A Brief Introduction to Hyperbolic Geometry for Machine Learning

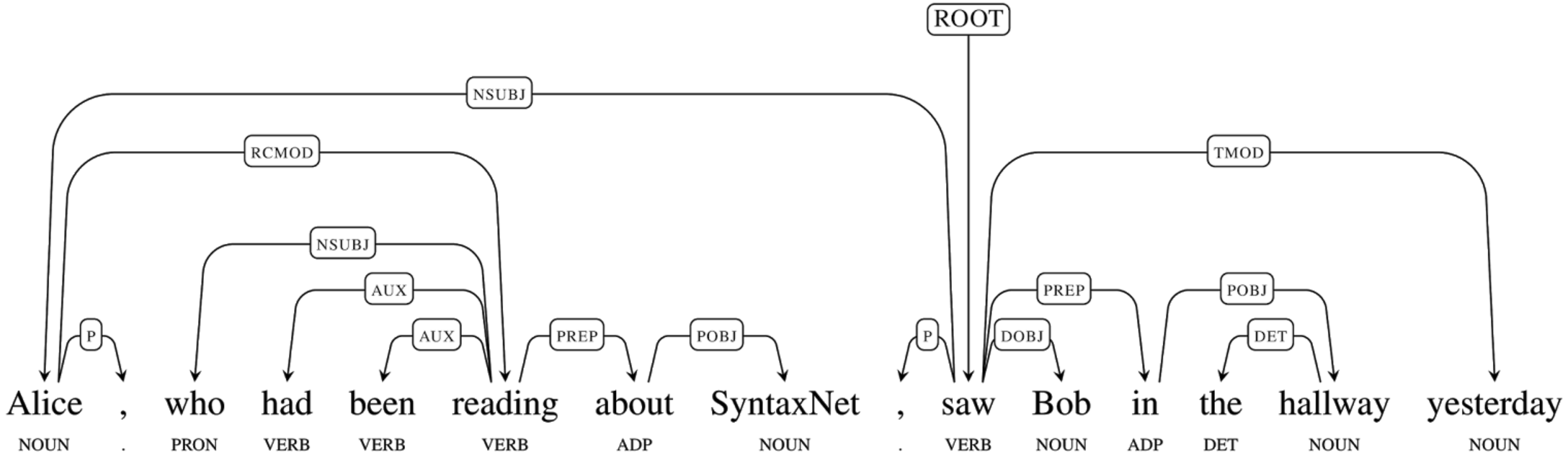
MACHINE INTELLIGENCE COMMUNITY

Justin Chen
April 13th, 2019

Motivations: Follow the Data

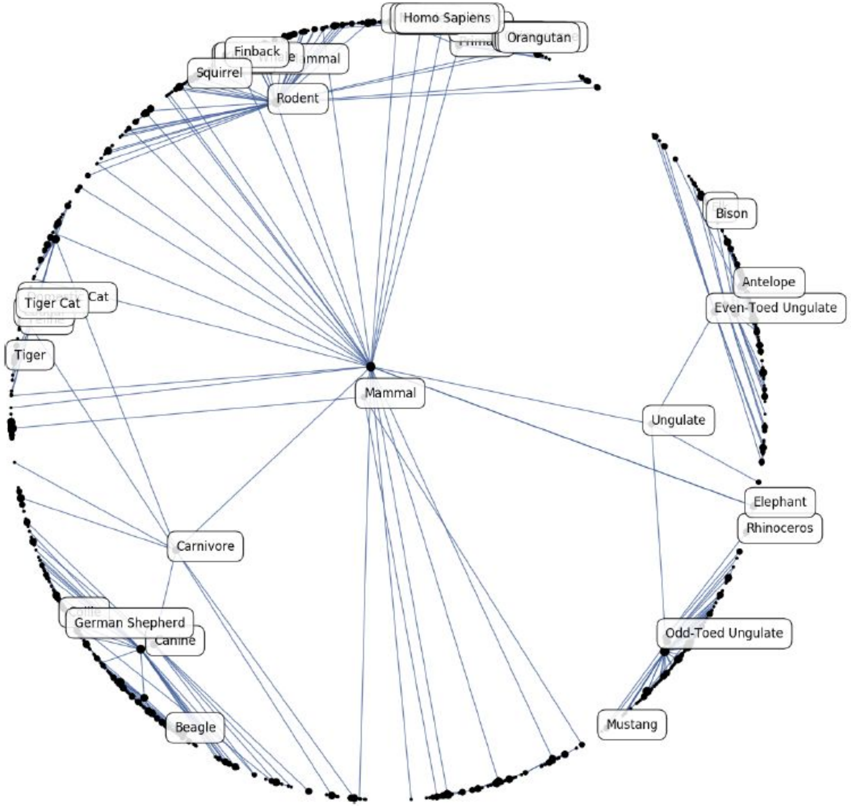
1. Machine learning is driven by data
2. Majority of algorithms and applications today rely on Euclidean data and Euclidean architectures
3. Reality is naturally non-Euclidean
    a. e.g. space-time, relations, semantic and compositional structures
4. Data that we assume to be Euclidean can be naturally represented in non-Euclidean spaces
    a. JOIN operation in MySQL
5. Next evolution of machine learning will focus on non-Euclidean models and data

Syntax Trees



https://ai.googleblog.com/2016/05/announcing-syntaxnet-worlds-most.html

Concept Hierarchies[8]



Euclidean and non-Euclidean Geometries

Absolute Geometry

All properties of geometry can be derived from **Euclid's first four postulates**.

Euclid’s First Postulate

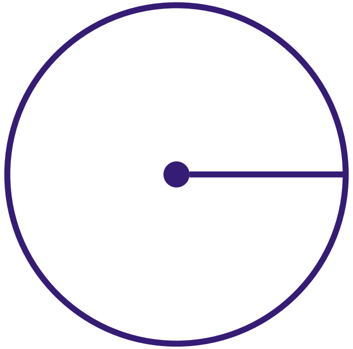
A **line segment** can be drawn by connecting two points with a **straight line**.

Euclid’s Second Postulate

A **straight line** can be extended **indefinitely** by appending straight line segments.

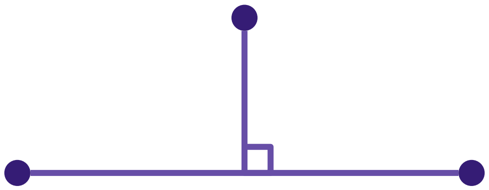
Euclid’s Third Postulate

A circle can be drawn using a straight line segment as the radius and one point as the center.



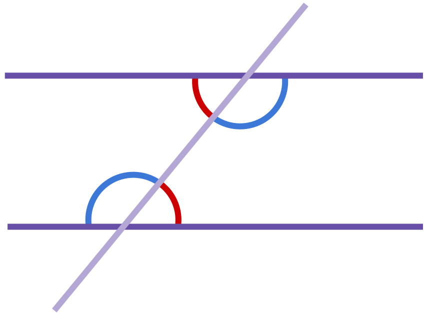
Euclid’s Fourth Postulate

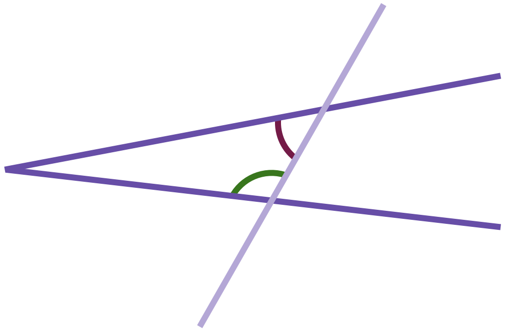
All **90 degree** angles are **congruent**.



Euclid's Fifth Postulate (Parallel Postulate)

If two lines intersect a third line and the sum of the interior angles on one side are less than 180 degrees, then the two lines are not parallel.





Euclid’s Fifth Postulate

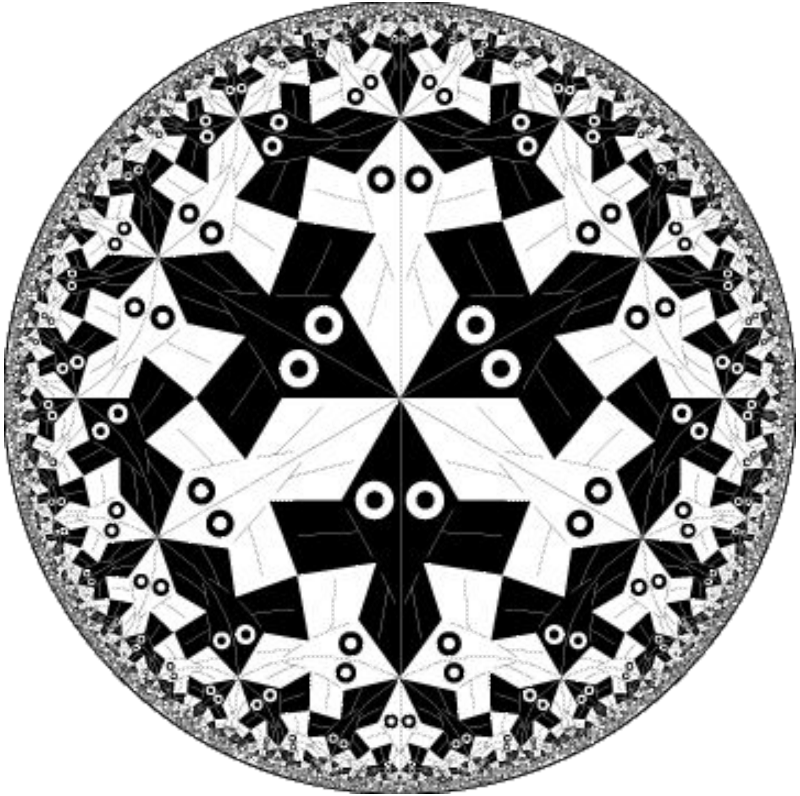
Cannot be proven as a theorem because there exists other geometries that are consistent without the 5th postulate

non-Euclidean geometries

Self-consistent geometries ignore Euclid’s Fifth Postulate

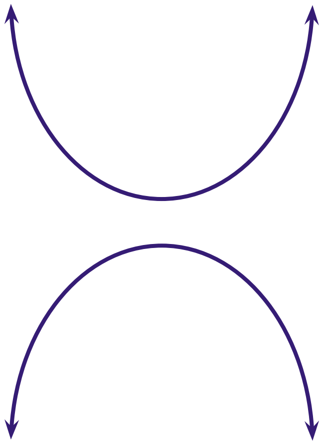
Hyperbolic Geometry

Circle Limit I, M.C. Escher

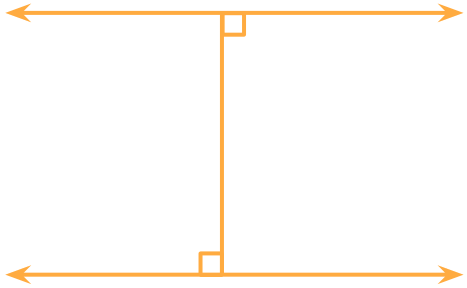


Types of Parallel Lines

Asymptotically parallel

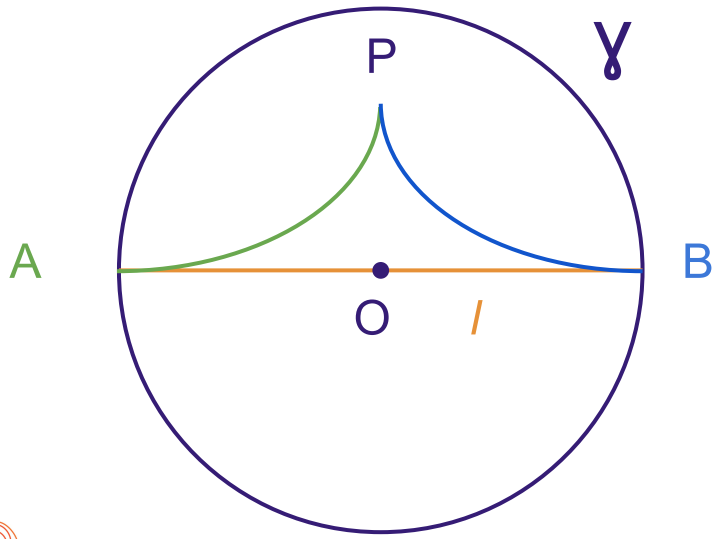


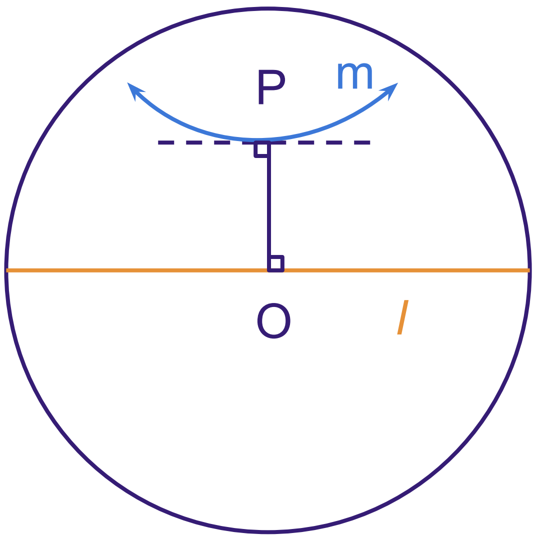
Divergently parallel



Asymptotically Parallel Lines

Two lines are parallel if they do not share any common points





Models of Hyperbolic Space

Models of Hyperbolic Geometry

1. Hyperbolic Hyperboloid
2. Beltrami-Klein
3. Poincaré Ball
4. Poincaré Half-space

All models are isometric.

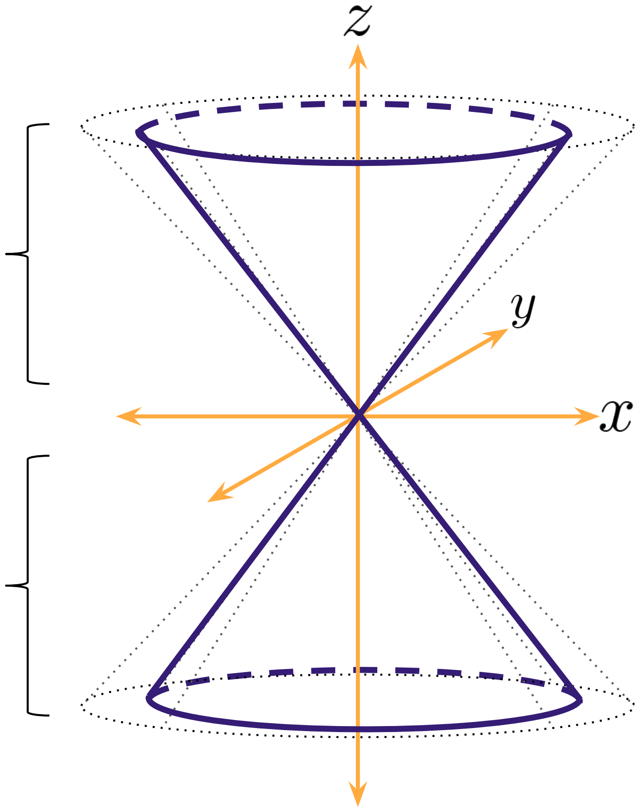
Double Cone

$$\underbrace{\frac{x^2}{a^2}}_{\text{Slope of sheet along x-axis [6]}} + \underbrace{\frac{y^2}{b^2}}_{\text{Slope of sheet along y-axis [6]}} - \underbrace{\frac{z^2}{c^2}}_{\text{Distance and slope between sheets [6]}} = 0$$

As a and b shrink towards zero, both cones flatten and translate towards each other.

As c tends towards zero, the two sheets flatten and translate towards each other.

Here A=B



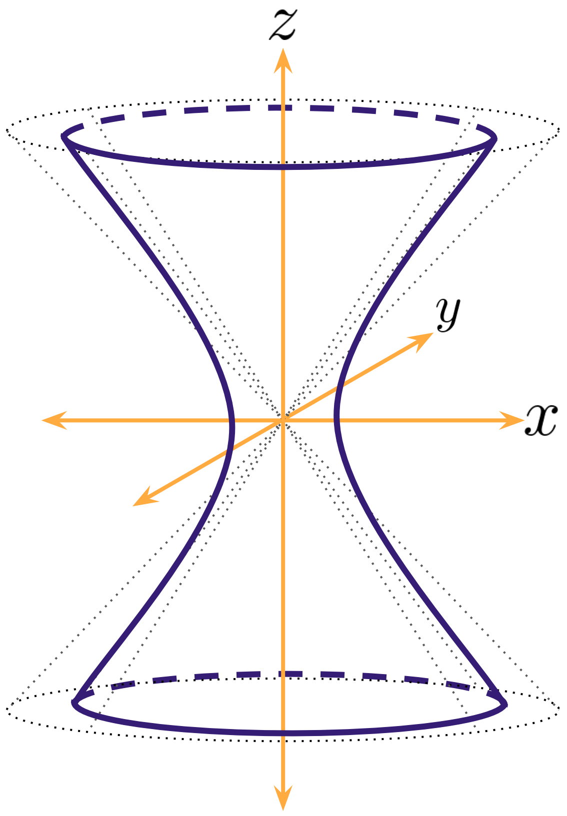
Hyperbolic Hyperboloid of 1 Sheet

$$\underbrace{\frac{x^2}{a^2}}_{} + \underbrace{\frac{y^2}{b^2}}_{} - \underbrace{\frac{z^2}{c^2}}_{} = 1$$

Slope of sheet along x-axis [6]

Slope of sheet along y-axis [6]

Distance and slope between sheets [6]



As a, b, and c tend towards zero, we recover the double cone.

Hyperbolic Hyperboloid of 2 Sheets

a.k.a. Minkowski model, Lorentz model

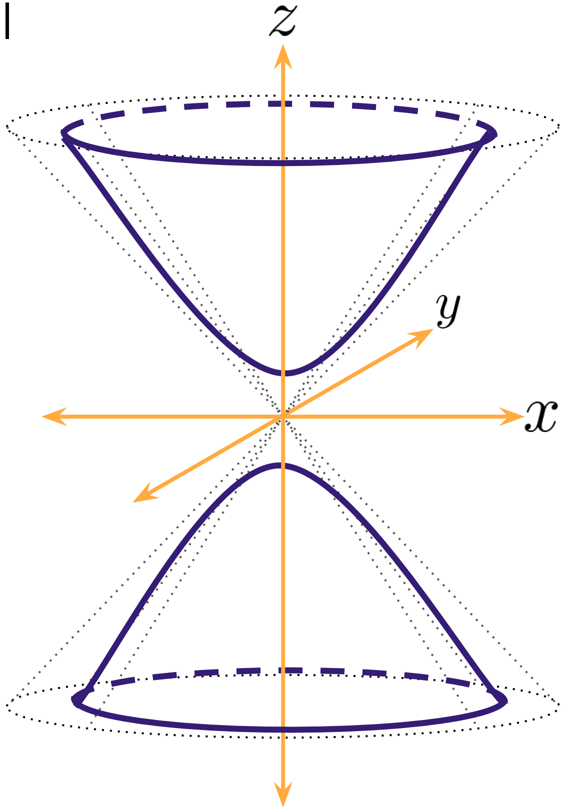
$$\underbrace{\frac{x^2}{a^2}}_{\text{Slope of sheet along x-axis [6]}} + \underbrace{\frac{y^2}{b^2}}_{\text{Slope of sheet along y-axis [6]}} - \underbrace{\frac{z^2}{c^2}}_{\text{Distance and slope between sheets [6]}} = -1$$

Forward/Future sheet

$\mathcal{S}^+ \; z \geq 1$

Backward/ Past sheet

$\mathcal{S}^- \; z \leq 1$



As a and b tend towards infinity and c tend towards zero, both sheets flatten and translate towards each other.

Beltrami-Klein Model

Represent points from hyperbolic space in a planar map

Project points from hyperboloid onto **interior** of disk in Euclidean plane. Points on boundary do not exist.

$$\mathcal{S} = \{(x, y) : x^2 + y^2 < 1\}$$

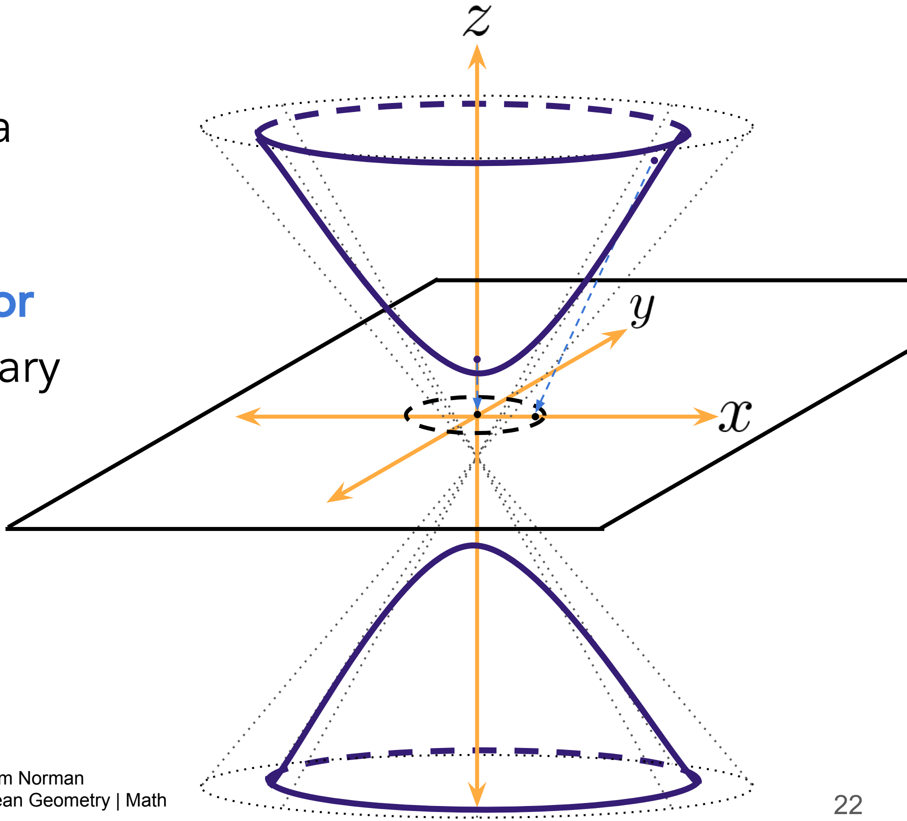


Diagram re-illustrated from Norman Wildberger's Non-Euclidean Geometry | Math History [2]

Beltrami-Klein Model

- First 4 of Euclid's postulates apply
- Lines are only represented as open chords
- Parallel postulate of Beltrami-Klein model
  - **Open chords** A and B are parallel to **open chord** C inside circle ɣ
  - Intersections outside of the circle ɣ are not part of this model
- Beltrami-Klein is **not conformal** to Euclidean space

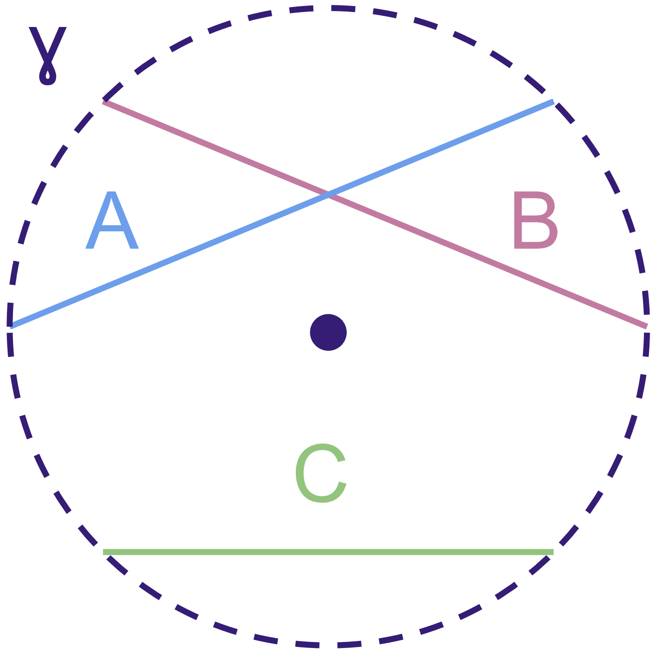


Diagram re-illustrated from Greenberg's Euclidean and Non-Euclidean Geometries Development and History, Chapter 7 [2]

Angles in Beltrami-Klein I

If either chord A or B is a **diameter** and intersect inside the circle, then A and B are **perpendicular** in the Beltrami-Klein model if and only if they are perpendicular in **Euclidean** space.

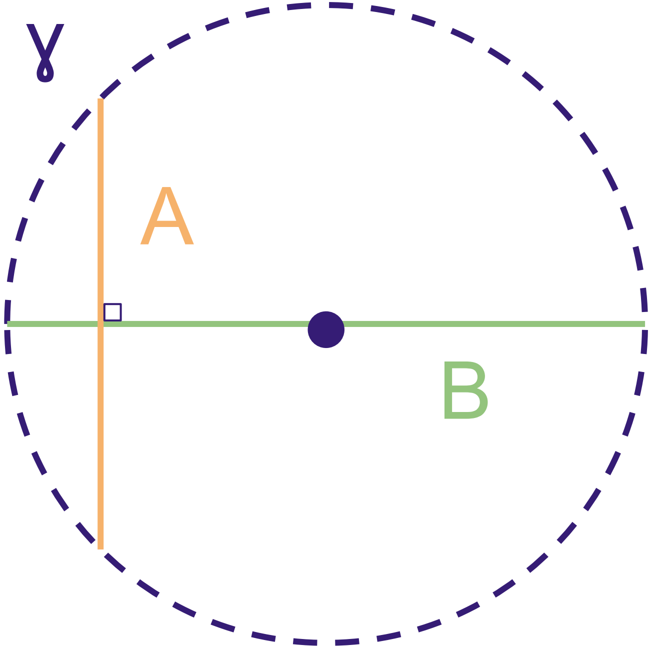


Diagram re-illustrated from Greenberg's Euclidean and Non-Euclidean Geometries Development and History, Chapter 7 [2]

Angles in Beltrami-Klein II

P(A) is the **pole** of A.

In the Beltrami-Klein, A and B are perpendicular.

This definition of perpendicularity is not equivalent to the Euclidean definition.

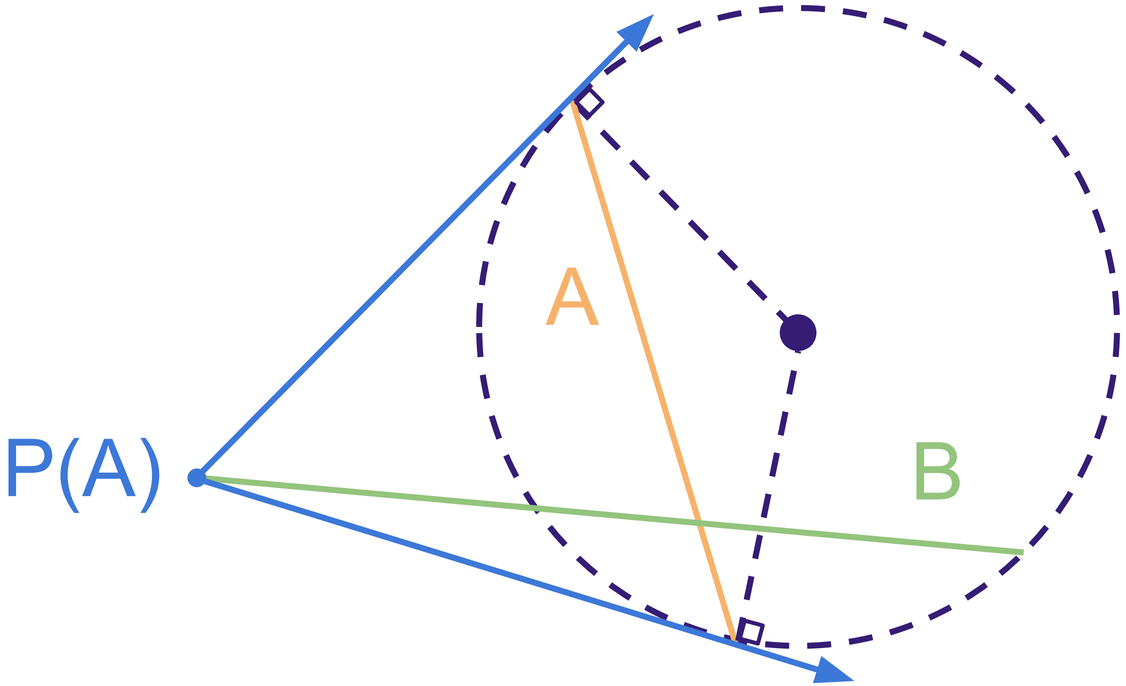
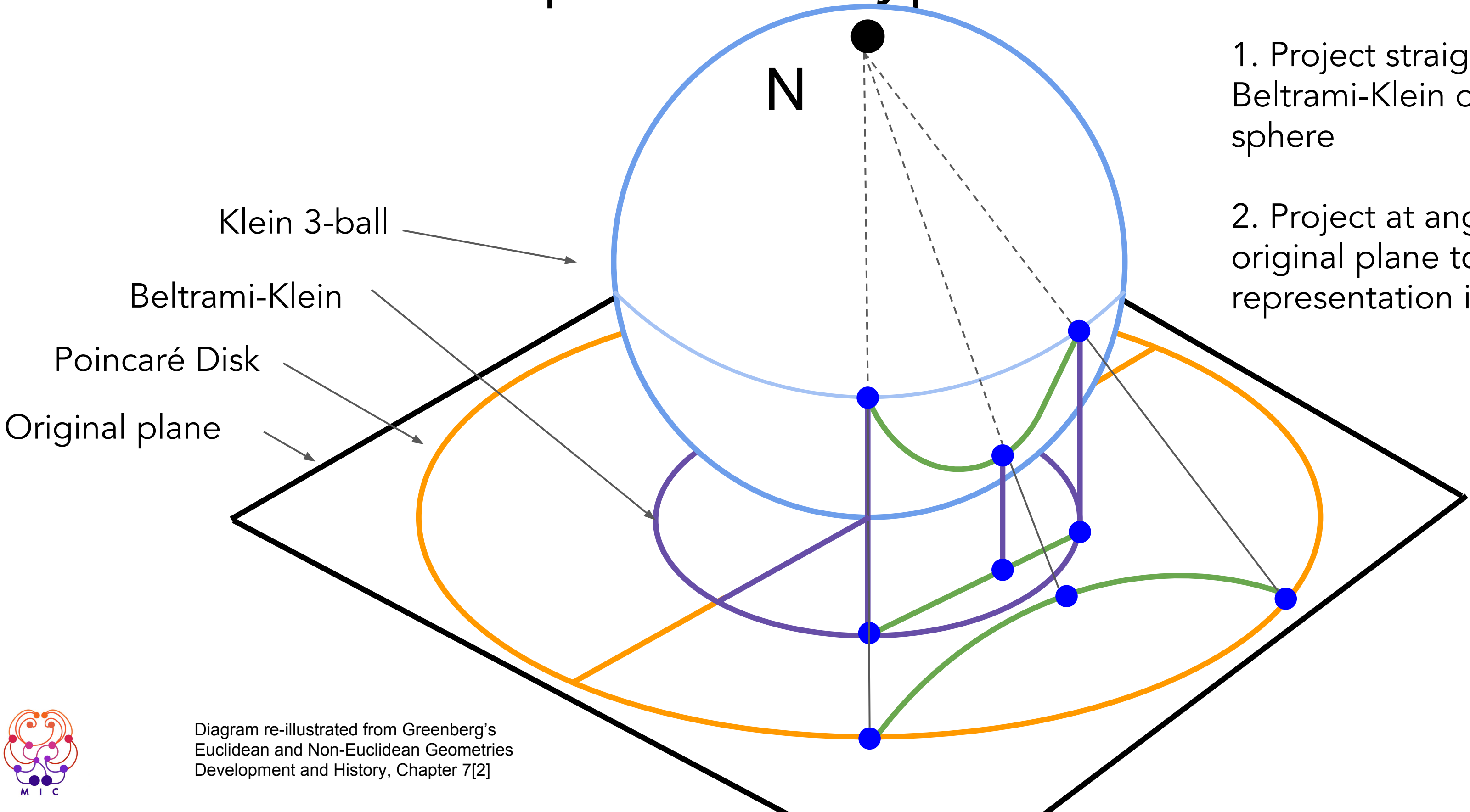


Diagram re-illustrated from Greenberg's Euclidean and Non-Euclidean Geometries Development and History, Chapter 7 [2]

Isomorphism of Hyperbolic Models



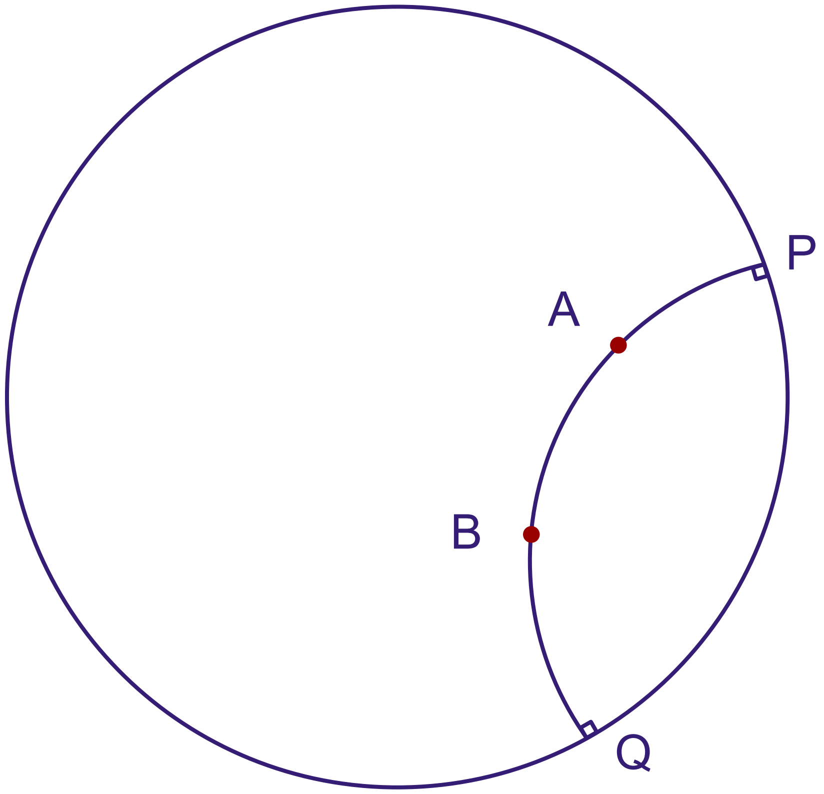
1. Project straight upward from Beltrami-Klein onto equator of sphere

2. Project at angle from N onto original plane to get isomorphic representation in Poincaré

Diagram re-illustrated from Greenberg's Euclidean and Non-Euclidean Geometries Development and History, Chapter 7[2]

Poincaré Model

Points on the Poincaré Disk



Ordinary point:
Any point within the circle

Ideal points:
Points on boundary. Because of this, they do not have corresponding points on the hyperbola.

Ideal points are infinitely far away relative to any ordinary point

Q, B, A, and P are **collinear points**

Poincaré Lines/ P-lines

All open chords *c* that pass through the origin represent lines

Hyperbolic line in the Poincaré model - all open arcs *m* are orthogonal to the circle

Geodesic - shortest path between two points on a Riemannian manifold

Any pair of points (a, b) lie on a unique geodesic.

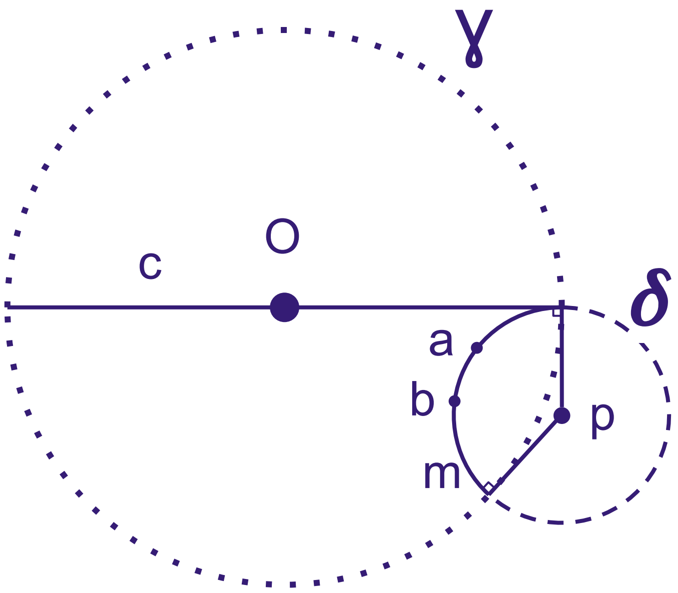
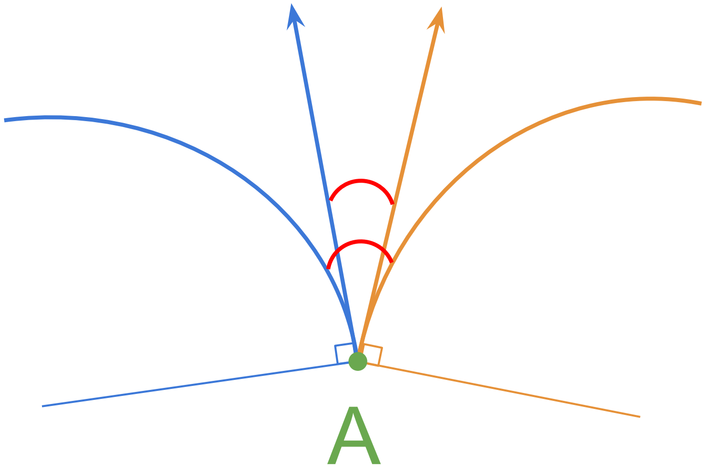


Diagram re-illustrated from Greenberg's Euclidean and Non-Euclidean Geometries Development and History, Chapter 7 [2]

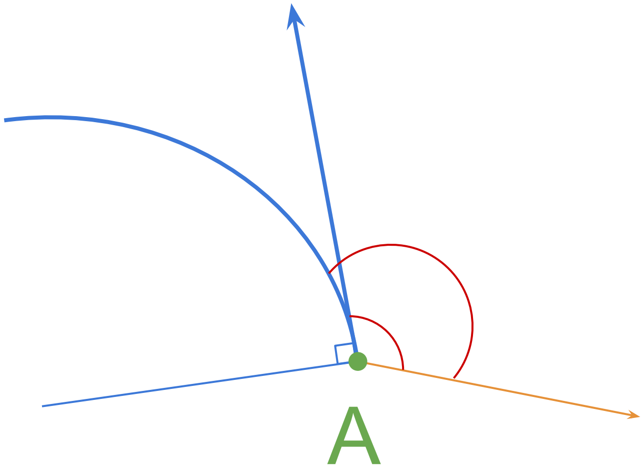
Congruence of angles I

If two directed circular arcs intersect at a point A, the degree between the arcs is equal to the degree between the tangent rays at A.



Congruence of angles II

If a directed circular arc intersects an ordinary ray at A, the number of degrees they form is equal to the number of degrees between the tangent ray and ordinary ray at A.



In some models of Hyperbolic space angles are **conformal** to Euclidean space i.e. Euclidean angles are preserved in hyperbolic space.

Cross Ratio

Projective invariant defined for **four collinear points**

Also called **double ratio** and **anharmonic ratio**

Segment lengths measured in Euclidean distance

$$(AB, PQ) = \frac{\|AP\|\|BQ\|}{\|BP\|\|AQ\|}$$

Properties of Cross Ratio

$$(AB, PQ) = \lambda$$

$$(AP, BQ) = 1 - \lambda$$

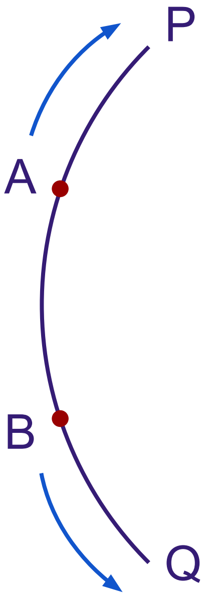
$$(AP, QB) = \frac{1}{1 - \lambda}$$

$$(AQ, PB) = \frac{\lambda}{\lambda - 1}$$

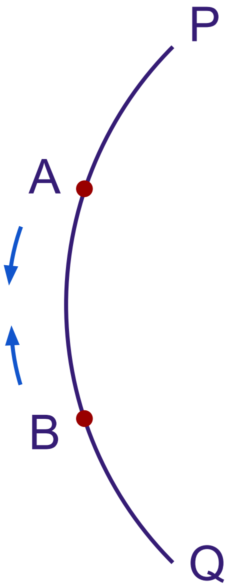
$$(AB, QP) = \frac{1}{\lambda}$$

$$(AQ, BP) = \frac{\lambda - 1}{\lambda}$$

Properties of Cross Ratio



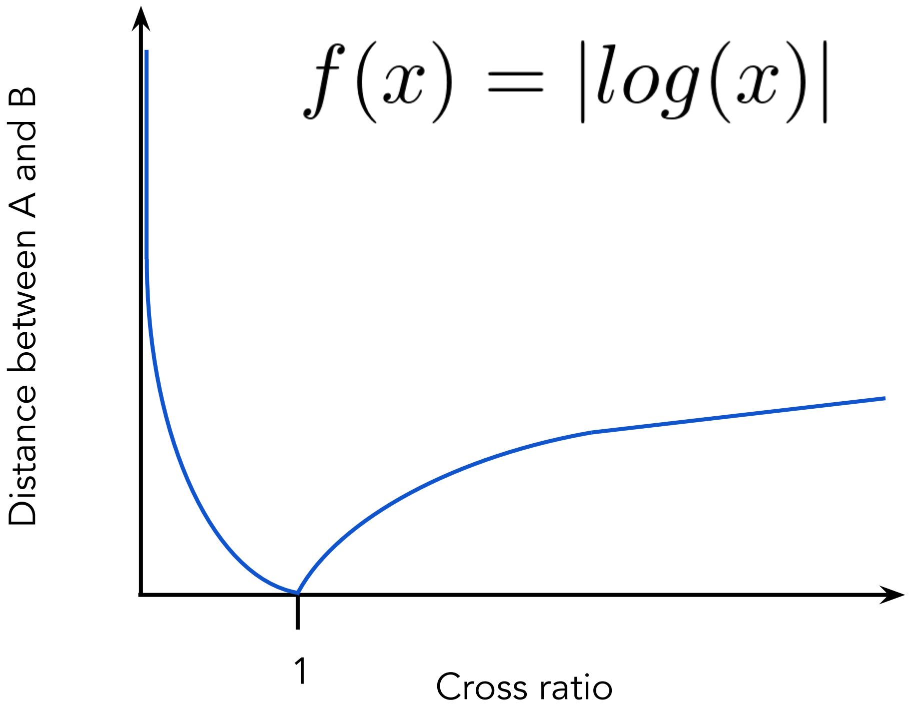
$$(AB, PQ) = 0$$



$$(AB, PQ) = 1$$

$$(AB, PQ) = \frac{\|AP\|\|BQ\|}{\|BP\|\|AQ\|}$$

Desired Behavior of Distance



Poincaré Distance

$$d(AB) = \left| log \frac{\|AP\| \|BQ\|}{\|BP\| \|AQ\|} \right|$$

Four ways to compute the Poincaré Distance, but one presented here for simplicity.

Poincaré Distance

"My point of view is that the disk has two ways to measure distance, the Euclidean way or the hyperbolic way. You can take the point of view that hyperbolic geometry takes place on the hyperboloid, where distance is measured by Euclidean distance. These points of view are equivalent: when you project the hyperboloid down to the disk, the E. distance on the hyperboloid becomes the hyperbolic distance on the disk."

Prof. Steve Rosenberg
Mathematics, Boston University

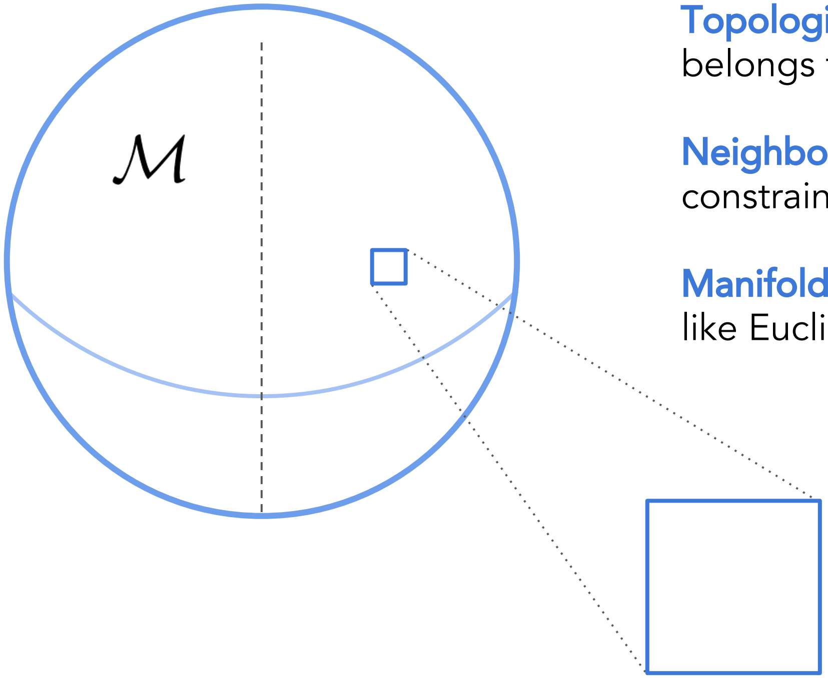
Optimization on Riemannian Manifolds

Goal of Riemannian Optimization

Learn parameters such that the inner product will constrain the points to the manifold $\mathcal{M}$

$$\forall \theta_i \in \theta : \|\theta_i\| < 1 \quad s.t. \langle \theta, x \rangle \in \mathcal{M}$$

Manifold

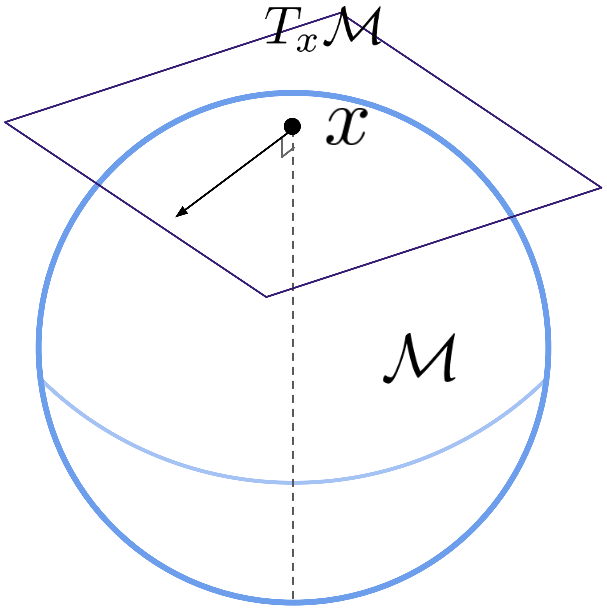


Topological space - a set of points each of which belongs to a neighborhood

Neighborhood - a set of points sharing the same constraints

Manifold - a topological space that locally behaves like Euclidean space

Metric Tensor



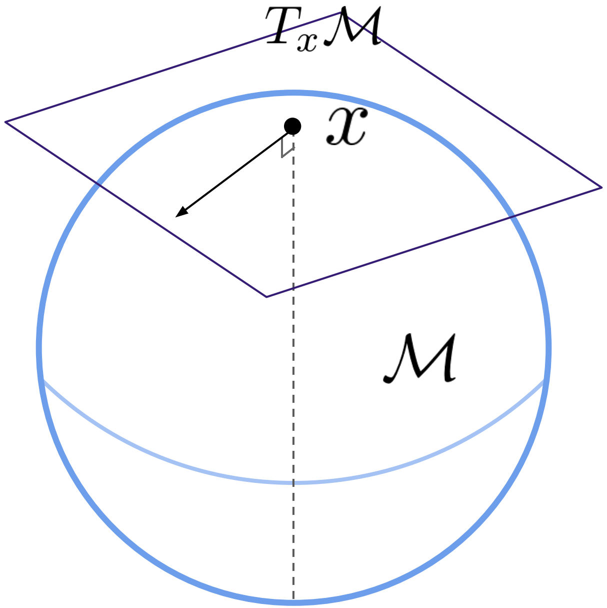
Tangent space - n-dimensional vector space $T_x\mathcal{M}$ that first order approximates the manifold $\mathcal{M}$ around point x.

Metric tensor - a function that takes two **tangent vectors** and performs **dot product** measuring an infinitesimal distance on the manifold.

$$g^E = \begin{bmatrix} 1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & 1 \end{bmatrix} \qquad \langle A, B \rangle = \begin{bmatrix} a_1 & \cdots & a_n \end{bmatrix} \begin{bmatrix} 1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & 1 \end{bmatrix} \begin{bmatrix} b_1 \\ \vdots \\ b_n \end{bmatrix}$$

Integrating the metric tensor gives distance between tangent vectors.

Riemannian Manifold



Poincaré metric

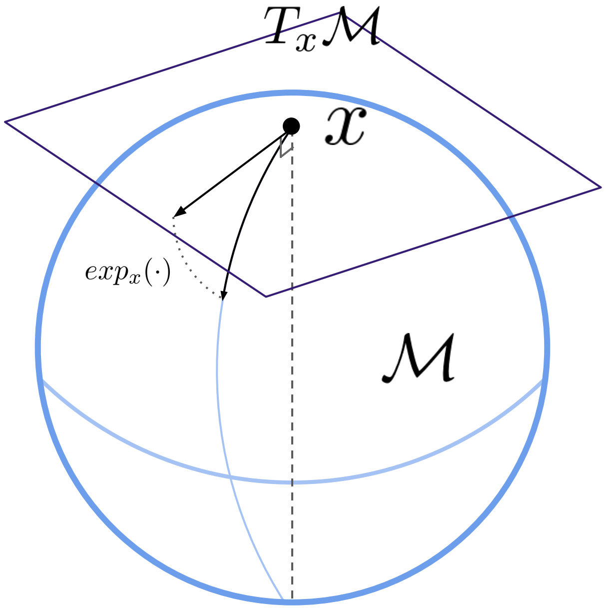
$$\mathcal{E} = \left(\frac{2}{1-\|x\|^2}\right)^2$$

Riemannian metric - a metric tensor that assigns positive values for all input. Constrain Euclidean metric tensor above with properties of a given manifold $\mathcal{E}$.

$$g_x(A,B) > 0 \ \forall x \in \mathbb{R}^+ \qquad g_x(A,B) = \mathcal{E} g^E$$

$$g_x = \left(\frac{2}{1-\|x\|^2}\right)^2 g^E$$

Retractions



Retraction - continuous mapping from a space onto a subspace that preserves position on the subspace.

Exponential map - **locally** defined retraction function that maps a point on a tangent space $T_x\mathcal{M}$ onto the Riemannian manifold $\mathcal{M}$

Riemannian Stochastic Gradient Descent (RSGD)

1. Compute **Euclidean gradient** ∇_E
2. Compute the **Riemannian gradient** ∇_R
   a. Project ∇_E onto the tangent space at the current parameter
      i. Projection is defined using the inner product corresponding to the manifold
3. Compute gradient step using exponential map of the Riemannian gradient [7]

$$\nabla_{\mathcal{R}} = g_x^{-1}(A, B)\nabla_E \mathcal{L}$$

$$\theta_{t+1} = exp_{\theta_t}(-\eta \nabla_R)$$

Hyperbolic Embeddings

Loss function

$$\nabla_E = \frac{\partial \mathcal{L}(\theta_t)}{\partial d(\theta_t, x)} \underbrace{\frac{\partial d(\theta_t, x)}{\partial \theta_t}}_{\text{Partial derivative of Poincaré distance}}$$

Euclidean gradient

Further Reading

1. Hilbert, David. The foundations of geometry. Open court publishing Company, 1902.
2. Greenberg, Marvin J. *Euclidean and non-Euclidean geometries: Development and history*. Macmillan, 1993.
3. Rogness, Jonathan, and Nathan Dunfield. “Interactive Gallery Of Quadric Surfaces.” IGQS: Hyperbolic Paraboloid, 2016, faculty.math.illinois.edu/~nmd/quadrics/hypparab.html.
4. Rogness, Jonathan, and Nathan Dunfield. “Interactive Gallery Of Quadric Surfaces.” IGQS: Hyperbolic Paraboloid, 2016, faculty.math.illinois.edu/~nmd/quadrics/hypparab.html.
5. De Sa, Christopher, et al. "Representation tradeoffs for hyperbolic embeddings." arXiv preprint arXiv:1804.03329 (2018).
6. Wilson, Benjamin, and Matthias Leimeister. "Gradient descent in hyperbolic space." arXiv preprint arXiv:1805.08207 (2018).
7. Bonnabel, Silvere. "Stochastic gradient descent on Riemannian manifolds." IEEE Transactions on Automatic Control 58.9 (2013): 2217-2229.
8. Nickel, Maximillian, and Douwe Kiela. "Poincaré embeddings for learning hierarchical representations." Advances in neural information processing systems. 2017.
9. Le, Matt, et al. "Inferring Concept Hierarchies from Text Corpora via Hyperbolic Embeddings." arXiv preprint arXiv:1902.00913 (2019).
10. Ganea, Octavian, Gary Bécigneul, and Thomas Hofmann. "Hyperbolic neural networks." Advances in Neural Information Processing Systems. 2018.
11. Tifrea, Alexandru, Gary Bécigneul, and Octavian-Eugen Ganea. "Poincar\'e GloVe: Hyperbolic Word Embeddings." arXiv preprint arXiv:1810.06546 (2018).
12. Nickel, Maximilian, and Douwe Kiela. "Learning continuous hierarchies in the lorentz model of hyperbolic geometry." *arXiv preprint arXiv:1806.03417* (2018).