Introduction to hyperbolic geometry

Nilava Metya

24 July 2021

1 The upper half plane and hyperbolic geometry

1.1 Introduction

In Euclidean geometry, the parallel postulate (or Playfair's axiom) is given by: *In a plane, given a line and a point not on it, there is at most one line passing through that point parallel to it.* Euclid questioned if the inventory for Euclidean lines (having five statements) could be reduced further. In particular, he asked if the parallel line postulate could be proved using the other four.
Hyperbolic geometry is a non-Euclidean geometry in which this axiom does not hold (but the other four do). The upper-half plane is the set $\mathbb{H} = \{z \in \mathbb{C} \mid \Im z > 0\}$. We will see how 'parallel lines' don't have the same behaviour in Hyperbolic geometry as in Euclidean geometry.

1.2 Riemann sphere

We denote the unit sphere in $\mathbb{R}^3$ by $\mathbb{S}^2 \coloneqq \left\{\boldsymbol{p} \in \mathbb{R}^3 : \|\boldsymbol{p}\|_2 = 1\right\}$. We note a couple of special points, namely, the north pole $\mathfrak{n}(0,0,1)$ and the south pole $\mathfrak{s}(0,0,-1)$.
Consider the map $\sigma : \mathbb{S}^2 \smallsetminus \{\mathfrak{n}\} \to \mathbb{C}$ given by $\sigma(x,y,z) = \dfrac{x+iy}{1-z}$. This is a stereographic projection from $\mathfrak{n}$. One must note that σ is a homeomorphism (where $\mathbb{S}^2$ is given the subspace topology of $\mathbb{R}^3$).

As we obtain $\mathbb{C}$ from $\mathbb{S}^2$ by removing a single point of $\mathbb{S}^2$, namely $\mathfrak{n}$, we can construct $\mathbb{S}^2$ by starting with $\mathbb{C}$ and adding a single point denoted by ∞. The *Riemann sphere* is defined as the union $\hat{\mathbb{C}} = \mathbb{C} \cup \{\infty\}$ with a complex manifold structure given on it, as follows. We can write $\mathbb{S}^2 = U \bigcup V$ where $U = \mathbb{S}^2 \smallsetminus \{\mathfrak{n}\}, V = \mathbb{S}^2 \smallsetminus \{\mathfrak{s}\}$. Note that these are open in $\mathbb{S}^2$. Define $\tau : V \to \mathbb{C}$ by $\tau(z) = \begin{cases} 0 & \text{if } z = \mathfrak{n} \\ \frac{1}{\sigma(z)} & \text{otherwise} \end{cases}$. We must note that the transition map $\tau\sigma^{-1}$ is simply $z \mapsto \frac{1}{z}$, which is biholomorphic.
The above construction of the Riemann sphere $\hat{\mathbb{C}}$ from the complex plane $\mathbb{C}$ is an example of one-point compactification.

We want to define open discts around the point ∞. On the sphere, these correspond to balls drawn around $\mathfrak{n}$, and under the map σ, they map to complements of closed balls on the plane $\mathbb{C}$. This forces us to define that a Euclidean disk of radius ε centred at $z = \infty \in \hat{\mathbb{C}}$ is $B_\varepsilon(\infty) \coloneqq \{w \in \mathbb{C} : |w| > \varepsilon\} \cup \{\infty\}$. So, open neighbourhoods around ∞ are just complements (in $\hat{\mathbb{C}}$) of compact sets $K \subseteq \mathbb{C}$.

1.3 Length of a curve in $\mathbb{H}$

Suppose we have a curve $\gamma : [0,1] \to \mathbb{R}^2$. We know that the length of this curve is given by $\int_0^1 |\dot\gamma(t)|\, dt$. In other words, if we have $\gamma = (x,y)$ then the length of the curve is $\int_0^1 \sqrt{\dot x(t)^2 + \dot y(t)^2}\, dt$. If s is the variable denoting the movement along the curve γ, then $(\Delta s)^2 \simeq (\Delta x)^2 + (\Delta y)^2$. Here Δx really means $x(t+\Delta t) - x(t)$. Similarly for y and s. Letting $\Delta t \to 0$ gives $|ds| = \sqrt{(dx)^2 + (dy)^2}$. In fact the length is really

$$\int_\gamma |ds| = \int_0^1 \sqrt{\left(\frac{dx}{dt}\right)^2 + \left(\frac{dy}{dt}\right)^2}\, dt$$

More generally denote the variable $\boldsymbol{p} = (x,y)$. Let $A : \mathbb{R}^2 \to \mathbb{R}$ be a positive definite form (hence symmetric), and consider $\|ds\|_A := \sqrt{\langle d\boldsymbol{p}, A\, d\boldsymbol{p}\rangle}$. This will allow us to measure length of curves:

$$\int_\gamma \|ds\|_A = \int_0^1 \sqrt{a(\gamma(t))\dot x(t)^2 + 2b(\gamma(t))\dot x(t)\dot y(t) + c(\gamma(t))\dot y(t)^2}\, dt$$

where $A(\boldsymbol{p}) = \begin{bmatrix} a(\boldsymbol{p}) & b(\boldsymbol{p}) \\ b(\boldsymbol{p}) & c(\boldsymbol{p}) \end{bmatrix}$ with the requirement that $\det A(\boldsymbol{p}) > 0 \forall \boldsymbol{p} \in \mathbb{R}^2$

Consider the upper half plane in $\mathbb{C}$, namely, $\mathbb{H} = \{z \in \mathbb{C} \mid \Im(z) > 0\}$. We may define the metric

$$|ds| := \frac{\sqrt{(dx)^2 + (dy)^2}}{y} = \frac{|dz|}{\Im z}.$$

With this metric, the length of a curve $\gamma : [0,1] \to \mathbb{H}$ is given by $\displaystyle\int_0^1 \frac{|\dot\gamma(t)|}{\Im\gamma(t)}\, dt.$

Example. Consider the horizontal line $\gamma : [0,1] \to \mathbb{H}$ given by $t \mapsto (kt, b)$ where $k \in \mathbb{R}$ is a fixed constant and $b > 0$. The length of this curve is

$$\int_0^1 \frac{\sqrt{\dot x(t)^2 + \dot y(t)^2}}{\Im\gamma(t)}\, dt = \int_0^1 \frac{|k|}{b}\, dt = \frac{|k|}{b}$$

Example. Consider the diagonal line $\gamma : [0,1] \to \mathbb{H}$ given by $t \mapsto (a + kt, b + \ell t)$ where $a \in \mathbb{R}, b > 0, k \geq 0, \ell > 0$ are constants. The length of this curve is

$$\int_0^1 \frac{\sqrt{\dot x(t)^2 + \dot y(t)^2}}{\Im\gamma(t)}\, dt = \int_0^1 \frac{\sqrt{k^2 + \ell^2}}{b + \ell t}\, dt = \sqrt{1 + \left(\frac{k}{\ell}\right)^2} \log\left(1 + \frac{\ell}{b}\right)$$

1.4 Geodesics in $\mathbb{H}$

A *geodesic* is a (piecewise differentiable) curve which minimizes local distances. More formally, a geodesic between two points $z_1, z_2 \in \mathbb{H}$ is a curve $\gamma : [0,1] \to \mathbb{H}$ such that $\gamma(0) = z_1, \gamma(1) = z_2$ and $\operatorname{len}(\gamma) = \min\left\{\operatorname{len}(\alpha) \;\middle|\; \alpha \in \mathbb{H}^{[0,1]}, \alpha \text{ is piecewise differentiable}, \alpha(0) = z_1, \alpha(1) = z_2\right\}$.

Fact. Say $\gamma = (x,y)$ minimizes $\displaystyle\int_a^b \mathscr{L}(\boldsymbol{p}(t), \dot{\boldsymbol{p}}(t))\, dt$, where $\mathscr{L}$ is smooth, then the Euler-Lagrange equation is satisfied:

$$\frac{\partial \mathscr{L}(\gamma, \dot\gamma)}{\partial \boldsymbol{p}} = \frac{d}{dt}\frac{\partial \mathscr{L}(\gamma, \dot\gamma)}{\partial \dot{\boldsymbol{p}}}$$

With this in mind, we can calculate the geodesics in our hyperbolic plane. Take $\mathscr{L} = \dfrac{\sqrt{\dot{x}(t)^2 + \dot{y}(t)^2}}{y(t)}$.

We have $\dfrac{\partial \mathscr{L}}{\partial x} = \dfrac{d}{dt}\dfrac{\partial \mathscr{L}}{\partial \dot{x}} \implies \dfrac{d}{dt}\left(\dfrac{\dot{x}}{y\sqrt{\dot{x}^2+\dot{y}^2}}\right) = 0 \implies \dfrac{\dot{x}}{y\sqrt{\dot{x}^2+\dot{y}^2}} = c$ for some constant c.

Also, $\dfrac{\partial \mathscr{L}}{\partial y} = \dfrac{d}{dt}\dfrac{\partial \mathscr{L}}{\partial \dot{y}} \implies \dfrac{\sqrt{\dot{x}^2+\dot{y}^2}}{-y^2} = \dfrac{d}{dt}\left(\dfrac{\dot{y}}{y\sqrt{\dot{x}^2+\dot{y}^2}}\right)$.

If we have $c = 0$ whence $x(t)$ is a constant, giving us a straight line. Otherwise we have the following:

$$\begin{aligned}\frac{\sqrt{\dot{x}^2+\dot{y}^2}}{-y^2\dot{x}} &= \frac{\frac{d}{dt}\left(\frac{\dot{y}}{y\sqrt{\dot{x}^2+\dot{y}^2}}\right)}{\dot{x}} = \frac{d}{dx}\left(\frac{c\dot{y}}{\dot{x}}\right) = cy'' \\ \implies \sqrt{1+(y')^2} &= \frac{-y''y^2}{y\sqrt{1+(y')^2}} \\ \implies \frac{1}{y} + \frac{(y')^2}{y} + y'' &= 0\end{aligned}$$

The above differential equation can be solved by substituting $u = y'$. The final solution comes out to be $(x-k)^2 + y^2 = r$ where r, k are some constants. This is clearly an (Euclidean) circle with radius r centered at $(k, 0)$. Since this circle passes through $z_1, z_2 \in \mathbb{H}$ (with $\Re(z_1) \neq \Re(z_2)$), k can be uniquely determined simply with ruler-compass construction - it's the intersection of the perpendicular bisector of the segment joining z_1, z_2 with the real axis. In fact, even if we have $\Re(z_1) = \Re(z_2)$ we will get that $x(t)$ is constant, whence the perpendicular bisector will meet the real axis at $\infty \in \hat{\mathbb{C}}$.

We have a result which makes this geometry 'nice':

Proposition 1.1. For each pair p and q of distinct points in $\mathbb{H}$, there exists a unique hyperbolic line l in $\mathbb{H}$ passing through p and q.

Proof. There are two cases to consider.
Suppose $\Re p = \Re q$. Then, the Euclidean line $L = \{z \in \mathbb{C} \mid \Re z = \Re p\}$ is perpendicular to the real axis and passes through both p and q. So, $l = \mathbb{H} \bigcap L$ does the job.
Suppose $\Re p \neq \Re q$. Simply join p, q and draw their perpendicular bisector. Suppose this meets the real axis at c. Then the circle $C = \partial B_{|p-c|}(c)$ is the unique Euclidean circle passing through p, q. So $l = C \bigcap \mathbb{H}$ is the required hyperbolic line. ∎

Definition 1.2. Two hyperbolic lines in $\mathbb{H}$ are parallel if they are disjoint.

Finally we see how Hyperbolic geometry is fundamentally different from Euclidean geometry.

Theorem 1.3. Let l be a hyperbolic line in $\mathbb{H}$, and let p be a point in $\mathbb{H}$ not on l. Then, there exist infinitely many distinct hyperbolic lines through p that are parallel to l.

Proof. There are two cases to consider.
Suppose l is contained in a vertical Euclidean line $L : x = k$. Take a point $x \in \mathbb{R}$ between $\Re p$ and k, and let A be the Euclidean circle centred on $\mathbb{R}$ that passes through x and p. We have already seen that such an Euclidean circle C is unique. We can do this for all x between $\Re p, k$.
Suppose l is contained in a Euclidean circle C which intersects the real axis at points $a < b$. WLOG $\Re p > b$. Take any point $x \in \mathbb{R}$ between b and $\Re p$. Let E be the Euclidean circle centred on $\mathbb{R}$ that passes through x and p. We can do this for infinitely many x between $b, \Re p$. ∎

1.5 Distance between points in $\mathbb{H}$

Consider points $z_0, z_1 \in \mathbb{H}$. Distance between these points is simply the length (under the hyperbolic metric) of the geodesic curve joining them.

For a moment assume that z_0, z_1 have unequal real parts. Then the perpendicular bisector of the segment $\overline{z_0 z_1}$ is $y(y_1 - y_0) + x(x_1 - x_0) + \frac{x_0^2+y_0^2-x_1^2-y_1^2}{2} = 0$. So, it meets the real axis at the point $\left(\frac{x_1^2+y_1^2-x_0^2-y_0^2}{2(x_1-x_0)}, 0\right)$. WLOG we can assume that this meeting point is $0 \in \mathbb{C}$. This is because translating along the real axis does not change the length (length of curve only depends on y): just translate everything by $\frac{x_1^2+y_1^2-x_0^2-y_0^2}{2(x_0-x_1)}$ along the horizontal axis. So, we are assuming that $|z_0| = |z_1| = r$ (say).

So all we are left to find is the distance between $z_{0,1}$ along the circle of radius r centered at 0. This curve is just given by $\gamma : [a, b] \to \mathbb{H}, t \mapsto r(\cos t, \sin t)$ where $\gamma(a) = z_0, \gamma(b) = z_1$. The length is

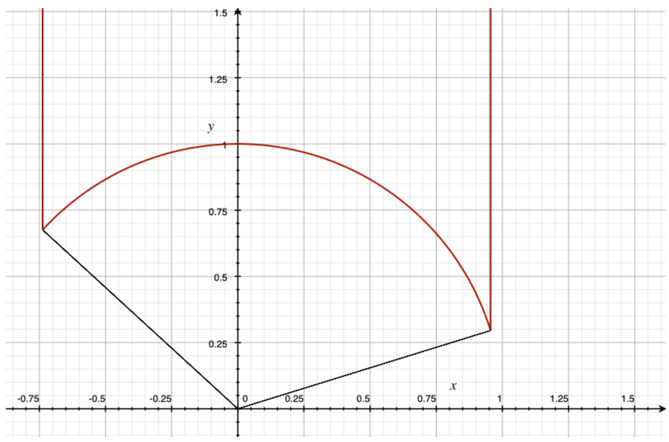
$$\int_a^b \frac{\sqrt{\dot{x}^2 + \dot{y}^2}}{y} dt = \int_a^b \frac{r}{r \sin t} dt = \int_a^b \csc t \, dt = \log\left(\frac{\cot a + \csc a}{\cot b + \csc b}\right) = \operatorname{arsinh}(\cot a) - \operatorname{arsinh}(\cot b)$$

Exercise 1.4. The distance between two points $z_1, z_2 \in \mathbb{H}$ is

$$2 \operatorname{arsinh} \sqrt{\frac{|z_1|^2 + |z_2|^2}{4\Im(z_1)\Im(z_2)}}$$

1.6 Triangles in $\mathbb{H}$

Triangles in $\mathbb{H}$ are determined the same way as Euclidean spaces. Choose any three distinct points and draw 'straight lines' between them. We look at a triangle with one *ideal point* for this purpose. An ideal point is a point at infinity (i.e., on $\partial\mathbb{H}$).



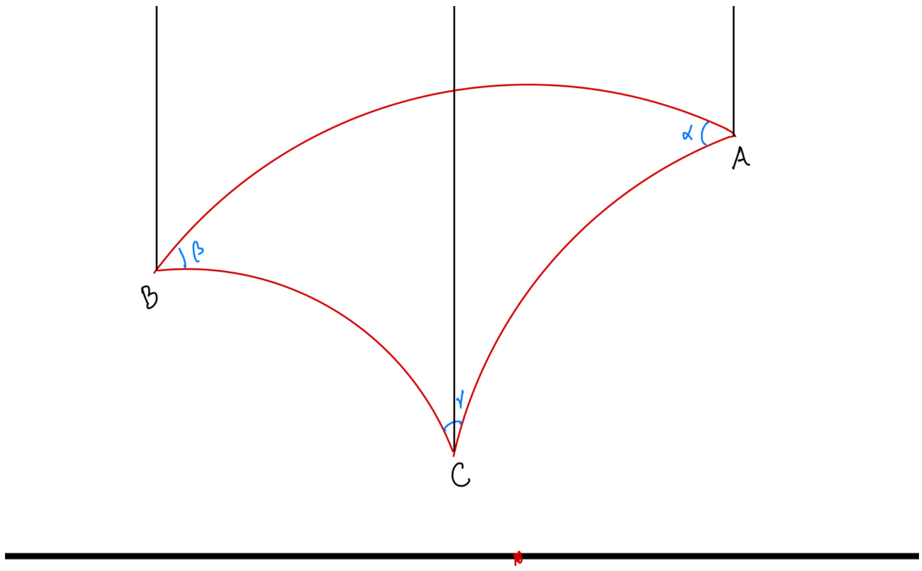
In this case, the angles of our triangle are $\alpha, \beta, 0$. α and β are the angles seen in quadrants I and II respectively. Under our hyperbolic metric, we have an area element (a small rectangle): $\frac{dx}{y}\frac{dy}{y} = \frac{dx\,dy}{y^2}$. We integrate this element over the entire triangle to find its area. instead of varying x, y we shall vary θ, y: Add thin vertical strips which meets the "circular" side at point $(r\cos\theta, r\sin\theta)$ (Here r is the Euclidean radius of the arc). So θ varies from α to $\pi - \beta$. And y varies from $r\sin\theta$

to ∞. All this goes through because of Fubbini's theorem. And we find our area to be

$$\mathscr{A} = \int_{\alpha}^{\pi-\beta} \left(\int_{r\sin\theta}^{\infty} \frac{r\sin\theta\, dy}{y^2} \right) d\theta = \int_{\alpha}^{\pi-\beta} d\theta = \pi - \alpha - \beta$$

.

Well, what about triangles with no ideal points? We will deduce that area using the above discussion.



This turns out to be (we don't write the third vertex if it is ∞)

$$\mathscr{A}(A,B,C) = \mathscr{A}(B,C) + \mathscr{A}(A,C) - \mathscr{A}(A,B) = \pi - \alpha - \beta - \gamma$$

It finally stands that $\mathscr{A} + \alpha + \beta + \gamma = \pi$.

2 The Poincaré disc

2.1 Möbius transformations

A *Möbius transformation* is a rational function $f : \mathbb{C} \to \mathbb{C}$ of the form

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

where $a,b,c,d \in \mathbb{C}$ are such that $ad - bc \neq 0$. It is not hard to see that we can scale the numerator and the denominator on the RHS expression of f and assume WLOG that $ad - bc = 1$. One may think of this as $f(z) = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{bmatrix} z \\ 1 \end{bmatrix}$. In the resultant 2×1 vector, the elements at positions $(1,1)$ and $(1,2)$ are the numerator and denominator respectively. It is a routine exercise to verify that such an expression is well-defined. Hereafter, we will identify all Möbius transformations as elements of $SL(2,\mathbb{C})$.

We can extend f to $\hat{f}:\hat{\mathbb{C}}\to\hat{\mathbb{C}}$ as follows.
If $c=0$ we say

$$\hat{f}(z)=\begin{cases} f(z) & \text{if } z\in\mathbb{C} \\ \infty & \text{if } z=\infty \end{cases}$$

If $c\neq 0$ we say

$$\hat{f}(z)=\begin{cases} f(z) & \text{if } z\in\mathbb{C}\setminus\left\{\frac{-d}{c}\right\} \\ \infty & \text{if } z=\frac{-d}{c} \\ \frac{a}{c} & \text{if } z=\infty \end{cases}$$

One can easily verify that $\hat{f}$ is a homeomorphism. In fact, the group of Möbius transformations $\hat{\mathbb{C}}\to\hat{\mathbb{C}}$ is $SL(2,\mathbb{C})/\{\pm I\}=PSL(2,\mathbb{C})$.

Fact. If $g:U\to\mathbb{C}$ is an injective holomorphic map (where $U\subset\mathbb{C}$ is a domain), then g is conformal.

Using this fact, we see the importance of Möbius transformations. Let $U=\mathbb{C}\setminus\left\{\hat{f}^{-1}(\infty)\right\}$. Now it is not hard to see that $f|_U$ is injective. So g is conformal.
It turns out that the set of meromorphic conformal automorphisms of $\hat{\mathbb{C}}$ is exactly $PSL(2,\mathbb{C})$.

2.2 The disc $\mathbb{D}$

We will be interested in the map $\varphi:=\left(z\mapsto\dfrac{z-i}{z+i}\right)$ which maps $\mathbb{H}$ to $\mathbb{D}$. Visually, $\mathbb{R}\cup\{\infty\}$ maps to the boundary of the unit circle $\infty\to 1, 1\mapsto -i, -1\mapsto i, 0\mapsto -1$. In fact this is an isometry between $\mathbb{H}$ and $\mathbb{D}=\{z\in\mathbb{C}:|z|<1\}$. We equip $\mathbb{D}$ with the metric $|ds|=\dfrac{2\,|dz|}{1-|z|^2}=\dfrac{2\sqrt{(dx)^2+(dy)^2}}{1-x^2-y^2}$.

Exercise 2.1. Check that the map $\varphi:\mathbb{H}\to\mathbb{D}$ is indeed an isometry, and $\varphi^{-1}=\psi:=\left(w\mapsto\frac{i(1+w)}{1-w}\right)$.

Exercise 2.2. $|a|_{\mathbb{D}}=\left|e^{i\theta}a\right|_{\mathbb{D}}\,\forall a\in\mathbb{D},\theta\in\mathbb{R}$.

From here it is clear that Euclidean circles, lying in $\mathbb{D}^2$, are indeed hyperbolic circles. Now take a circle in $\mathbb{D}^2$ of Euclidean radius r. The Poincaré disc radius of this circle is simply $d_{\mathbb{H}}\left(\psi(0),\psi(r)\right)=d_{\mathbb{H}}\left(i,i\dfrac{1+r}{1-r}\right)=\log\dfrac{1+r}{1-r}$. So, if we have n circles in $\mathbb{D}$ which are uniformly spaced (w.r.t. hyperbolic metric), say with radii $d,2d,\ldots,nd$, they correspond to Euclidean radii $\left\{\dfrac{e^{kd}-1}{e^{kd}+1}\right\}_{k=1}^{n}$. To imagine what the "look like", as k becomes large, the Euclidean radii are $\simeq 1-2e^{-kd}$, i.e., they 'get exponentially close'.