circ p^{-1})(x) = x$, which is true if and only if $(n \circ p^{-1})(x) = p^{-1}(x)$ (ie, if and only if $p^{-1}(x)$ is a fixed point of n). Thus, there exists a bijection between the fixed points of m and the fixed points of n, implying that the fixed point sets have the same cardinality. □

Definition 2.24. A Möbius transformation is **parabolic** if it has exactly one fixed point.

Remark 2.25. Any parabolic transformation m is conjugate to $n(z) = z + 1$, called the **standard form** of m. This can be found computationally starting from the assumption that m has one fixed point.

Now consider a Möbius transformation m with exactly two fixed points, x and y. Suppose q is a transformation such that $q(x) = 0$ and $q(y) = \infty$. Then $(q \circ m \circ q^{-1})(0) = 0$ and $(q \circ m \circ q^{-1})(\infty) = \infty$. The standard form of m is $(q \circ m \circ q^{-1})(z) = n(z) = az$ for some $a \in \overline{\mathbb{C}} \setminus \{0, 1\}$.

Remark 2.26. If p and q are transformations satisfying $p(x) = q(x) = 0$ and $p(y) = q(y) = \infty$, the multipliers a and b for $p \circ m \circ p^{-1} = az$ and $q \circ m \circ q^{-1} = bz$ are the same. The proof of this statement is left as an exercise for the reader.

Definition 2.27. A Möbius transformation is **elliptic** if it has a standard form with multiplier $|a| = 1$. We can write $a = e^{2i\theta}$ for some $\theta \in (0, \pi)$. The standardized transformation represents a rotation by 2θ about the origin.

Definition 2.28. A Möbius transformation is **loxodromic** if it has a standard form with multiplier $|a| \neq 1$. We can write $a = r^2 e^{2i\theta}$ for some $r \in \mathbb{R}^+ \setminus \{1\}$ and $\theta \in [0, \pi)$. The standardized transformation represents a composition of a dilation by r^2 and a rotation by 2θ about the origin.

Moving past pure Möbius transformations, we encounter a larger group of homeomorphisms that preserve circles in $\overline{\mathbb{C}}$. In fact, it *is* the group of circle-preserving homeomorphisms in $\overline{\mathbb{C}}$, a result that can be reached through arduous means not reproduced here [1].

Definition 2.29. The **general Möbius group**, denoted **Möb**, is the group generated by Möb^+ and $C(z) = \overline{z}$. Every nontrivial element of Möb can be expressed as a composition

$$C \circ m_1 \circ C \circ m_2 \circ ... \circ C \circ m_k,$$

where $k \geq 1$ and $m_i \in \text{Möb}^+$ for all $i \in \{1, ..., k\}$.

Möb inherits transitivity on distinct triples, circles, and discs in $\overline{\mathbb{C}}$ from Möb^+, but not unique transitivity on triples.

3. The upper half plane $\mathbb{H}$ and the Poincaré disc $\mathbb{D}$

Definition 3.1. The **upper half plane** $\mathbb{H} = \{z \in \mathbb{C} \mid \text{Im}(z) > 0\}$ is the projection of $\mathbb{H}^2$ onto the portion of the complex plane containing all points with a positive imaginary component. Geodesics in $\mathbb{H}$ are either vertical lines intersecting the real axis or semicircles intersecting the real axis at right angles. Said real axis is actually the one point compactification of the real axis, $\overline{\mathbb{R}} = \mathbb{R} \cup \{\infty\}$, which forms the **boundary at infinity** of $\mathbb{H}$.

Definition 3.2. The measure of an angle between two hyperbolic curves is equal to the measure of the Euclidean angle between the straight lines tangent to the curves at their intersection.

Theorem 3.3. *The subgroup* $\text{Möb}(\mathbb{H}) = \{m \in \textit{Möb} \mid m(\mathbb{H}) = \mathbb{H}\}$ *sends hyperbolic lines to hyperbolic lines in* $\mathbb{H}$.

Proof. We know that all elements of Möb send circles to circles and that all hyperbolic lines in $\mathbb{H}$ are intersections of $\mathbb{H}$ with circles in $\overline{\mathbb{C}}$ (including straight lines) that meet $\overline{\mathbb{R}}$ at right angles. It can be shown that Möb($\mathbb{H}$) preserves angles between circles in $\overline{\mathbb{C}}$ [1], which completes the proof. □

Theorem 3.4. Möb($\mathbb{H}$) *acts transitively on* $\mathbb{H}$.

Proof. Let $x \in \mathbb{H}$, and write $x = a + bi$, where $a, b \in \mathbb{R}$ and $b > 0$. Consider the transformation $m(z) = \frac{1}{b}(z - a)$, which is equal to the composition of $p(z) = z - a$ and $q(z) = \frac{1}{b}z$, both of which are elements of Möb($\mathbb{H}$). Observe that m sends every $x \in \mathbb{H}$ to $i \in \mathbb{H}$. Then by Lemma 2.17, Möb($\mathbb{H}$) acts transitively on $\mathbb{H}$. □

Proposition 3.5. Möb($\mathbb{H}$) *acts transitively on the set of hyperbolic lines in* $\mathbb{H}$.

Proof. The proof of this proposition results from constructing a transformation that maps an arbitrary hyperbolic line to the positive imaginary axis and applying Lemma 2.17. □

Many metrics exist for the Euclidean plane, from the standard Euclidean metric to the so-called taxicab metric to the nearly meaningless discrete metric. Although the discrete metric can technically also be applied to the hyperbolic plane, it's basically useless beyond determining whether two points are in the exact same location. Thus we move towards a more informative **hyperbolic metric**.

Definition 3.6. Let $P : [a, b] \to \mathbb{H}$ parameterize a piecewise smooth (C^1) path in the upper half plane $\mathbb{H}$. The hyperbolic length of P is

$$\text{length}_{\mathbb{H}}(P) = \int_a^b \frac{1}{\text{Im}(P(t))} \; |P'(t)| \; dt = \int_a^b \frac{1}{\text{Im}(z)} \; |dz|.$$

Proposition 3.7. *Lengths are invariant under Möb*($\mathbb{H}$)*; that is, for any* $P : [a, b] \to \mathbb{H}$ *and any* $m \in$ *Möb*($\mathbb{H}$),

$$\textit{length}_{\mathbb{H}}(P) = \textit{length}_{\mathbb{H}}(m \circ P).$$

Proof. Computational. Expanding each length integral according to Definition 3.6 eventually results in equality. □

Definition 3.8. Let $x, y \in \mathbb{H}$, and let $\Gamma(x, y)$ be the set of all piecewise C^1 paths $P : [a, b] \to \mathbb{H}$ such that $P(a) = x$ and $P(b) = y$. Then the **hyperbolic distance** between x and y is

$$\text{d}_{\mathbb{H}}(x, y) = \inf\{\text{length}_{\mathbb{H}}(P) \mid P \in \Gamma(x, y)\}.$$

One way to explicitly calculate the distance between points x and y in $\mathbb{H}$ is as follows:

Find a transformation $m \in$ Möb($\mathbb{H}$) such that $m(x) = ai$ and $m(y) = bi$ for some $a, b > 0$. Then

$$\text{d}_{\mathbb{H}}(x, y) = \text{d}_{\mathbb{H}}(ai, bi) = \left| \ln \frac{b}{a} \right|.$$

Another way is to express x and y in terms of their real and imaginary components: $x = t + ui$ and $y = v + wi$ with $u, w > 0$. If $t = v$, the distance is just $\left| \ln \frac{w}{u} \right|$ as in the previous method. Otherwise,

$$\text{d}_{\mathbb{H}}(x, y) = \left| \ln \left| \frac{(t - c - r)w}{(v - c - r)v} \right| \right|,$$

where c is the Euclidean center and r is the Euclidean radius of the Euclidean circle containing the hyperbolic line connecting x and y.

See Chapter 3 of [1] for more details about the derivation of the hyperbolic metric. We will take it for granted that $(\mathbb{H}, \mathrm{d}_{\mathbb{H}})$ is a metric space and that the distance formulas work.

Another model of the hyperbolic plane contained in $\mathbb{C}$ is on the smaller side.

Definition 3.9. The **Poincaré disc** $\mathbb{D} = \{z \in \mathbb{C} \mid |z| < 1\}$, also known as the conformal disc model, is a projection of $\mathbb{H}^2$ onto the unit circle.[1] Geodesics in $\mathbb{D}$ follow circular paths intersecting the boundary of the disc at right angles. They are only straight when they cross the center of $\mathbb{D}$.

The Möbius transformation

$$m(z) = \frac{z - i}{z + i}$$

takes points from $\mathbb{H}$ to $\mathbb{D}$; the inverse function that maps $\mathbb{D}$ to $\mathbb{H}$ is

$$m^{-1}(z) = \frac{-iz - i}{z - 1}.$$

The hyperbolic metric is invariant under these transformations, so all geodesics in $\mathbb{H}$ are geodesics in $\mathbb{D}$ and vice versa; the boundary at infinity of both $\mathbb{D}$ and $\mathbb{H}$ is $\overline{\mathbb{R}}$.

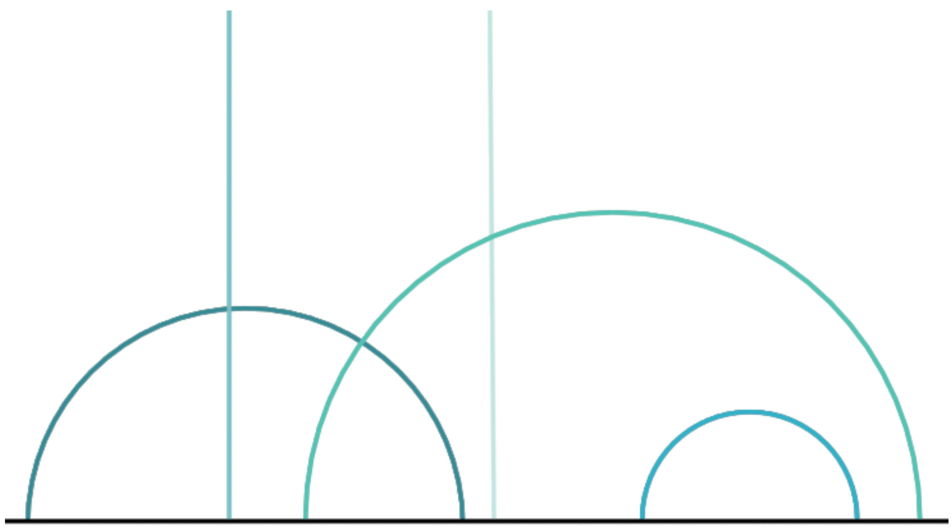

FIGURE 1. The upper half plane and some geodesics. The vertical geodesics pictured both go to ∞. The horizontal black line in this figure and the black circle in Figure 2 represent $\overline{\mathbb{R}}$.

[1] A slightly less popular disc model is the Klein disc [2], a.k.a. the projective model, where geodesics are all straight lines. However, angles and circles are distorted by this projection, so some of its uses are limited.

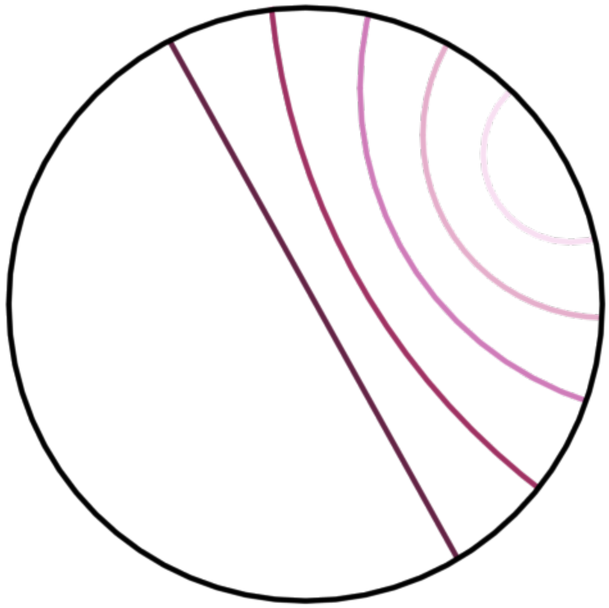

FIGURE 2. The Poincaré disc and some geodesics.

4. Convexity

Convexity in the hyperbolic plane is analogous to Euclidean convexity in that convex sets contain all minimal paths between their elements. The only difference is the nature of geodesics in $\mathbb{H}^2$. We will encounter convexity again in the next section when we define hyperbolic polygons.

Definition 4.1. A subset X of the hyperbolic plane is **convex** if for all $x, y \in X$, the closed hyperbolic line segment ℓ_{xy} connecting them is contained in X.

Remark 4.2. Hyperbolic lines, line segments, and rays are all convex since they must contain every closed line segment in themselves.

Proposition 4.3. *The intersection of an arbitrary collection of convex subsets of the hyperbolic plane is convex.*

Proof. Let $\{X_a\}_{a\in I}$ be a collection of convex subsets of the hyperbolic plane indexed by I. Fix x and y in the intersection $X = \bigcap_{a\in I} X_a$. Then x and y are contained in each X_a, so by convexity, ℓ_{xy} is contained in every X_a. Then $\ell_{xy} \subset \bigcap_{a\in I} X_a$. Thus, X is convex. □

Any hyperbolic line ℓ divides the hyperbolic plane into two components. These components, when singled out, form some of the most elementary convex sets in $\mathbb{H}^2$.

Definition 4.4. Each component in the complement of a hyperbolic line ℓ is an **open half plane**, and the line that determines a given half plane is its **bounding**

line. The union of an open half plane with its bounding line is a **closed half plane**.

Proposition 4.5. Möb($\mathbb{H}$) *acts transitively on the set of half planes in* $\mathbb{H}$.

Proof. The proof of this proposition results from constructing a transformation that maps an arbitrary half plane to $\mathbb{H}$ and applying Lemma 2.17. □

Proposition 4.6. *Half planes are convex.*

Proof. This statement can be proved by demonstrating convexity for a specific open half plane, noting that Möb($\mathbb{H}$) acts transitively on half planes and preserves convexity, implying that every open half plane is convex. The proof for closed half planes is similar. □

Definition 4.7. The **convex hull** of a set $X \subset \mathbb{H}^2$, written conv(X), is the intersection of all convex sets in the hyperbolic plane containing X.

Theorem 4.8. *A closed set* $X \subset \mathbb{H}^2$ *is convex if and only if it can be expressed as the intersection of a collection of half planes.*

Proof. By Proposition 4.6, a single half plane is convex, so by Proposition 4.3, the intersection of an arbitrary collection of half planes will be convex.

A sketch of the converse proof [1]: Let X be a closed convex subset of the hyperbolic plane. It can be shown that because X is convex, for any point $z \notin X$ in the upper half plane, there exists exactly one $x \in X$ such that $\mathrm{d}_{\mathbb{H}}(z, x)$ is the infimum of all distances between z and elements of X.

The hyperbolic line passing through x perpendicular to the hyperbolic line connecting z and x is the bounding line of an open half plane O_z containing z as well as this half plane's complement, a closed half plane C_z not containing z. It can be shown by contradiction that C_z must contain all of X, or else there will exist a point $y \in X$ such that $\mathrm{d}_{\mathbb{H}}(z, x) > \mathrm{d}_{\mathbb{H}}(z, y)$. Then X can be expressed as the intersection of all possible C_z determined by points not in X, or

$$X = \bigcap_{\substack{z \in \mathbb{H} \\ z \notin X}} C_z.$$

□

5. Hyperbolic polygons and the Gauss-Bonnet formula

The definition of a polygon in $\mathbb{H}^2$ depends first and foremost on a concept not needed for polygons in Euclidean geometry, to prevent the existence of polygons excessively dense with sides.

Definition 5.1. A collection $\mathcal{H} = \{H_a\}_{a \in I}$ of half planes in the hyperbolic plane is **locally finite** if for any $z \in \mathbb{H}^2$, there exists $\epsilon > 0$ such that the open hyperbolic disc $D_\epsilon(z)$ centered at z with radius ϵ has a nonempty intersection with only finitely many bounding lines corresponding to half planes in $\mathcal{H}$.

Remark 5.2. Finite collections of half planes have at most $n \in \mathbb{N}$ bounding lines associated with n half planes, so any open disc $D_\epsilon(z)$ intersects at most n bounding lines. Some countably infinite collections are locally finite, but no uncountably infinite collection is locally finite (try proving this as an exercise if you like).

Definition 5.3. A **hyperbolic polygon** is the intersection of a locally finite collection of closed half planes in the hyperbolic plane.

Since a hyperbolic polygon takes the intersection of closed, convex sets, by Proposition 4.3, it is also closed and convex. Note that a given polygon has many possible expressions.

As previously mentioned, this basic definition prevents the existence of hyperbolic polygons that are "too" full of edges, but it still allows for innumerable unintuitive polygons. Here are two on $\mathbb{D}$ (pardon the slightly incorrect geodesics):

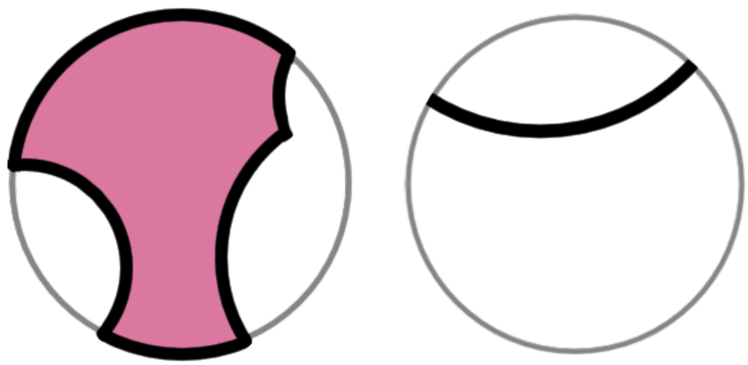

FIGURE 3. Left: a hyperbolic 3-gon with infinite area. Right: a hyperbolic 2-gon (the intersection of two closed half planes with the same bounding line) with no area.

Definition 5.4. A hyperbolic polygon is **degenerate** if it has an empty interior.

Proposition 5.5. *All degenerate polygons are a hyperbolic line, closed hyperbolic ray, closed hyperbolic line segment, or single point.*

Proof. This can be proved by contraposition. Suppose a hyperbolic polygon P contains three points not on the same line; it can be shown that these three points determine a region with positive area, implying that P has a nonempty interior by convexity. □

Definition 5.6. Let P be a hyperbolic polygon contained in a half plane determined by a hyperbolic line ℓ. If $P \cap \ell$ is a point, that point is called a **vertex** of P. Otherwise, if $P \cap \ell$ is nonempty, then $P \cap \ell$ is called a **side** of P.

Definition 5.7. A vertex v of a hyperbolic polygon P is **ideal** if two sides of P approach v at infinity.

For the remainder of this section, we only consider nondegenerate polygons.

Definition 5.8. A hyperbolic polygon is **reasonable** if it does not contain an open half plane.

Definition 5.9. A hyperbolic polygon is **compact** if it is reasonable and has no ideal vertices. This means only closed line segment sides are permitted, and there will be an equal number of vertices and sides.

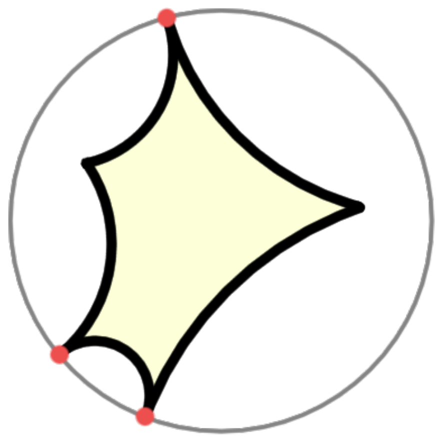

FIGURE 4. A reasonable, noncompact hyperbolic 5-gon with three ideal vertices drawn in $\mathbb{D}$...

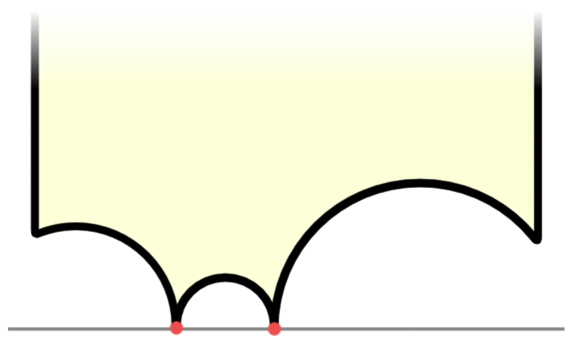

FIGURE 5. ...and in $\mathbb{H}$.

Proposition 5.10. *A compact hyperbolic polygon P is equal to the convex hull of its vertices V.*

Proof. P is convex and contains its vertices, so $\mathrm{conv}(V) \subset P$. On the other hand, $\mathrm{conv}(V)$ contains all the line segments joining pairs of vertices in V, so $\mathrm{conv}(V)$ contains the boundary of P. Fix x in the interior of P. Any hyperbolic line passing through x must intersect ∂P twice. These intersection points determine a closed line segment containing x, and by convexity must be contained in $\mathrm{conv}(V)$. Thus, x is contained in the convex hull of V and $\mathrm{int}(P) \cup \partial P = P \subset \mathrm{conv}(V)$. □

Definition 5.11. A hyperbolic polygon is an **ideal *n*-gon** if it has n sides, n ideal vertices, and is reasonable.

Definition 5.12. The **interior angle** of a hyperbolic polygon at a vertex is the angle between the two sides of the polygon that meet at that vertex.

Remark 5.13. An ideal vertex has an interior angle measure of 0.

Definition 5.14. A hyperbolic polygon is **regular** if it is compact and has equal side lengths and interior angles.

We now turn to the more numerical properties of hyperbolic polygons.

Squaring the arc length integrand from Definition 3.6 gives the integrand for hyperbolic area.

Definition 5.15. Let X be a subset of $\mathbb{H}$ and let $z = x + iy$. The hyperbolic area of X is

$$\int_X \frac{1}{\operatorname{Im}(z)^2}\, dx\, dy = \int_X \frac{1}{y^2}\, dx\, dy.$$

Remark 5.16. Hyperbolic area is invariant under Möb($\mathbb{H}$).

Again, we will take the preceding statements for granted; see [1] for superior rigor and transparency.

Lemma 5.17. *A hyperbolic triangle with one ideal vertex and interior angles α and β associated with its other two vertices has area $\pi - (\alpha + \beta)$.*

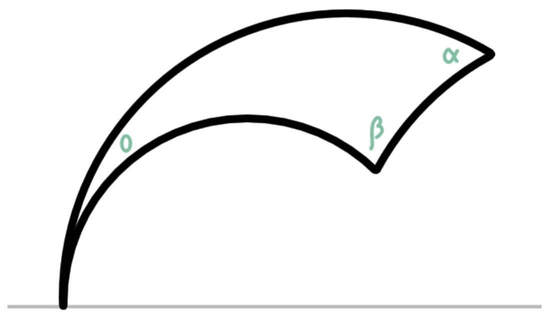

FIGURE 6

Proof. By transitivity of Möb($\mathbb{H}$), it suffices to show that this formula holds for the hyperbolic triangle with an ideal vertex at ∞ and two vertices at coordinates $e^{i\theta}$ and $e^{i\phi}$, where $0 \leq \theta < \phi \leq \pi$.

A direct calculation for the area of this triangle (including the substitution $x = \cos u$) yields the following:

$$\begin{aligned} \int_\Delta \frac{1}{y^2}\, \mathrm{d}x\, \mathrm{d}y &= \int_{\cos(\phi)}^{\cos(\theta)} \int_{\sqrt{1-x^2}}^{\infty} \frac{1}{y^2}\, \mathrm{d}y\, \mathrm{d}x \\ &= \int_{\cos(\phi)}^{\cos(\theta)} \frac{1}{\sqrt{1-x^2}}\, \mathrm{d}x \\ &= \int_{\phi}^{\theta} -\mathrm{d}u \\ &= \phi - \theta. \end{aligned} \tag{5.18}$$

Since the interior angle measure of any ideal vertex is 0, the interior angle of the vertex at $e^{i\theta}$ is $\theta = \alpha$, and the interior angle of the vertex at $e^{i\phi}$ is $\pi - \phi = \beta$, the area of the triangle can be expressed as $\pi - (\alpha + \beta)$. □

Corollary 5.19. *An ideal hyperbolic triangle has area* π*.*

Proof. This follows directly from setting $\alpha = \beta = 0$ in Lemma 5.17. □

Theorem 5.20. *A hyperbolic triangle with interior angles* α*,* β*, and* γ *has area* $\pi - (\alpha + \beta + \gamma)$*.*

Proof. To create a triangle like in Lemma 5.17, we extend one side of a given triangle (represented by the shaded area in Figure 7) into the real axis and draw a hyperbolic ray from the remaining vertex to where the extension meets the real axis. This new triangle has interior angles $0, \alpha$, and something slightly larger than β—we denote the difference by δ. Its area includes the triangle of interest as well as another triangle with one ideal vertex, this time with interior angles $0, \delta$, and $\pi - \gamma$.

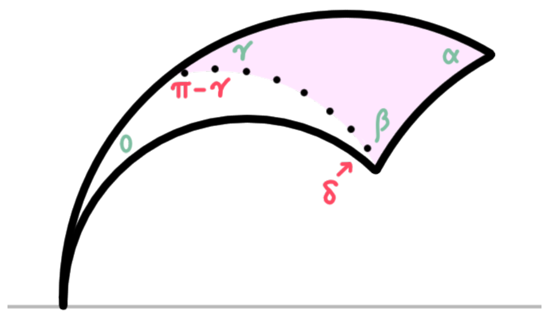

Figure 7

Then by Lemma 5.17, the largest triangle has area

$$\pi - (\alpha + (\beta + \delta))$$

and the unshaded triangle has area

$$\pi - (\delta + (\pi - \gamma)).$$

Therefore, the area of the shaded triangle must be their difference:

$$(\pi - \alpha - \beta - \delta) - (\gamma - \delta) = \pi - (\alpha + \beta + \gamma).$$

□

Theorem 5.20 is known as the Gauss-Bonnet theorem, and from repeated applications, we can obtain the general formula for the area of a reasonable hyperbolic polygon. We simply break down a given polygon into hyperbolic triangles and sum their areas.

Theorem 5.21. *A reasonable hyperbolic polygon with interior angles* α_1*,* α_2*, ...,* α_n *has area*

$$(n-2)\pi - (\alpha_1 + \alpha_2 + ... + \alpha_n) = (n-2)\pi - (\sum_{k=1}^{n} \alpha_k).$$

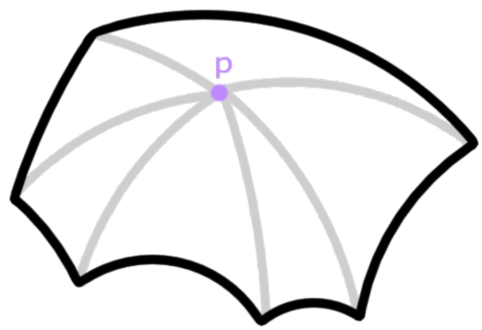

FIGURE 8. A hyperbolic 6-gon split into triangles.

Proof. Without loss of generality, suppose α_i and $\alpha_{(i+1)(\text{mod } n)}$ are adjacent interior angles for each $i \in \{1, ..., n\}$.

For all $i \in \{1, ..., n\}$, let v_i be the vertex associated with interior angle α_i.

Take any point p in the interior of P. There is a unique hyperbolic line segment joining p to each v_k, which we denote ℓ_i. Because P is convex, each ℓ_i is contained in P; because every v_i and $v_{(i+1)(\text{mod } n)}$ are adjacent vertices, a hyperbolic triangle is determined by $\ell_{v_i}, \ell_{v_{(i+1)(\text{mod } n)}}$, and the side of P connecting the two vertices. Denote this triangle T_i, and denote its interior angle at p by t_i.

Observe that this splitting creates n triangles whose areas sum to the area of P. Furthermore, the sum of all the triangles' interior angles not at p is $\sum_{k=1}^{n} \alpha_k$, while the sum of all the triangles' interior angles at p is

$$\sum_{k=1}^{n} t_k = 2\pi.$$

We know from Theorem 5.20 that the area of any hyperbolic triangle is π minus the sum of its interior angles. Thus, the total area of these n triangles must be $n\pi$ minus all the interior angles of every T_i, or

$$n\pi - (\sum_{k=1}^{n} t_k + \sum_{k=1}^{n} \alpha_k) = n\pi - 2\pi - \sum_{k=1}^{n} \alpha_k = (n-2)\pi - \sum_{k=1}^{n} \alpha_k. \tag{5.22}$$

□

And that's that!

Acknowledgments

Numerous and profuse thanks to my mentor Oliver Wang, for providing invaluable guidance and possessing incredible patience throughout this program; to Dr. Peter May, for organizing this REU again this summer; to Stephanie Oh, for agreeing to embark on this adventure and keeping me from wallowing in unproductivity; to Bonnie Hu, for being an excellent roommate and centipede extermination assistant; and to my parents, who had enough faith in my survival skills to release their only youngest child into the Chicagoan mathematical wilderness for two and a half months. I wrote a paper for you.

References

[1] James W. Anderson. Hyperbolic Geometry. Springer-Verlag London. 2005.
[2] Kathryn Mann. DIY hyperbolic geometry. https://math.berkeley.edu/ kpmann/DIYhyp.pdf.
[3] Charles Walkden. Hyperbolic Geometry. https://personalpages.manchester.ac.uk/staff/charles.walkden/hyperbolic-geometry/hyperbolic_geometry.pdf.