COMP760: Geometry and Generative Models
Week 7: Hyperbolic Geometry

Notes by: Joey Bose and Prakash Panangaden

1

HYPERBOLIC GEOMETRY

Historically, hyperbolic space arose as an example of a geometry that is internally consistent but breaks Euclid's parallel postulate. Within the Riemannian geometry framework, hyperbolic spaces are manifolds with constant negative curvature K and are of particular interest for embedding hierarchical structures. There are multiple models of n-dimensional hyperbolic space, such as the hyperboloid $\mathbb{H}^n_K$, also known as the Lorentz model, or the Poincaré ball $\mathbb{P}^n_K$. We will briefly outline these two representations of hyperbolic geometry while noting that the former is primarily used as a visualization tool and the latter for most numerical computations.

1.1 Lorentz Model

An n-dimensional hyperbolic space, is the unique, complete, simply-connected n-dimensional Riemannian manifold of constant negative curvature, $K < 0$. The Lorentz or hyperboloid model, $\mathbb{H}^n_K$, is perhaps one of the most convenient representation of hyperbolic space, since it is equipped with relatively simple explicit formulas and useful numerical stability properties [Nickel and Kiela, 2018]. Henceforth, for notational clarity, we use boldface font to denote points on the hyperboloid manifold. We choose the 2D Poincaré disk $\mathbb{P}^2_1$ to visualize hyperbolic space because of its conformal mapping to the unit disk. The Lorentz model embeds hyperbolic space $\mathbb{H}^n_K$ within the $n+1$-dimensional Minkowski space, defined as the manifold $\mathbb{R}^{n+1}$ equipped with the following inner product:

$$\langle \mathbf{x}, \mathbf{y} \rangle_{\mathcal{L}} := -x_0 y_0 + x_1 y_1 + \cdots + x_n y_n, \tag{1.1}$$

which has the type $\langle \cdot, \cdot \rangle_{\mathcal{L}} : \mathbb{R}^{n+1} \times \mathbb{R}^{n+1} \to \mathbb{R}$. It is indefinite, which may come as a shock to those used to the traditional definition of inner product from linear algebra, but this is a well-defined notion extensively used in special relativity for over a century. It is common to denote this space as $\mathbb{R}^{1,n}$ to emphasize the distinct role of the zeroth coordinate. In the Lorentz model, we model hyperbolic space as the (upper sheet of) the hyperboloid embedded in Minkowski space. It is a remarkable fact that though the Lorentzian metric is indefinite, the induced Riemannian metric $g_{\mathbf{x}}$ on the unit hyperboloid is positive definite Ratcliffe [1994]. The n-Hyperbolic space with curvature K and origin $\mathbf{o} = (1/K, 0, \ldots, 0)$, is a

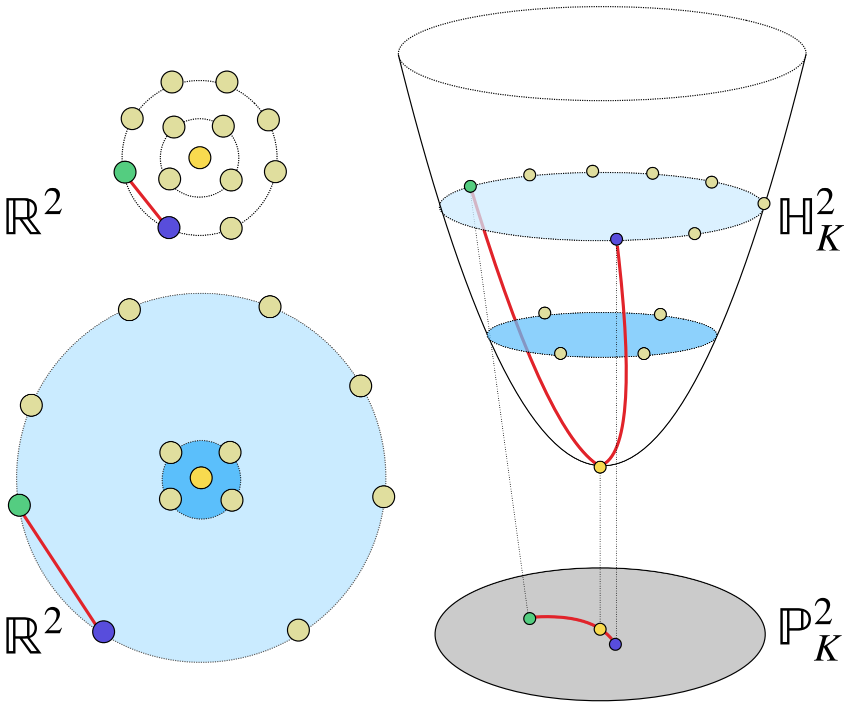

Figure 1.1: The shortest path between a given pair of node embeddings in $\mathbb{R}^2$ and hyperbolic space as modelled by the Lorentz model $\mathbb{H}^2_K$ and Poincaré disk $\mathbb{P}^2_K$. Unlike Euclidean space, distances between points grow exponentially as you move away from the origin in hyperbolic space, and thus the shortest paths between points in hyperbolic space go through a common parent node (i.e., the origin), giving rise to hierarchical and tree-like structure. Figure and caption taken from Bose et al. [2020].

Riemannian manifold $(\mathbb{H}^n_K, g_{\mathbf{x}})$ where,

$$\mathbb{H}^n_K := \{\mathbf{x} \in \mathbb{R}^{n+1} : \langle \mathbf{x}.\mathbf{x} \rangle_{\mathcal{L}} = 1/K,\ x_0 > 0,\ K < 0\}.$$

Equipped with this, the induced distance between two points $(\mathbf{x}, \mathbf{y})$ in $\mathbb{H}^n_K$ is given by

$$d(\mathbf{x}, \mathbf{y})_{\mathcal{L}} := \frac{1}{\sqrt{-K}} \operatorname{arccosh}(-K \langle \mathbf{x}, \mathbf{y} \rangle_{\mathcal{L}}). \tag{1.2}$$

This is obtained by inverting the relation for η described above.

The tangent space to the hyperboloid at the point $\mathbf{p} \in \mathbb{H}^n_K$ can also be described as an embedded subspace of $\mathbb{R}^{1,n}$. It is given by the set of points satisfying the orthogonality relation with respect to the Minkowski inner product $\mathcal{T}_{\mathbf{p}}\mathbb{H}_K^n := \{u : \langle u, \mathbf{p}\rangle_{\mathcal{L}} = 0\}$.

Of special interest are vectors in the tangent space at the origin of $\mathbb{H}_K^n$ whose norm under the Minkowski inner product is equivalent to the conventional Euclidean norm. That is $v \in \mathcal{T}_{\mathbf{o}}\mathbb{H}_K^n$ is a vector such that $v_0 = 0$ and $||\mathbf{v}||_{\mathcal{L}} := \sqrt{\langle \mathbf{v}, \mathbf{v}\rangle_{\mathcal{L}}} = ||\mathbf{v}||_2$. Thus *at the origin* the partial derivatives with respect to the ambient coordinates, $\mathbb{R}^{n+1}$, define the covariant derivative.

Projections. Starting from the extrinsic view by which we consider $\mathbb{R}^{n+1} \supset \mathbb{H}_K^n$, we may project any vector $x \in \mathbb{R}^{n+1}$ on to the hyperboloid using the shortest Euclidean distance:

$$\mathrm{proj}_{\mathbb{H}_K^n}(x) = \frac{x}{\sqrt{-K}||x||_{\mathcal{L}}}. \tag{1.3}$$

Furthermore, by definition a point on the hyperboloid satisfies $\langle \mathbf{x}, \mathbf{x}\rangle_{\mathcal{L}} = 1/K$ and thus when provided with n coordinates $\hat{x} = (x_1, \dots, x_n)$ we can always determine the missing coordinate to get a point on $\mathbb{H}_K^n$ using the relation: $x_0 = \sqrt{||\hat{x}||_2^2 + \frac{1}{K}}$.

Exponential Map. The exponential map takes a vector, v, in the tangent space of a point $\mathbf{x} \in \mathbb{H}_K^n$ to a point on the manifold—i.e., $\mathbf{y} = \exp_{\mathbf{x}}^K(v) : \mathcal{T}_{\mathbf{x}}\mathbb{H}_K^n \to \mathbb{H}_K^n$ by moving a unit length along the *geodesic*, γ (straightest parametric curve), uniquely defined by $\gamma(0) = \mathbf{x}$ with direction $\gamma'(0) = v$. The closed form expression for the exponential map is then given by

$$\exp_{\mathbf{x}}^K(v) = \cosh\Big(\frac{||v||_{\mathcal{L}}}{R}\Big)\mathbf{x} + \sinh\Big(\frac{||v||_{\mathcal{L}}}{R}\Big)\frac{Rv}{||v||_{\mathcal{L}}}, \tag{1.4}$$

where we used the *generalized radius* $R = 1/\sqrt{-K}$ in place of the curvature.

Logarithmic Map. As the inverse of the exponential map, the logarithmic map takes a point, $\mathbf{y}$, on the manifold back to the tangent space of another point $\mathbf{x}$. In the Lorentz model this is defined as

$$\log_{\mathbf{x}}^K \mathbf{y} = \frac{\operatorname{arccosh}(\alpha)}{\sqrt{\alpha^2 - 1}}(\mathbf{y} - \alpha\mathbf{x}), \quad \text{where } \alpha = K\langle \mathbf{x}, \mathbf{y}\rangle_{\mathcal{L}}. \tag{1.5}$$

Parallel Transport. The parallel transport for two points $\mathbf{x}, \mathbf{y} \in \mathbb{H}_K^n$ is a map that carries the vectors in $v \in \mathcal{T}_{\mathbf{x}}\mathbb{H}_K^n$ to corresponding vectors at $v' \in \mathcal{T}_{\mathbf{y}}\mathbb{H}_K^n$ along the geodesic. That is vectors are connected between the two tangent spaces such that the covariant derivative is unchanged. Parallel transport is a map that preserves the metric, i.e.,

$\langle \mathrm{PT}^K_{\mathbf{x}\to\mathbf{y}}(v), \mathrm{PT}^K_{\mathbf{x}\to\mathbf{y}}(v')\rangle_{\mathcal{L}} = \langle v, v'\rangle_{\mathcal{L}}$ and in the Lorentz model is given by

$$\begin{aligned}\mathrm{PT}^K_{\mathbf{x}\to\mathbf{y}}(v) &= v - \frac{\langle \log^K_{\mathbf{x}}(\mathbf{y}), v\rangle_{\mathcal{L}}}{d(\mathbf{x},\mathbf{y})_{\mathcal{L}}}(\log^K_{\mathbf{x}}(\mathbf{y}) + \log^K_{\mathbf{y}}(\mathbf{x})) \\ &= v + \frac{\langle \mathbf{y}, v\rangle_{\mathcal{L}}}{R^2 - \langle \mathbf{x}, \mathbf{y}\rangle_{\mathcal{L}}}(\mathbf{x}+\mathbf{y}),\end{aligned} \tag{1.6}$$

where α is as defined above. Another useful property is that the inverse parallel transport simply carries the vectors back along the geodesic and is simply defined as $(\mathrm{PT}^K_{\mathbf{x}\to\mathbf{y}}(v))^{-1} = \mathrm{PT}^K_{\mathbf{y}\to\mathbf{x}}(v)$.

1.2 Probability Distributions on Hyperbolic Spaces

We present the following way to define distributions on Hyperbolic spaces using the Lorentz model. For certain distributions such as the Riemannian normal we can find equivalent instantiations in the Poincaré model as well but this is generally not true.

Riemannian Normal. The first is the Riemannian normal Pennec [2006], Said et al. [2014], which is derived from maximizing the entropy given a Fréchet mean μ and a dispersion parameter σ. Specifically, we have $\mathcal{N}_{\mathcal{M}}(\mathbf{z}|\mu, \sigma^2) = \frac{1}{Z}\exp\left(-d_{\mathcal{M}}(\mu, \mathbf{z})^2/2\sigma^2\right)$, where $d_{\mathcal{M}}$ is the *induced distance* and Z is the normalization constant Said et al. [2014], Mathieu et al. [2019].

Restricted Normal. One can also restrict sampled points from the normal distribution in the ambient space to the manifold. One example is the Von Mises distribution on the unit circle and its generalized version, i.e., Von Mises-Fisher distribution on the hypersphere Davidson et al. [2018].

Wrapped Normal. We can define a wrapped normal distribution [Falorsi et al., 2019, Nagano et al., 2019], which is obtained by (1) sampling from $\mathcal{N}(0, I)$ and then transforming it to a point $v \in \mathcal{T}_{\mathbf{o}}\mathbb{H}^n_K$ by concatenating 0 as the zeroth coordinate; (2) parallel transporting the sample v from the tangent space at $\mathbf{o}$ to the tangent space of another point $\boldsymbol{\mu}$ on the manifold to obtain u; (3) mapping u from the tangent space to the manifold using the exponential map at $\boldsymbol{\mu}$. Sampling from such a distribution is straightforward and the density can be obtained via the change of variable formula,

$$\log p(\mathbf{z}) = \log p(v) - (n-1)\log\left(\frac{\sinh(\|u\|_{\mathcal{L}})}{\|u\|_{\mathcal{L}}}\right), \tag{1.7}$$

where $p(\mathbf{z})$ is the wrapped normal distribution and $p(v)$ is the tangent space normal distribution at the origin $\mathbf{o}$.

1.3 Poincaré Ball Model

The Poincaré Ball model of hyperbolic geometry is one of the many models of this geometry and it is defined as the Riemannian manifold $\mathbb{P}^n_K = (\mathcal{B}^n_K, \mathfrak{g}_K)$, where $\mathcal{B}^n_K$ is the open ball of radius $R = \frac{1}{\sqrt{|K|}}$ and $\mathfrak{g}_K$ is the Riemannian metric tensor defined as:

$$\mathfrak{g}_K(x) = (\lambda^K_x)^2 \mathfrak{g}^e(x), \tag{1.8}$$

where $\mathfrak{g}^e$ in equation 1.8 is the standard Euclidean metric and $\lambda^K_x = \frac{2}{1+K||x||^2}$. The induced volume form under the Poincaré Ball model invokes a density $\frac{d\mathrm{Vol}}{d\mathrm{Leb}} = \sqrt{|G(x)|} = (\lambda^K_x)^n$.

Note: This invariant measure is absolutely continuous w.r.t. to the Lebesgue measure.

One of the key advantages of the Poincaré Ball model over other models of hyperbolic geometry is as a tool for visualization. Specifically, this phenomenon is realized in 2-dimensions and as a result it is appropriately named the Poincaré disk model which is conformal to the 2-dimensional unit circle. An interesting property of this model is that all straight lines form circular arcs contained within the unit circle, and only the lines that bisects the unit circle by passing through the origin are conventionally straight as one would describe them in Euclidean space. This is because distances distort away from the origin as the space is growing exponentially rather than linearly as is the case in Euclidean spaces. Figure 1.2 visualizes three parallel (also called "ultraparallel") lines in the Poincaré disk model.

Mobius Addition. Unlike Euclidean spaces the Poincaré Ball model does not admit regular addition. Instead, we can perform an analogue to addition termed *Mobius addition* which comes from the theory of *gyrovector spaces* [Ungar, 2008], which can be viewed in a similar analagy to how vector spaces are used in Euclidean geometry. Formally, Mobius addition is defined as:

$$x \oplus_K y = \frac{(1 - 2K\langle x, y\rangle - K||y||^2_2)x(1 + K||x||^2_2)y}{1 - 2K\langle x, y\rangle + K^2||x||^2_2\ ||y||^2_2} \tag{1.9}$$

Mobius addition on the unit disk is neither commutative nor associative, and as a result it is fairly distinct from addition on familiar vector spaces. Armed with Mobius addition we can now define familiar concepts in Riemannian geometry such as exponential and logarithmic maps, parallel transport and distances. All of this is presented in a matter of fact manner, without further derivation.

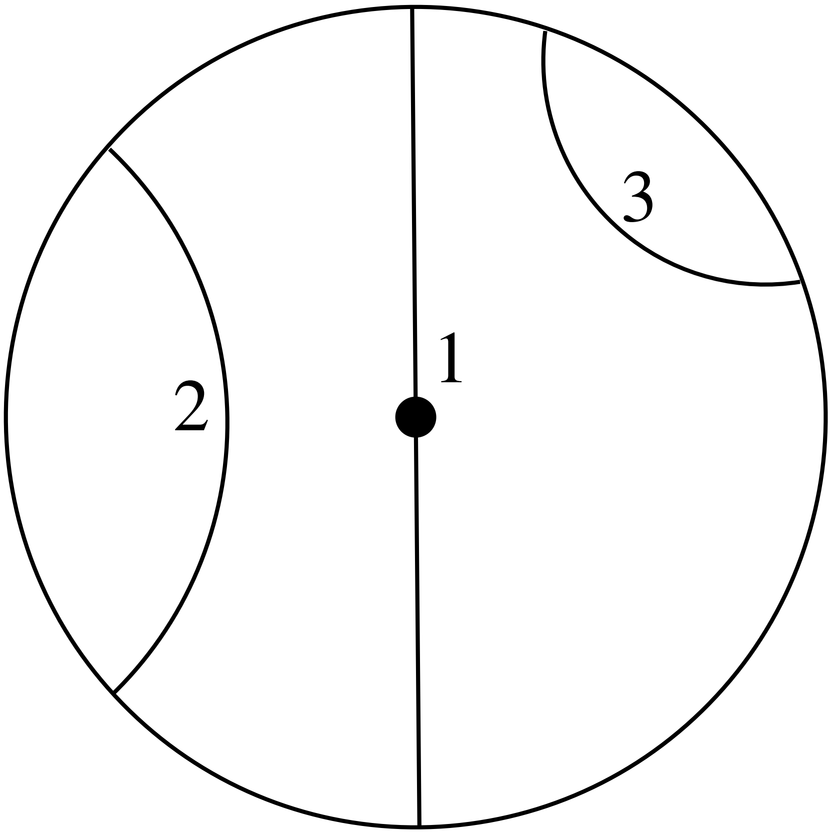

Figure 1.2: A visualization of 3 parallel lines in the Poincaré disk model.

Exponential Map. The exponential map can be conveniently written using Mobius addition as follows:

$$\exp_x^K(v) = x \oplus_K \left(\tanh\left(\sqrt{-K}\frac{\lambda_x^K ||v|||_2}{2} \right) \frac{v}{\sqrt{-K}||v||_2} \right). \tag{1.10}$$

Logarithmic Map. Correspondingly, the logarithmic map can also be expressed using Mobius addition:

$$\log_x^K(v) = \frac{2}{\sqrt{-K}\lambda_x^K} \tanh^{-1}\left(\sqrt{-K}||-x \oplus_K v||_2 \right) \frac{-x \oplus_K v}{||-x \oplus_K v||_2}. \tag{1.11}$$

Parallel Transport. To introduce parallel transport in $\mathbb{P}_K^n$ we first need to introduce the notion of gyration:

$$\mathrm{gyr}[x, y]v = -(x \oplus_K y) \oplus_K (x \oplus_K (y \oplus_K v)). \tag{1.12}$$

The hyperbolic geometric interpretation of gyration measures the extent to which Mobius addition $\oplus_K$ deviates from associativity for points $x, y, v \in \mathcal{B}_K^n$. Having defined gyration we can now write down parallel transport,

$$\mathrm{PT}_{x\to y}^K(v) = \frac{\lambda_x^K}{\lambda_y^K}\mathrm{gyr}[y, -x]v. \tag{1.13}$$

Distances. Finally, we can also write the distance between two points x, y on $\mathbb{P}_K^n$:

$$d_{\mathbb{P}}(x, y) = \frac{1}{\sqrt{-K}}\cosh^{-1}\left(1 - \frac{2K||x-y||_2^2}{(1+K||x||_2^2)(1+K||y||_2^2)}\right). \tag{1.14}$$

Bibliography

A. J. Bose, A. Smofsky, R. Liao, P. Panangaden, and W. L. Hamilton. Latent Variable Modelling with Hyperbolic Normalizing Flows. *arXiv:2002.06336 [cs, stat]*, Feb. 2020. URL http://arxiv.org/abs/2002.06336.

T. R. Davidson, L. Falorsi, N. De Cao, T. Kipf, and J. M. Tomczak. Hyperspherical variational auto-encoders. *arXiv preprint arXiv:1804.00891*, 2018.

L. Falorsi, P. de Haan, T. R. Davidson, and P. Forré. Reparameterizing Distributions on Lie Groups. *arXiv:1903.02958 [cs, math, stat]*, Mar. 2019. URL http://arxiv.org/abs/1903.02958.

E. Mathieu, C. Le Lan, C. J. Maddison, R. Tomioka, and Y. W. Teh. Continuous hierarchical representations with poincaré variational auto-encoders. In *Advances in neural information processing systems*, pages 12544–12555, 2019.

Y. Nagano, S. Yamaguchi, Y. Fujita, and M. Koyama. A wrapped normal distribution on hyperbolic space for gradient-based learning. In *International Conference on Machine Learning*, pages 4693–4702, 2019.

M. Nickel and D. Kiela. Learning continuous hierarchies in the lorentz model of hyperbolic geometry. *arXiv preprint arXiv:1806.03417*, 2018.

X. Pennec. Intrinsic statistics on riemannian manifolds: Basic tools for geometric measurements. *Journal of Mathematical Imaging and Vision*, 25(1):127, 2006.

J. G. Ratcliffe. *Foundations of Hyperbolic Manifolds*. Number 149 in Graduate Texts in Mathematics. Springer-Verlag, 1994.

S. Said, L. Bombrun, and Y. Berthoumieu. New riemannian priors on the univariate normal model. *Entropy*, 16(7):4015–4031, 2014.

A. A. Ungar. A gyrovector space approach to hyperbolic geometry. *Synthesis Lectures on Mathematics and Statistics*, 1(1):1–194, 2008.