
Hyperbolic Geometry

JIM RICHARDSON and ASIF ZAMAN

October 21, 2010

1 Introduction

In the open unit disk \mathbb{D} , we are often concerned with the standard Euclidean geometry, where the arclength element is $|dz|$ and the metric is $d(x, y) = |x - y|$ for $x, y \in \mathbb{D}$. This geometry is the most natural to consider for most situations as it possesses many agreeable properties, namely:

- (i) The shortest distance between two points is a unique straight line.
- (ii) Given a point x and a line L , there is a unique line containing x and parallel to L .
- (iii) The angle sum of a triangle is π .
- (iv) The circumference of a circle of radius r in \mathbb{D} is $2\pi r$.
- (v) The curvature of \mathbb{D} is 0 (i.e. the unit disk is flat).

Moreover, rotations and translations are distance-preserving in the Euclidean metric. This feature is a fundamental characteristic of Euclidean geometry as it matches our intuition with everyday life. On the other hand, the conformal self-maps of \mathbb{D} do not always preserve this metric, despite the fact that these are well-behaved functions.

If these automorphisms preserved distance under some metric, we would likely have many other interesting properties. Hence, our goal is to answer the following questions:

Question 1: *Does there exist a metric for which conformal self-maps on \mathbb{D} are distance-preserving?*

Question 2: *If so, how does the new geometry compare with properties (i) to (v)?*

2 Hyperbolic Metric

Definition 2.1 (Isometry). For metric spaces X and Y , a function $f : X \rightarrow Y$ is an *isometry* if $d_Y(f(x), f(y)) = d_X(x, y)$ for all $x, y \in X$; in other words, f is distance-preserving.

For our purposes, we will study the case when $X = Y = \mathbb{D}$ and $d_X = d_Y$. Isometries convey significant information about the metric space and its structure. Recall that the specifying the metric is equivalent to specifying the arclength element used to define the length of any smooth curve in \mathbb{D} . What metric should be imposed so that conformal automorphisms of \mathbb{D} are isometries? The answer comes from the result of Schwarz–Pick’s Lemma. Recall

Theorem 2.2 (Schwarz-Pick Lemma). *If $f : \mathbb{D} \rightarrow \mathbb{D}$ is analytic, then for all $z \in \mathbb{D}$*

$$|f'(z)| \leq \frac{1 - |f(z)|^2}{1 - |z|^2} \quad (2.1)$$

Further, we have equality above if and only if $f \in \text{Aut}(\mathbb{D})$.

Hence, if we denote $w = f(z)$ where $f \in \text{Aut}(\mathbb{D})$, then from (2.1), we have

$$\left| \frac{dw}{dz} \right| = \frac{1 - |w|^2}{1 - |z|^2} \quad (2.2)$$

Rewriting in differential form,

$$\frac{|dw|}{1 - |w|^2} = \frac{|dz|}{1 - |z|^2} \quad (2.3)$$

This strongly suggests an appropriate definition for our desired metric.

Definition 2.3 (Hyperbolic length). For any smooth curve γ in \mathbb{D} , define the *hyperbolic length* of γ to be

$$L(\gamma) = 2 \int_{\gamma} \frac{|dz|}{1 - |z|^2} \quad (2.4)$$

The factor of 2 is included to force the curvature to be -1 (see Section 4). Now, observe that from (2.3) we have for $f \in \text{Aut}(\mathbb{D})$ that

$$L(f \circ \gamma) = 2 \int_{f \circ \gamma} \frac{|dz|}{1 - |z|^2} = 2 \int_{\gamma} \frac{|dw|}{1 - |w|^2} = L(\gamma) \quad (2.5)$$

Thus, the hyperbolic length of γ is invariant under conformal automorphisms of \mathbb{D} .

Definition 2.4 (Hyperbolic distance). For $x, y \in \mathbb{D}$, define the *hyperbolic distance* $\rho(x, y)$ from x to y to be

$$\rho(x, y) := \inf_{\gamma} L(\gamma) \quad (2.6)$$

where the infimum is over all smooth curves γ in \mathbb{D} from x to y .

From this definition, there is one immediate concern: for any $x, y \in \mathbb{D}$, is the hyperbolic distance $\rho(x, y)$ achieved by some smooth curve, and if so, is the curve unique? The answer to both of these questions is *yes*, but the proof will be given later (see Theorem 3.2).

For now, let us assume that for any $x, y \in \mathbb{D}$ such a curve γ exists. Given this fact, it is easy to verify that ρ satisfies the properties of a metric for \mathbb{D} . Moreover, conformal self-maps of \mathbb{D} preserve this metric since they preserve hyperbolic length. Thus, we have affirmatively answered Question 1, and so this metric gives us our desired *hyperbolic geometry* in \mathbb{D} . The metric space (\mathbb{D}, ρ) is also known as the *hyperbolic disk*.

Consider the transformed curve $f \circ \gamma$ where $f : \mathbb{D} \rightarrow \mathbb{D}$ is analytic. When does $f \circ \gamma$ have the same hyperbolic length of γ ? Can the length decrease, or increase? It turns out the hyperbolic length cannot have increased. We state this result as a theorem.

Theorem 2.5. *Every analytic function $f : \mathbb{D} \rightarrow \mathbb{D}$ is a contraction mapping with respect to the hyperbolic metric ρ . That is,*

$$\rho(f(x), f(y)) \leq \rho(x, y), \quad x, y \in \mathbb{D} \tag{2.7}$$

The inequality is strict for all $x \neq y$, except when f is a conformal self-map of \mathbb{D} in which case we have equality for all $x, y \in \mathbb{D}$.

Proof. By Theorem 3.2, we may let γ be the unique distance minimizing curve from x to y . In other words, $L(\gamma) = \rho(x, y)$. Hence, $f \circ \gamma$ is a curve from $f(x)$ to $f(y)$. Thus,

$$\begin{aligned} \rho(f(x), f(y)) &\leq 2 \int_{f \circ \gamma} \frac{|dw|}{1 - |w|^2} && \text{by definition of } \rho \\ &= 2 \int_{\gamma} \frac{|f'(z)||dz|}{1 - |f(z)|^2} && \text{reparametrizing} \\ &\leq 2 \int_{\gamma} \frac{|dz|}{1 - |z|^2} && \text{by Schwarz-Pick's Lemma} \\ &= \rho(x, y) \end{aligned}$$

From Schwarz-Pick's Lemma, we have that equality occurs if and only if f is a conformal self-map of \mathbb{D} . □

Upon review of the theorem and its proof, it becomes apparent that it is merely a reformulation of the Schwarz-Pick Lemma. Nonetheless, it still sheds light on the nature of our hyperbolic geometry.

3 Geodesics

In Euclidean geometry, as stated in property (i), a straight line is the shortest path between two points. We would like to generalize this notion for an arbitrary metric on a surface.

Definition 3.1 (Geodesic). In a Riemannian manifold, a locally distance-minimizing path between two points x, y is called a *geodesic* from x to y .

Formally speaking, this is not the precise definition of a geodesic, but the idea will suffice for our purposes. Now, it is certainly not necessary that geodesics exist. Even if they did,

the uniqueness of a geodesic may not be guaranteed. Luckily for us, we have their existence and uniqueness in our space (\mathbb{D}, ρ) as given in the following theorem.

Theorem 3.2. *For any two distinct points $z_1, z_2 \in \mathbb{D}$ there is a unique path of minimum hyperbolic length connecting the two points. If both points are nonzero, the path is the arc of the circle that passes through the points and is orthogonal to the unit circle. Otherwise, the path is the straight line between the two points.*

Proof. Take $z_1, z_2 \in \mathbb{D}$. Without loss, we may assume $z_2 \neq 0$. We have that all conformal self maps of the unit disk are of the form

$$f(z) = e^{i\theta} \frac{z - \alpha}{1 - \alpha^* z} \quad (3.1)$$

Let $f(z)$ be a conformal self map of \mathbb{D} such that $f(z_1) = 0$. If necessary, compose f with a rotation Θ so that the composition $g = \Theta \circ f$ has $g(z_2) = r$ with $r > 0$. Note that if $z_1 = 0$, then g is simply a rotation.

In general, $g(z)$ is still a conformal self-map of \mathbb{D} . In particular, g maps paths orthogonal to the unit disk to paths orthogonal to the unit disk, g preserves the hyperbolic length of paths, and g maps the set of circles and lines to the set of circles and lines.

Hence, it suffices to show that the path of minimum hyperbolic length between 0 and r is the ordinary Euclidean straight line. To see why, notice that applying g^{-1} to this straight line gives, for $z_1 \neq 0$, the arc of a circle orthogonal to the unit circle and, for $z_1 = 0$, another straight line through 0.

Suppose $\gamma(t) = x(t) + iy(t)$, $0 \leq t \leq 1$ is a smooth curve starting at 0 and ending at r . First, assume $y(t) \neq 0$ for some $t \in [0, 1]$. If we define $\alpha(t) := \text{Re}(\gamma(t)) = x(t)$, then

$$\int_{\alpha} \frac{|dz|}{1 - |z|^2} = \int_0^1 \frac{|dx(t)|}{1 - x(t)^2} < \int_0^1 \frac{|dx(t)|}{1 - |\gamma(t)|^2} \leq \int_{\gamma} \frac{|dz|}{1 - |z|^2} \quad (3.2)$$

where the first inequality is strict since $|y(t)| > 0$ over some interval and $y(t)$ is continuous. Thus, we conclude $L(\alpha) < L(\gamma)$, so we may assume $y(t) = 0$ for all $t \in [0, 1]$.

If $x(t)$ is decreasing on some interval, we may redefine $\tilde{x}(t)$ by removing any interval $[a, b] \subseteq [0, 1]$ where $x(a) = x(b)$. Then $L(\tilde{x}) < L(x)$, so we may assume $x(t)$ is always increasing. \square

This resolves property (i): geodesics exist and are unique but they are *not* necessarily straight lines in \mathbb{D} . The geodesics of \mathbb{D} are given by arcs of circles that are perpendicular to the unit circle. In fact, upon further inspection of the proof and result, we have also resolved property (ii). Given a point and geodesic in \mathbb{D} , we can give *infinitely many* lines parallel to the geodesic. Note that two lines are considered *parallel* if they are disjoint.

Without loss, by a construction similar to the proof of Theorem 3.2, we may assume the given line is the real axis in \mathbb{D} and, by rotation, the point z_0 is in the lower half plane. If we consider a geodesic passing through z_0 and any z in the lower half plane, then the geodesic

must remain entirely in the lower half plane. In this way, we have defined infinitely many geodesics through z_0 disjoint from the real axis.

4 Curvature

Curvature is a very important idea in geometry and a subject of much research. We give a brief introduction to the curvature of 2 manifolds, more specifically surfaces in 3 dimensional space.

4.1 Curvature of a path

Suppose we have a smooth curve $\gamma : [a, b] \rightarrow \mathbb{R}^2$. How do we make sense of its curvature? Imagine we have two such paths γ_1 and γ_2 .

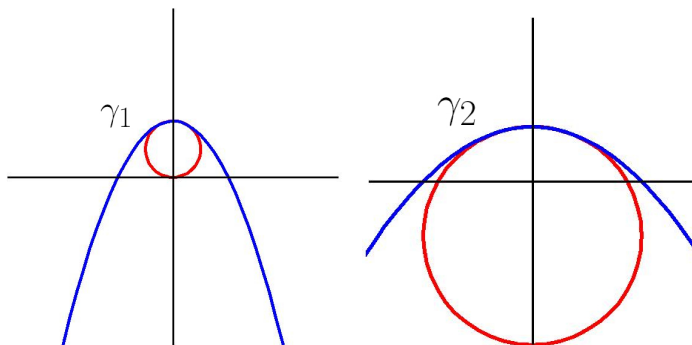


Figure 1: Two paths of different curvature

The first thing we notice is that γ_1 is a lot more curved at t_0 than γ_2 . The second thing we notice is that a circle intersecting t_0 has been drawn on each curve, and moreover in each case it seems to be the largest circle that intersects t_0 without intersecting somewhere else on the curve. The third thing we notice is that the circle on γ_1 is a lot smaller than the circle on γ_2 . We can use this to define curvature.

Definition 4.1 (Curvature of a path). The *curvature of a path* γ at a point t_0 is $\kappa(t_0) := \frac{1}{R}$ where R is the radius of the largest circle that intersects t_0 but doesn't go through the rest of γ . When no such circle exists, let $\kappa(t_0) = 0$.

Alternatively, we may define this as

Definition 4.2 (Curvature of a path). The *curvature of a path* γ at a point t_0 is $\kappa(t_0) := |\gamma''(t_0)|$ provided γ has unit speed parametrization.

We may also adopt the convention that the curvature is negative if the circle is "below" the line and positive if the circle is "above" the line, as shown in Figure 2.

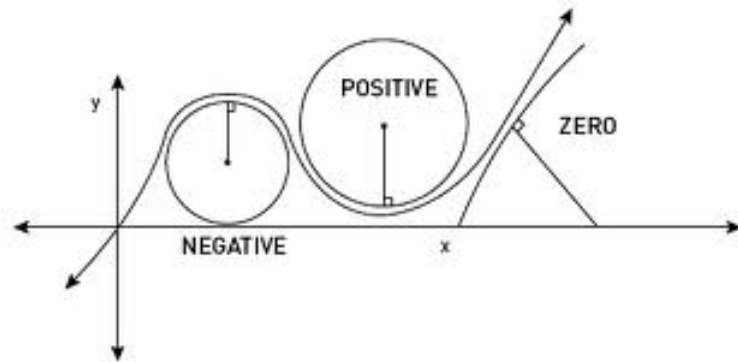


Figure 2: Positive and negative curvature

4.2 Curvature of a surface

Armed with our definition of curvature of a path, we may define the Gaussian curvature of a surface at a point. Consider a point p_0 and a plane normal to our surface that goes through this point. The intersection of this plane with our surface forms a path and we may compute the curvature of this path (see Figure 3).

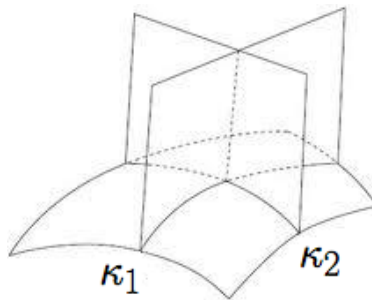


Figure 3: The principle curvatures (image from [5])

We can do this for any plane normal to the surface at p_0 and therefore we may define the maximum and minimum curvatures κ_1 and κ_2 at point p_0 , called the principle curvatures. The *Gaussian curvature* at p_0 is then defined as $G(p_0) := \kappa_1 \kappa_2$. We have three basic types of Gaussian curvature: positive, negative and zero. Examples of surfaces where every point has positive, negative and zero curvature are: the sphere, the hyperbolic plane and the euclidean plane respectively (see Figure 4).

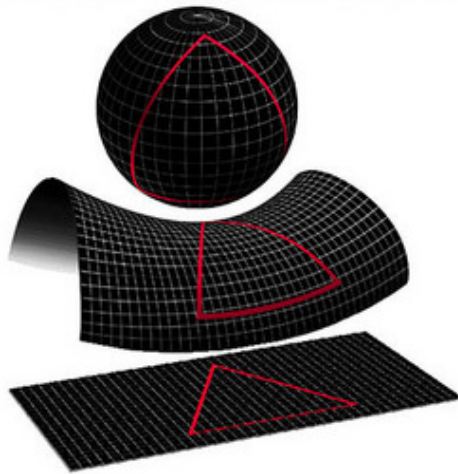


Figure 4: Positive, negative and zero Gaussian curvature (image from [6])

Defined for each of these surfaces is the idea of a geodesic, and hence geodesic triangle. Figure 4 also shows a geodesic triangle drawn on each of our surfaces. By investigating this picture closely we see that on the euclidean plane the angles in a triangle sum to exactly π , the angles of a triangle on a sphere sum to strictly larger than π and the angles of a triangle on the hyperbolic plane sum to strictly less than π , as we have already shown. If we could show that the hyperbolic disk has negative curvature everywhere, it would be an immediate consequence that the angle sum of triangles is strictly less than π .

4.3 Curvature of the Hyperbolic Disk

With our weird metric space (\mathbb{D}, ρ) , we might ask the following question: what is the curvature of this space and does this even make sense? At first glance, we can imagine \mathbb{D} as a flat disk sitting in 3 dimensions, in which case its Gaussian curvature would be zero, but this is not the case.

There is a famous result called the “Theorema Egregium” which says that the Gaussian curvature of a surface is an intrinsic property (see [4] for details). What is meant by this is that an ant crawling along the surface – unaware that the surface is curving in three dimensions – should still be able to measure its curvature. More formally, the Gaussian curvature of an embedded smooth surface is invariant under the local isometries. This will be useful for thinking about our hyperbolic space.

By viewing the hyperbolic plane from the top, we can consider the ant’s perspective. From this view, it is impossible to tell the surface is curving in three dimensions. It essentially looks like a flat plane with the geodesics given by strange curved paths – much like the metric space (\mathbb{D}, ρ) we have defined. How can we determine the curvature of the hyperbolic disk?

Based on Figure 4, studying angle sums of triangles should give us some intuition about

the curvature of our space. Let us define triangles in our new geometry in the obvious way and resolve property (iii).

Definition 4.3 (Geodesic Triangle). A *geodesic triangle* is formed by the three geodesics between three distinct points $z_1, z_2, z_3 \in \mathbb{D}$ called vertices.

Proposition 4.4. *The angles in a geodesic triangle on \mathbb{D} sum to strictly less than π .*

Proof. Suppose we have a geodesic triangle with vertices $z_1, z_2, z_3 \in \mathbb{D}$. Let $w = f(z)$ be a conformal self map of \mathbb{D} with $f(z_1) = 0$. Then our triangle is mapped to a triangle with vertices $0, w_2 = f(z_2), w_3 = f(z_3) \in \mathbb{D}$.

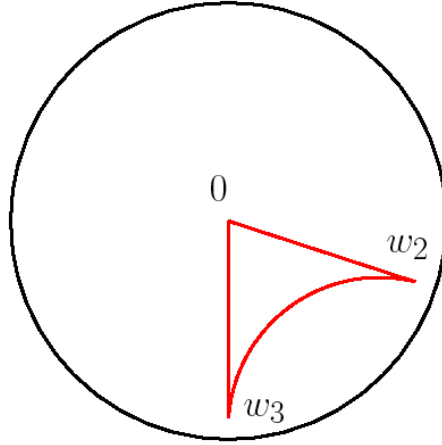


Figure 5: Geodesic triangle in \mathbb{D}

f preserves angles so it suffices to sum the angles in this new triangle. But this new triangle has two sides given by Euclidean straight lines and the other given by a circle perpendicular to the unit circle. Since this circle bends inwards to the triangle the angles must sum to strictly less than π . \square

From this result, we expect that \mathbb{D} will have negative Gaussian curvature, and so the hyperbolic disk is actually describing some surface. What is the surface and how can we compute its curvature? To do so, we first need a method of computing the Gaussian curvature of a surface that is intrinsic (i.e. does not rely on the fact that the surface is sitting in 3 dimensional space). We state, without proof, the following theorem (see [2] for details).

Theorem 4.5. *The Gaussian curvature of a surface at a point p can be computed as*

$$G(p) = \lim_{r \rightarrow 0^+} (2\pi r - C(r, p)) \frac{3}{\pi r^3} \quad (4.1)$$

where $C(r, p)$ is the circumference of the geodesic circle, centre p and radius r .

It is important to note that the *geodesic circle* implies that the radius r is with respect to our hyperbolic metric ρ . With this theorem, we may prove the following wonderful result.

Theorem 4.6. *The Gaussian curvature of (\mathbb{D}, ρ) is -1 everywhere.*

Proof. Let $p_0 \in \mathbb{D}$. We wish to compute $G(p_0)$. By applying an appropriate conformal self-map of \mathbb{D} , we may assume without loss that $p_0 = 0$.

Consider the geodesic ball $B_H(r) = \{z \in \mathbb{D} : \rho(0, z) = r\}$. By Theorem 3.2, a geodesic connecting $z \in \mathbb{D}$ to 0 is given by a straight line. Therefore, $B_H(r)$ is an ordinary Euclidean ball of some radius R . In other words, $|z| = R$ for any $z \in B_H(r)$.

Let us compute R . Let $z_0 \in B_H(r)$ and define γ be the geodesic from 0 to z_0 , and so $\gamma(t) := z_0 t$, $0 \leq t \leq 1$. Hence,

$$\begin{aligned} r &= 2 \int_{\gamma} \frac{|dz|}{1 - |z|^2} = 2 \int_0^1 \frac{R dt}{1 - R^2 t^2} \\ &= \int_0^1 \frac{R}{1 - Rt} dt + \int_0^1 \frac{R}{1 + Rt} dt \\ &= -\log(1 - Rt) + \log(1 + Rt) \Big|_0^1 \\ &= \log \left(\frac{1 + R}{1 - R} \right) \end{aligned}$$

Solving for R , we find $R = \frac{e^r - 1}{e^r + 1}$. Therefore $B_H(r) = B(R)$ where B is the Euclidean circle, centre 0 and radius $R = \frac{e^r - 1}{e^r + 1}$. We may then compute $C(0, r)$. Let Γ be the path $\Gamma(t) := R e^{it}$, $0 \leq t \leq 2\pi$, then

$$\begin{aligned} C(0, r) &= 2 \int_{\gamma} \frac{|dz|}{1 - |z|^2} = 2 \int_0^{2\pi} \frac{R dt}{1 - R^2} \\ &= 4\pi \cdot \frac{R}{1 - R^2} \\ &= 4\pi \cdot \frac{\frac{e^r - 1}{e^r + 1}}{1 - \left(\frac{e^r - 1}{e^r + 1}\right)^2} \\ &= 2\pi \cdot \frac{1}{2} \frac{e^{2r} - 1}{e^r} \\ &= 2\pi \cdot \frac{e^r - e^{-r}}{2} = 2\pi \sinh r \end{aligned}$$

Expanding $\sinh(r)$ into its Taylor series, we have $2\pi \sinh(r) = 2\pi r + \frac{\pi r^3}{3} + \mathcal{O}(r^5)$. Thus,

$$\begin{aligned}
G(0) &= \lim_{r \rightarrow 0^+} \left[(2\pi r - C(r, 0)) \frac{3}{\pi r^3} \right] \\
&= \lim_{r \rightarrow 0^+} \left[\left(2\pi r - 2\pi r - \frac{\pi r^3}{3} + \mathcal{O}(r^5) \right) \frac{3}{\pi r^3} \right] \\
&= - \lim_{r \rightarrow 0^+} (1 + \mathcal{O}(r^2)) \\
&= -1
\end{aligned}$$

which completes the proof. □

References

- [1] J. Anderson, *Hyperbolic Geometry*, 2nd ed., Springer, (2005).
- [2] T. Feeman, *Portraits of the earth: a mathematician looks at maps*, AMS Bookstore, (2002), pp. 35–40.
- [3] T. Gamelin, *Complex analysis*, Springer, (2001), pp. 266–270.
- [4] K. Gauss, *General Investigations of Curved Surfaces - Unabridged*, (transl. by A. Hiltebeitel, J. Morehead) Watchmaker Publishing, (2007).
- [5] *Geological Society of America*. Retrieved October 20, 2010. available from: <http://geology.gsapubs.org/content/35/6/519/F1.expansion.html>
- [6] *NASA Imagine*. Retrieved October 20, 2010, available from: <http://imagine.gsfc.nasa.gov/Images/map/geometry.jpg>
- [7] B. Iversen, *Hyperbolic Geometry*, Cambridge University Press, (1992).
- [8] S. Katok, *Fuchsian Groups*, The University of Chicago Press, (1992).