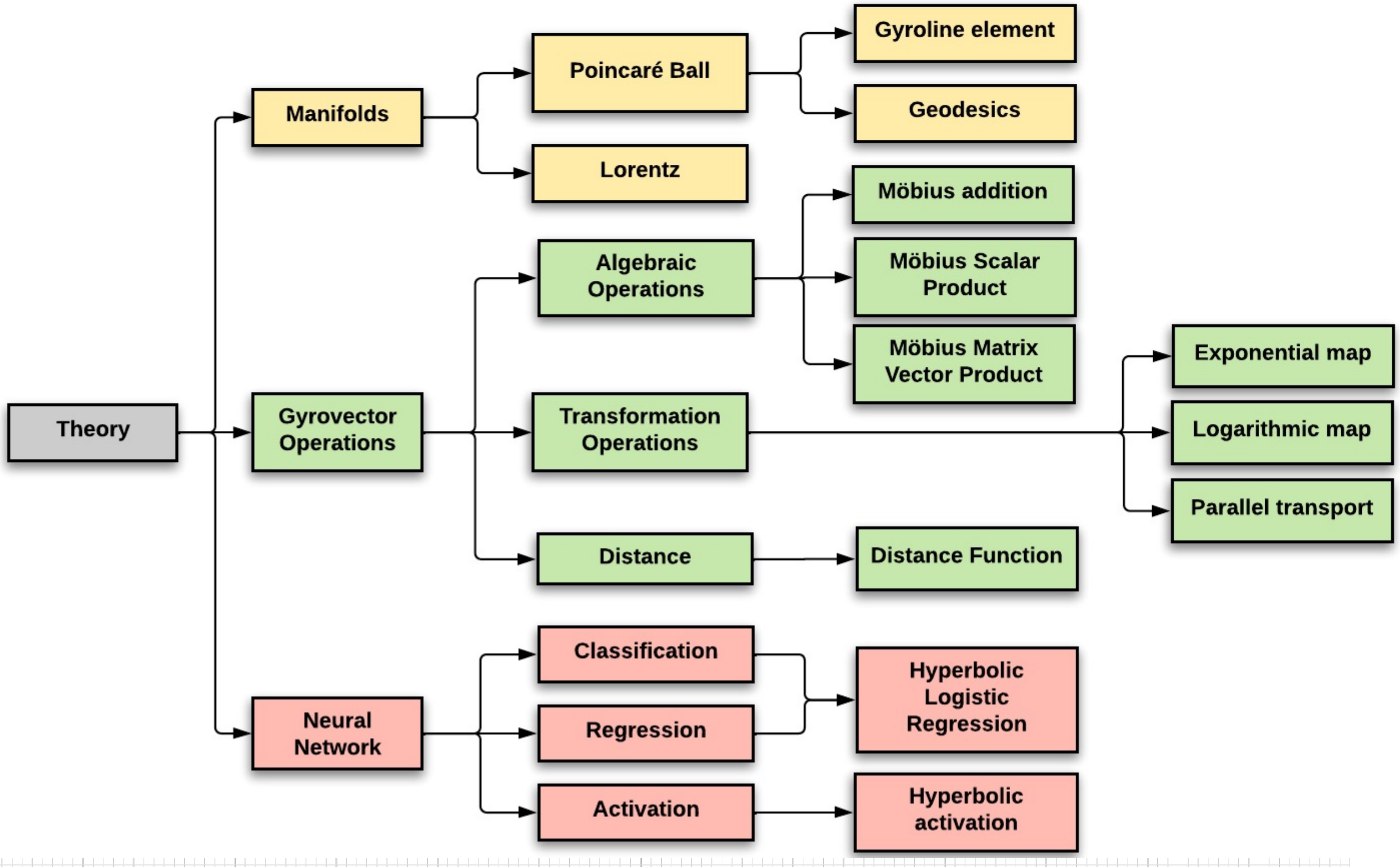
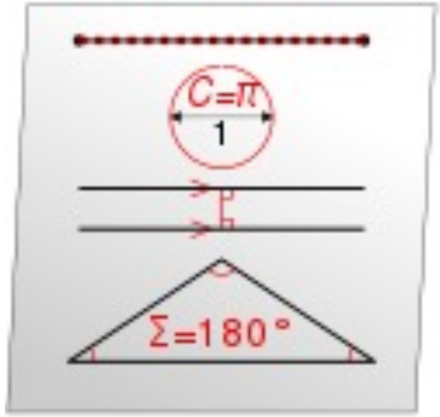
Part 2: Theory



What is Hyperbolic Space?

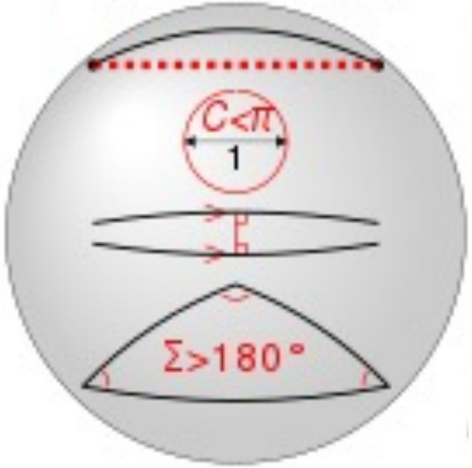
Euclidean vs Elliptic vs Hyperbolic

Euclidean*
(Curvature = 0)



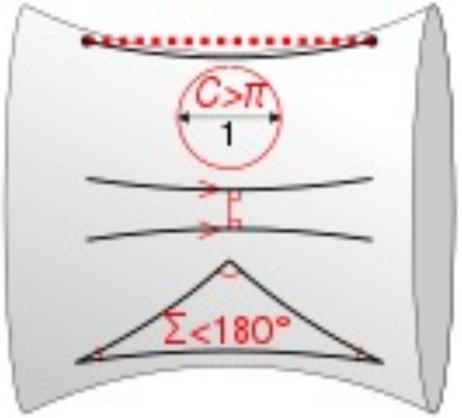
- Circumference/Diameter $= \pi$
- One parallel line
- Sum of angles in a $\Delta = 180°$

Elliptic*
(Curvature > 0)



- Circumference/Diameter $< \pi$
- No Parallel lines.
- Sum of angles in a $\Delta > 180°$

Hyperbolic*
(Curvature < 0)

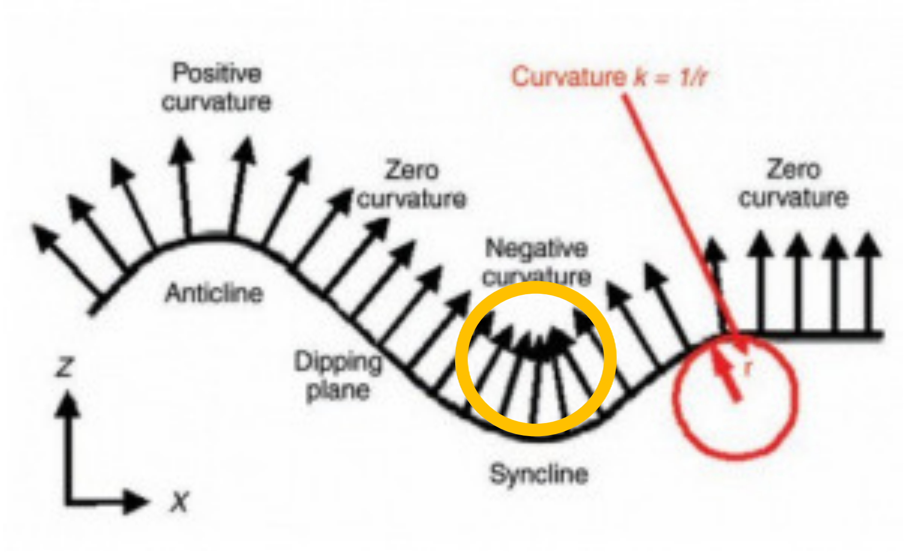


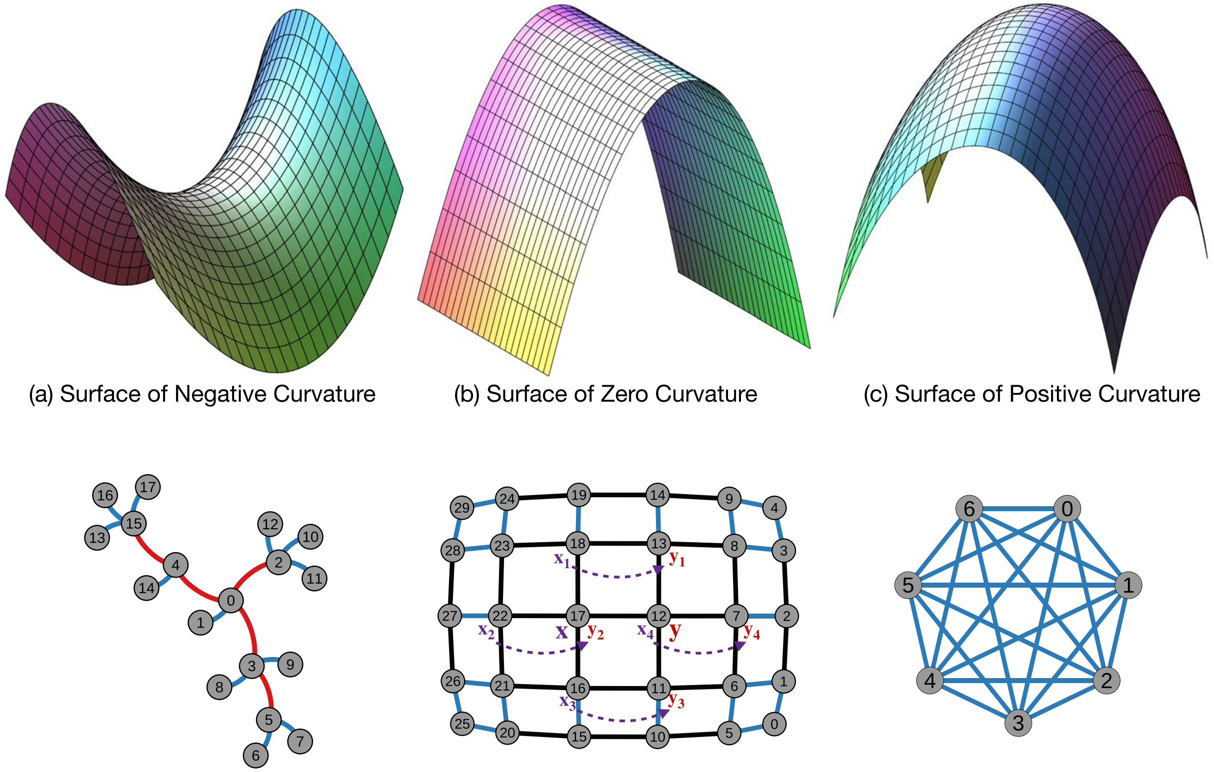
- Circumference/Diameter $> \pi$
- Infinitely many parallel lines
- Sum of angles in a $\Delta < 180°$

An n-dimensional hyperbolic space is an n-dimensional complete Riemannian manifold with a constant negative sectional curvature.

* https://thereaderwiki.com/en/Hyperbolic_geometry

Curvature





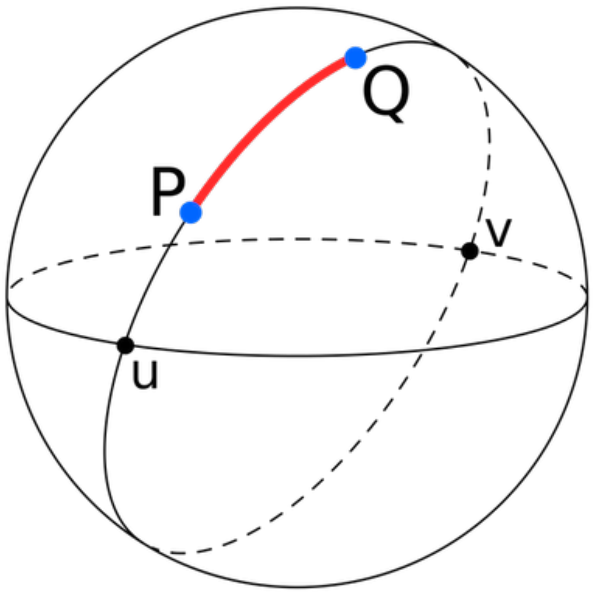
(a) Surface of Negative Curvature (b) Surface of Zero Curvature (c) Surface of Positive Curvature

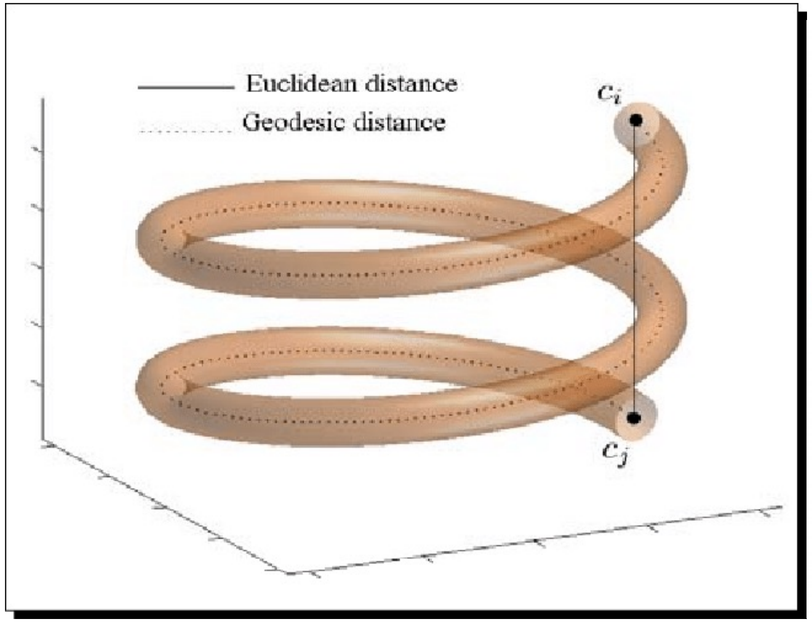
(d) Negative Curvature (e) Zero Curvature (f) Positive Curvature

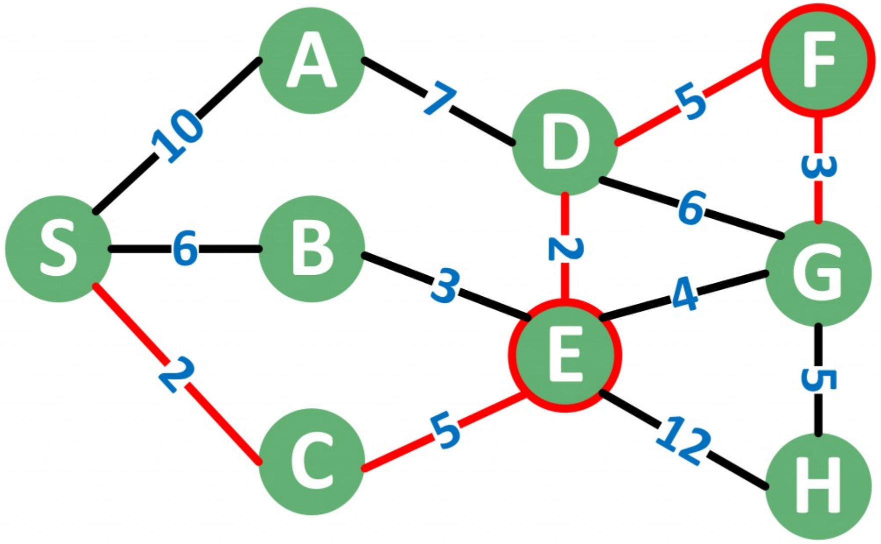
From *Community Detection on Networks with Ricci Flow*, 2019

Given a dataset, estimate curvature.
Learn suitable curvature for a given task and dataset.
Measures: Gromov Hyperbolicity, Ricci Curvature

Geodesics







Wikipedia
https://www.researchgate.net/figure/Euclidean-vs-geodesic-distance-on-a-nonlinear-manifold_fig3_2894469
https://www.baeldung.com/cs/shortest-path-to-nodes-graph

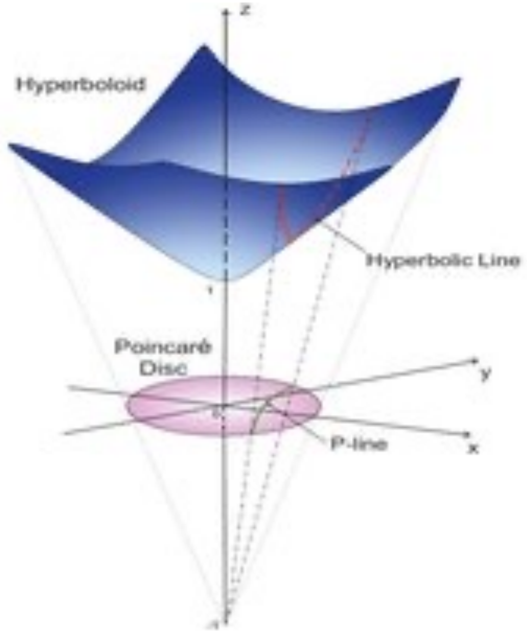
Geodesic: Some kind of shortest distance between two points on the manifold. Applies to graphs too.

Geodesics are generalization of straight lines in Euclidean geometry.
eodesics are constant velocity curves which are locally distance minimizin

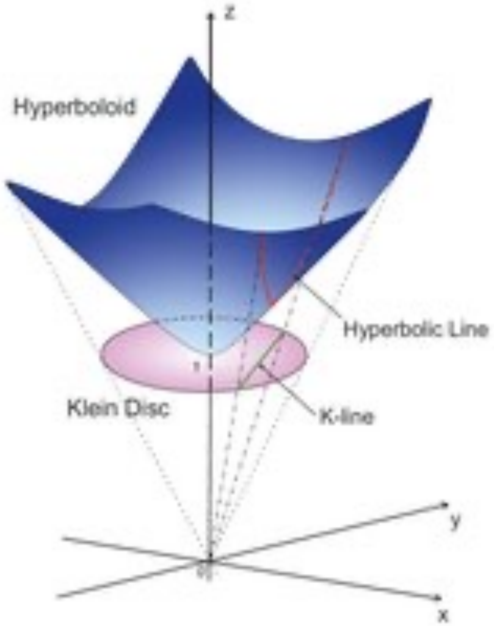
What is Hyperbolic Space?

Isometric models

Defining operations on the large hyperbolic space is difficult. Different projections of the Hyperbolic space are used for different purposes. The projections are isometric to each other, implying, both satisfy the negative curvature property of hyperbolic space.



Poincaré Disk*
Projection on circular disk at origin
Point of projection: (0,0,-1)



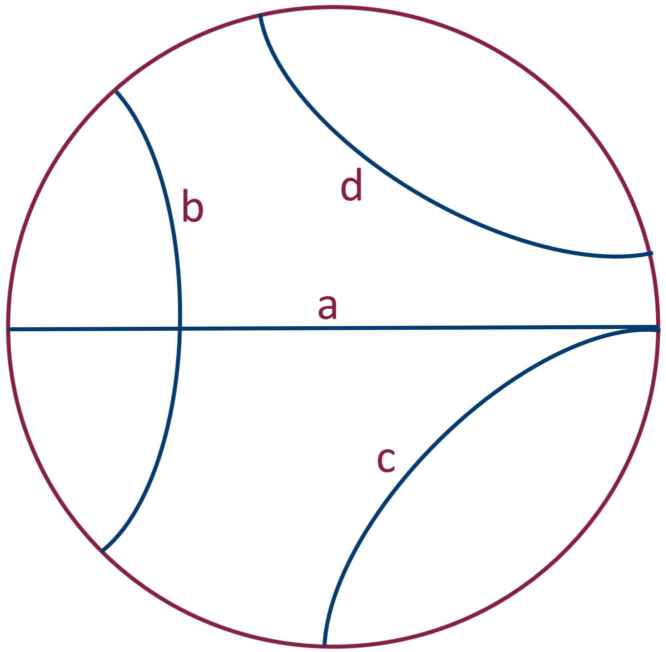
Beltrami-Klein Model*
Projection on circular disk at (0,0,1)
Point of projection: Origin

Poincaré Disk and Klein Disk are isometric representations of hyperboloid sheet.

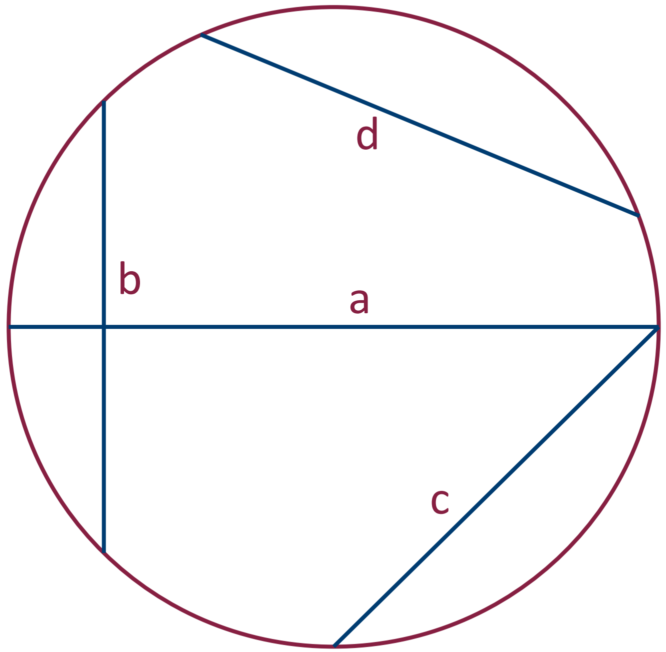
* W. Peng, T. Varanka, A. Mostafa, H. Shi and G. Zhao, "Hyperbolic Deep Neural Networks: A Survey," in *IEEE Transactions on Pattern Analysis and Machine Intelligence.*

What is Hyperbolic Space?

Isometric Models



Poincaré Disk
Preserves Angles
Purpose: Good for capturing hierarchy



Beltrami-Klein Model
Preserves Convexity
Purpose: Good for aggregation of values (Attention models)

Use a representation appropriate for the task: visualization, aggregation, etc.

Development of Neural Networks

Requirements

The basic numerical representation in DNNs is vectors/tensors. The underlying operations on these objects are:

- Distances between points and angles between vectors
- Addition (by extension subtraction) of vectors
- Scalar Multiplication of vectors
- Matrix Multiplication and Concatenation
- Applying Function (for activations and transformation)
- SoftMax - Multinomial Logistic Regression (MLR)

To adopt DNNs to a new space, we need to define well-known operations in the new space.

Poincaré Ball Manifold

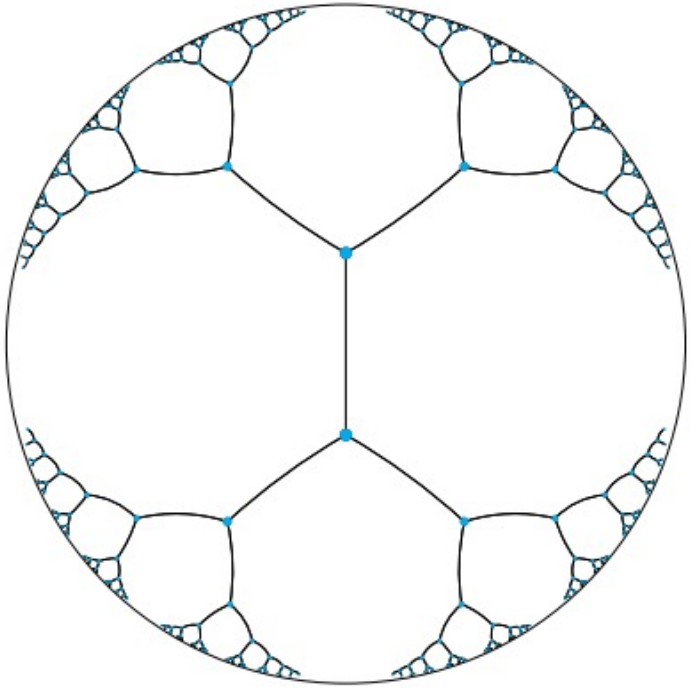
Riemannian Metric

1. Volume of the ball increases exponentially with respect to the radius: Circumference: $2\pi \sinh(r) = O(e^x)$
2. In a hierarchy, the number of nodes increases exponentially with depth.

Riemannian metric helps with the definition of hyperbolic operations; $g^H = \lambda_x^2 g^E, where\ \lambda_x = \frac{2}{1-\|x\|^2},\ and\ g^E = I^n$

Suitable for capturing hierarchies and visualization.

Hyperbolic spaces have exponential capacity.
With *depth ∝ radius*, we can model hierarchy.



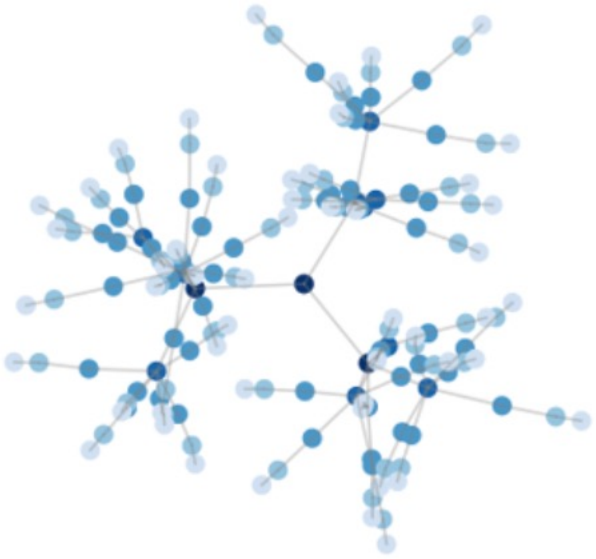
Tree embedded in a Poincaré Disk*
All line segments have the same length

*https://bjlkeng.github.io/posts/hyperbolic-geometry-and-poincare-embeddings/

Poincaré Ball Manifold

Euclidean vs Poincaré Ball Embeddings

DISEASE dataset: Synthetic dataset on disease propagation. The network is trained for Link Prediction. Node Colors represent node depth.



Euclidean Embeddings*
(GCN, DISEASE dataset)

Poincaré Embeddings*
(HGCN, DISEASE dataset)

*Chami, Ines, et al. "Hyperbolic graph convolutional neural networks." Advances in neural information processing systems 32 (2019).

Euclidean Space

Lines, angles, and shortest distance

The Euclidean space is defined by the vector space model.

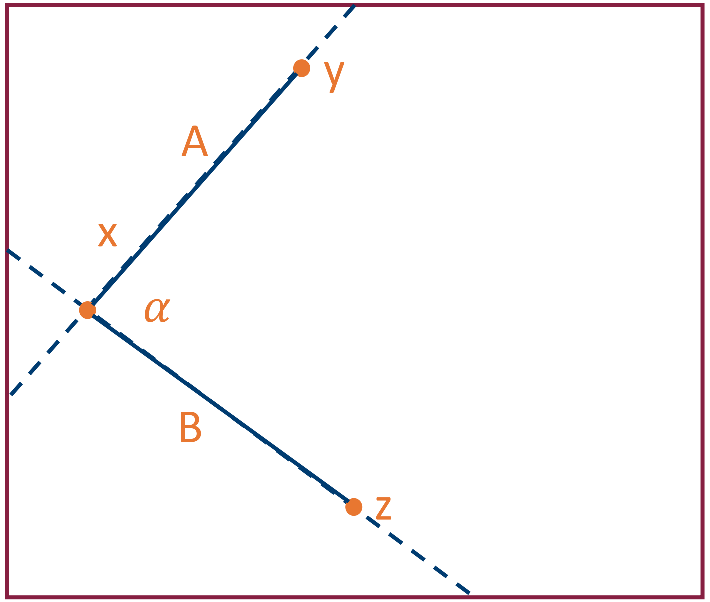
Lines:

$\gamma_A = x + (y - x)t \ (-\infty < t < \infty)$

$\gamma_B = x + (z - x)t \ (-\infty < t < \infty)$

Angles: Angle between vectors.

$$\cos\alpha = \left\langle \frac{y - x}{\|y - x\|}, \frac{z - x}{\|z - x\|} \right\rangle$$



Vectors:

$\gamma_{x \to y} = y - x$

$\gamma_{x \to z} = z - x$

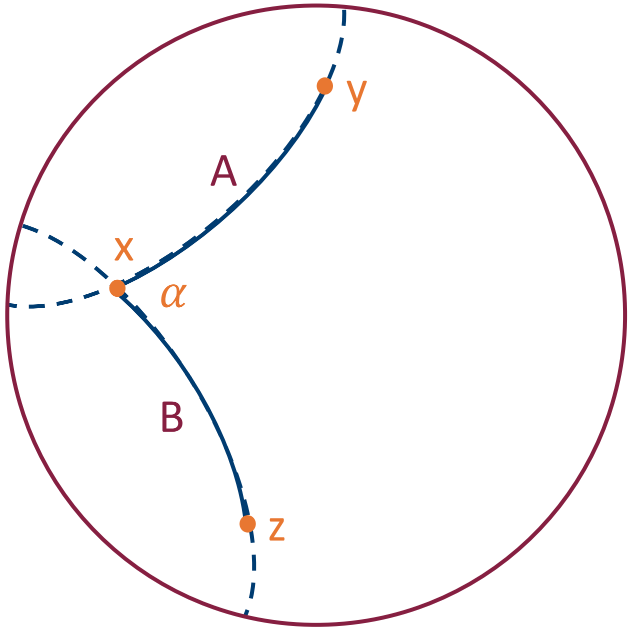
Geodesic length:

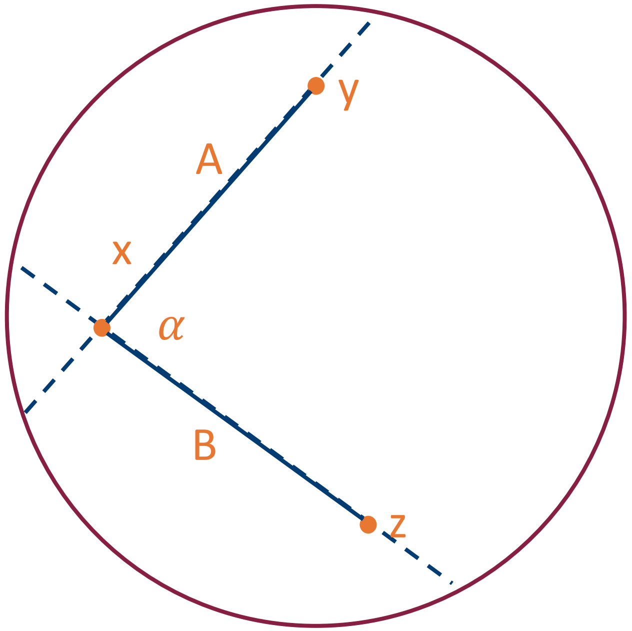
$A(x,y) = \|y - x\|$

$B(x,z) = \|z - x\|$

Hyperbolic Space

Gyrolines in Poincaré Ball and Klein Model





Gyrolines: Line in the gyrovector space.

$$\gamma_A = x \oplus (\ominus\, x \oplus y) \otimes t \quad (-\infty < t < \infty)$$

$$\gamma_B = x \oplus (\ominus\, x \oplus z) \otimes t \quad (-\infty < t < \infty)$$

Gyrolines are algebraic counterparts of Geodesics.
What are the closed form expressions?

Hyperbolic Space

Gyroline and Geodesic Distance

The Euclidean geometry are defined in the vector space model,
For hyperbolic geometry on Poincaré ball, we rely on the gyrovector space model.

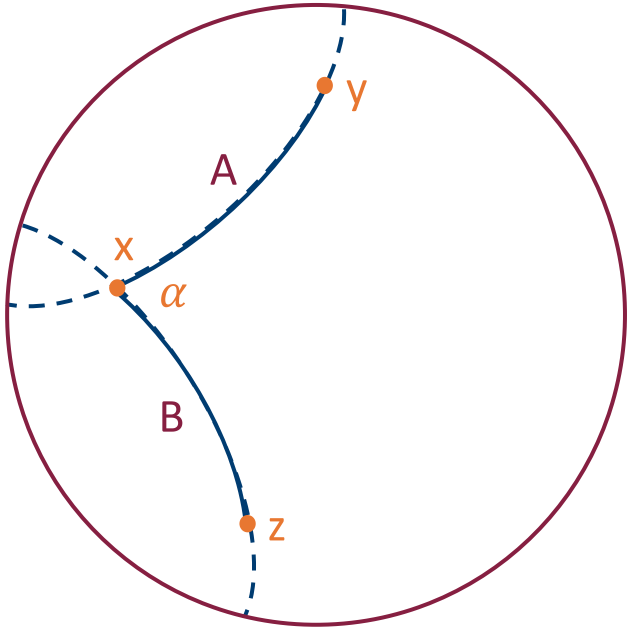
Gyrolines: Line in the gyrovector space.

$$\gamma_A = x \oplus (\ominus x \oplus y) \otimes t \quad (-\infty < t < \infty)$$

$$\gamma_B = x \oplus (\ominus x \oplus z) \otimes t \quad (-\infty < t < \infty)$$

Angles: Angle between segments.

$$\cos\alpha = \left\langle \frac{\ominus x \oplus y}{\| \ominus x \oplus y \|}, \frac{\ominus x \oplus z}{\| \ominus x \oplus z \|} \right\rangle$$



Segment (solid): Vector in gyrovector space.

$$\gamma_{x \to y} = \ominus x \oplus y$$

$$\gamma_{x \to z} = \ominus x \oplus z$$

Geodesic length: Length of vector/Shortest distance between two points in gyrovector space.

$$A(x,y) = \| \ominus x \oplus y \|$$

$$B(x,z) = \| \ominus x \oplus z \|$$

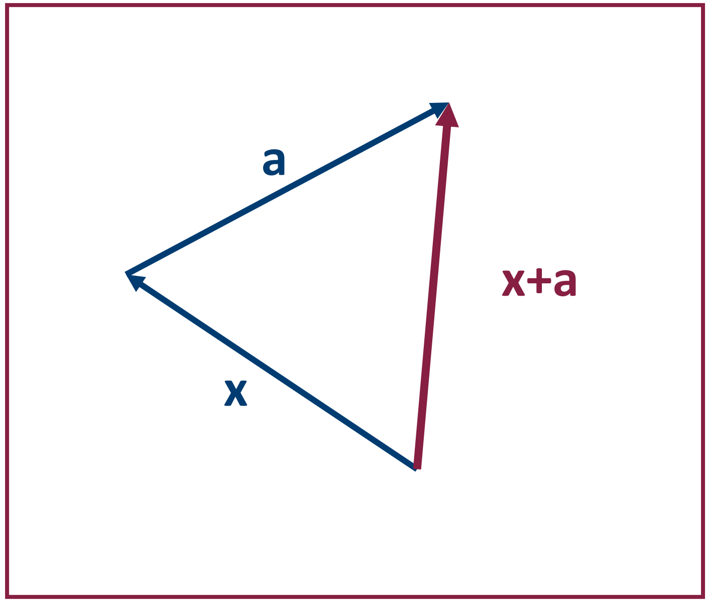
Gyrovector space is similar in notation to vector space.
Not Commutative or associative in the usual sense.

Poincaré Ball Manifold

Euclidean Addition

Euclidean Addition $x{+}a$, can be defined as the translation of **vector x** by **vector a**.

$$x + a = (x_1{+}a_1,\ x_2{+}a_2)$$



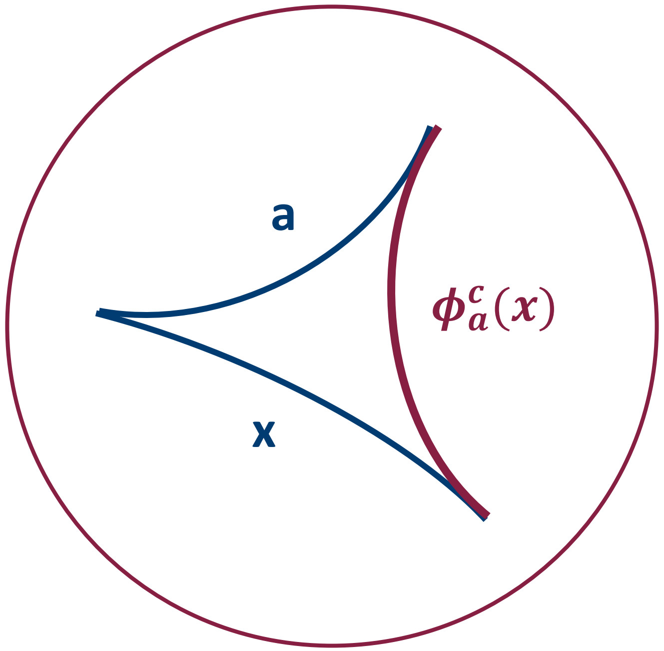
Note: A translation vector is a type of transformation that moves a figure in the coordinate plane from one location to another.

Poincaré Ball Manifold

Gyro addition through Möbius Transformations

Such translation in Poincaré Ball is possible through Möbius Transformation;

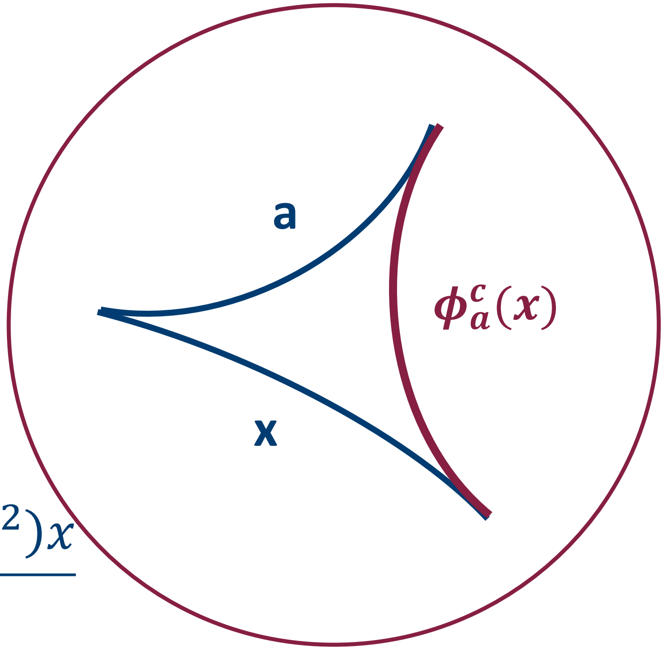
$$f(a, x) = \phi_a^c(x)$$



Poincaré Ball Manifold

Gyro addition through Möbius Transformations

Such translation in hyperbolic space is possible through Möbius Transformation which for the Poincaré ball is;



$$f(a, x) = \phi_a^c(x)$$

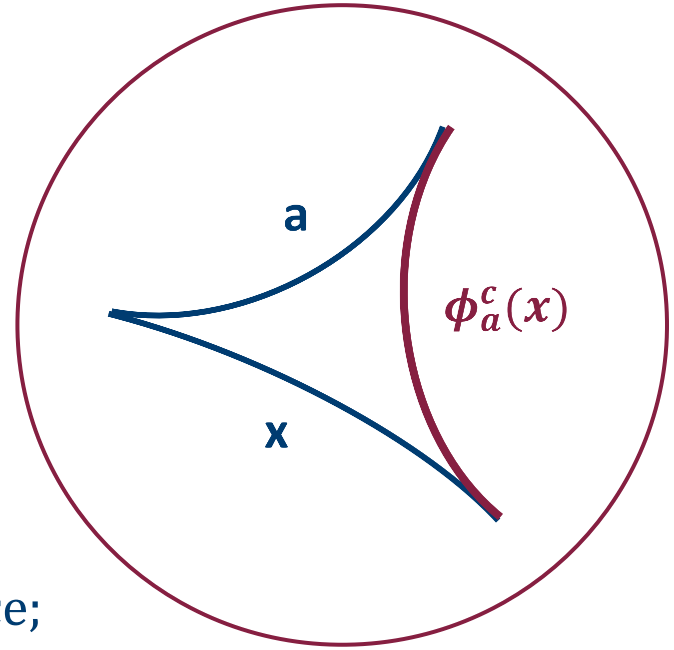
$$\phi_a^c(x) = \frac{x + ca}{c\tilde{a}x + 1} = \frac{(1 + 2ca \cdot x + c \parallel x \parallel^2)a + (1 - c \parallel a \parallel^2)x}{1 + 2ca \cdot x + c^2 \parallel a \parallel^2 \parallel x \parallel^2}$$

Poincaré Ball Manifold

Gyro addition through Möbius Transformations

Möbius Addition:

$$x \oplus_c a = \frac{(1 + 2ca \cdot x + c \parallel x \parallel^2)a + (1 - c \parallel a \parallel^2)x}{1 + 2ca \cdot x + c^2 \parallel a \parallel^2 \parallel x \parallel^2}$$



Similarly, Möbius subtraction is handled as;

$$x \ominus_c a = x \oplus_c (-a)$$

Note that zero-curvature gives back the Euclidean space;

$$a \oplus_0 x = x + a$$

$$x \ominus_0 a = x - a$$

Poincaré Ball Manifold

Möbius scalar multiplication

Möbius scalar product:

$$r \otimes_c x = x \oplus_c x \oplus_c \dots \oplus_c x \text{ (r times)}$$
$$= \frac{1}{\sqrt{c}} \tanh\left(r \tanh^{-1}(\sqrt{c} \parallel x \parallel)\right) \frac{x}{\|x\|}$$

Note that zero-curvature gives back the Euclidean space;

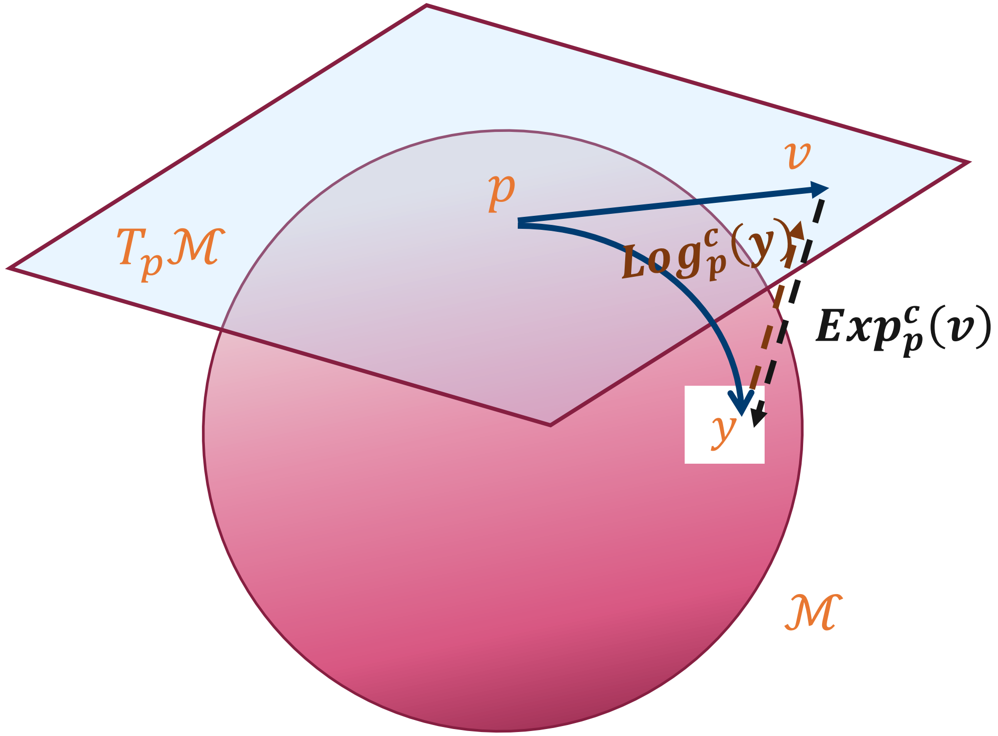
$$\lim_{c \to 0} r \otimes_c x = rx$$

Möbius operations provide the basis for vector addition and scalar multiplication on Poincaré Ball.
How do we perform such operations on a general Riemannian manifold?
A general Riemannian manifold does not have a coordinate representation.

Poincaré Ball Manifold
Transformations: Local Tangent Space

For defining operations in hyperbolic space, we utilize the tangent plane which consists of velocity vectors and is Euclidean.

To move between a Poincaré ball manifold $\mathcal{M}$ and local tangent plane $T_p\mathcal{M}$ at point p,
we use the logarithmic map $\mathrm{Log}_p(y) : \mathcal{M} \to T_p\mathcal{M}$ and
exponential map $\mathrm{Exp}_p(v) : T_p\mathcal{M} \to \mathcal{M}$



$$Log_p^c(y) = \frac{2}{\sqrt{c}\lambda_p^c} \tanh^{-1}(\sqrt{c}\| -p \oplus_c y \|) \frac{-p \oplus_c y}{\| -p \oplus_c y \|}$$

$$Exp_p^c(v) = p \oplus_c \left(\tanh\left(\sqrt{c} \frac{\lambda_x^c \| v \|}{2} \right) \frac{v}{\sqrt{c} \| v \|} \right)$$

In $\mathbb{R}^n$, $Exp_p(v) = p + v$

Tangent space is Euclidean and admits coordinate representation. Tangent space is different from the manifold.

Poincaré Ball Manifold

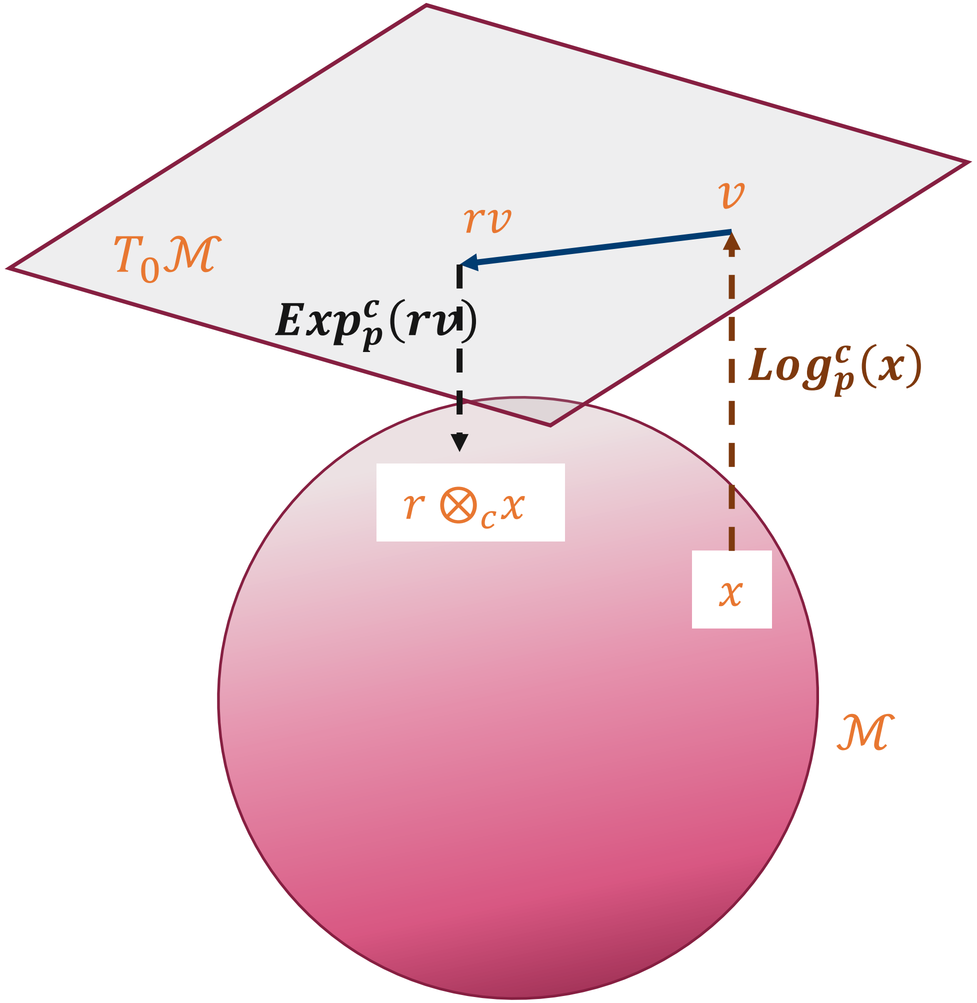
Transformations: Local Tangent Space

Note that, tangent plane behaves like a Euclidean, so a Reformulation of Möbius scalar multiplication is valid:

$$r \otimes_c x = Exp_0^c\left(rLog_0^c(x)\right)$$

Möbius matrix-vector product;

$$[M_1 \quad M_2]\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = M_1 \otimes_c x_1 \oplus_c M_2 \otimes_c x_2$$



Poincaré Ball Manifold

Neural Networks: Activation Function

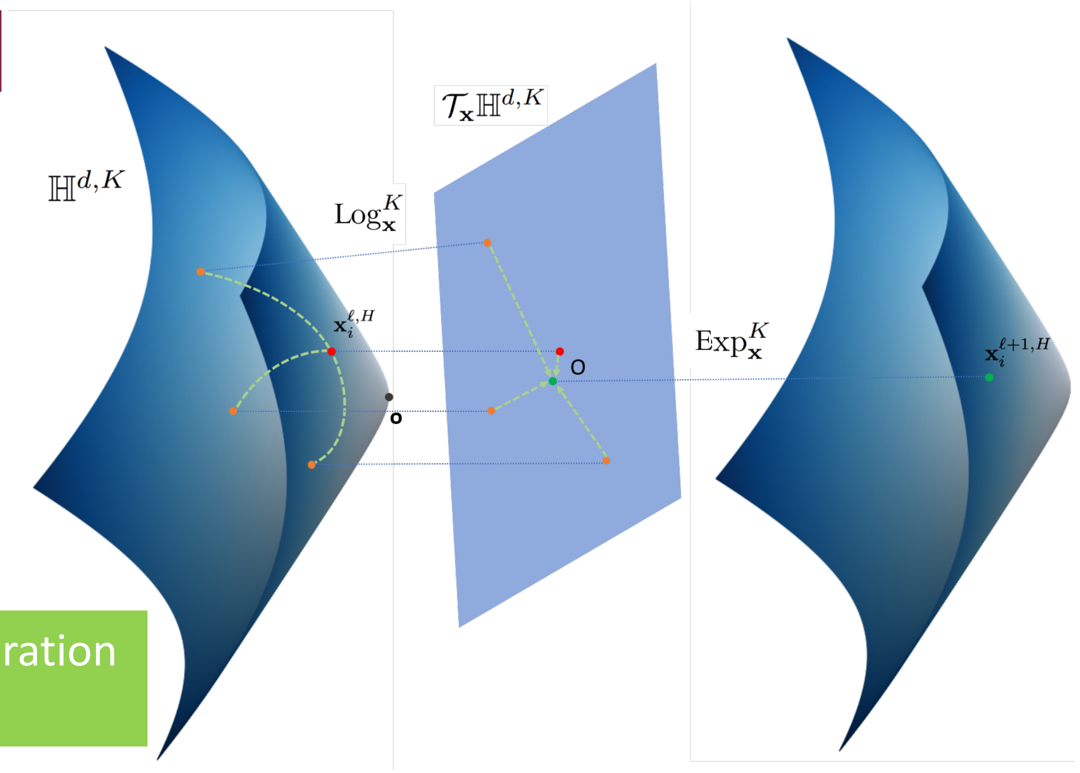
Using a Euclidean non-linear activation $f(x): \mathbb{H}^d \rightarrow \mathbb{H}^d$ in hyperbolic space;

Map to tangent space, so Euclidean functions are valid

Apply Euclidean function to the mapped input

Map the output back to the Poincaré Ball

$$f^{\otimes c}(x) = Exp_0^c(f(Log_0^c(x)))$$



General Recipe: Map to tangent space, perform the operation like aggregation, etc. and map back to manifold.

Image Source: Chami, Ines, et al. "Hyperbolic graph convolutional neural networks." Advances in neural information processing systems 32 (2019).

Poincaré Ball Manifold

Neural Networks: Distance Function

We can use the mapping function to compute hyperbolic distances too;

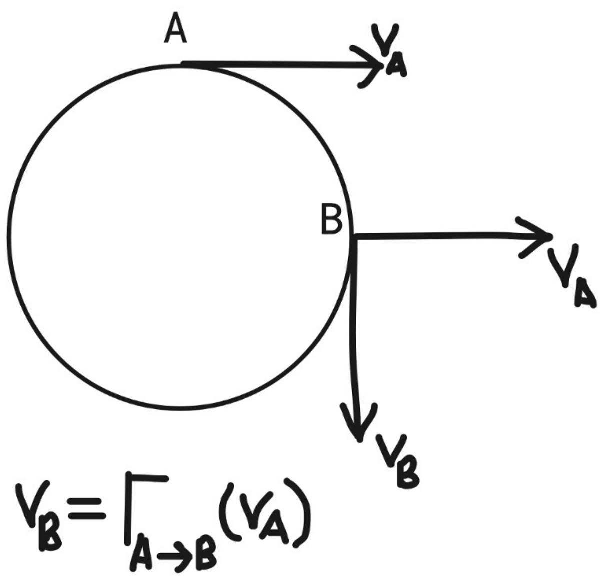
$$d_{\mathbb{R}}(x, y) = \| y - x \|$$

$$d_{\mathbb{H}}(x, y) = Exp_0^c(\| -x \oplus_c y \|)$$

$$d_{\mathbb{H}}(x, y) = \frac{2}{\sqrt{c}} \tanh^{-1}(\sqrt{c} \| -x \oplus_c y \|)$$

Poincaré Ball Manifold

Neural Networks: Parallel Transport



https://miro.medium.com/max/1400/1*iB5lwZrALNHssRYtRdO2wA.png

How do we move around hyperbolic space?

Poincaré Ball Manifold

Neural Networks: Parallel Transport

To move a point from one tangent plane to another, we use Parallel Transport.

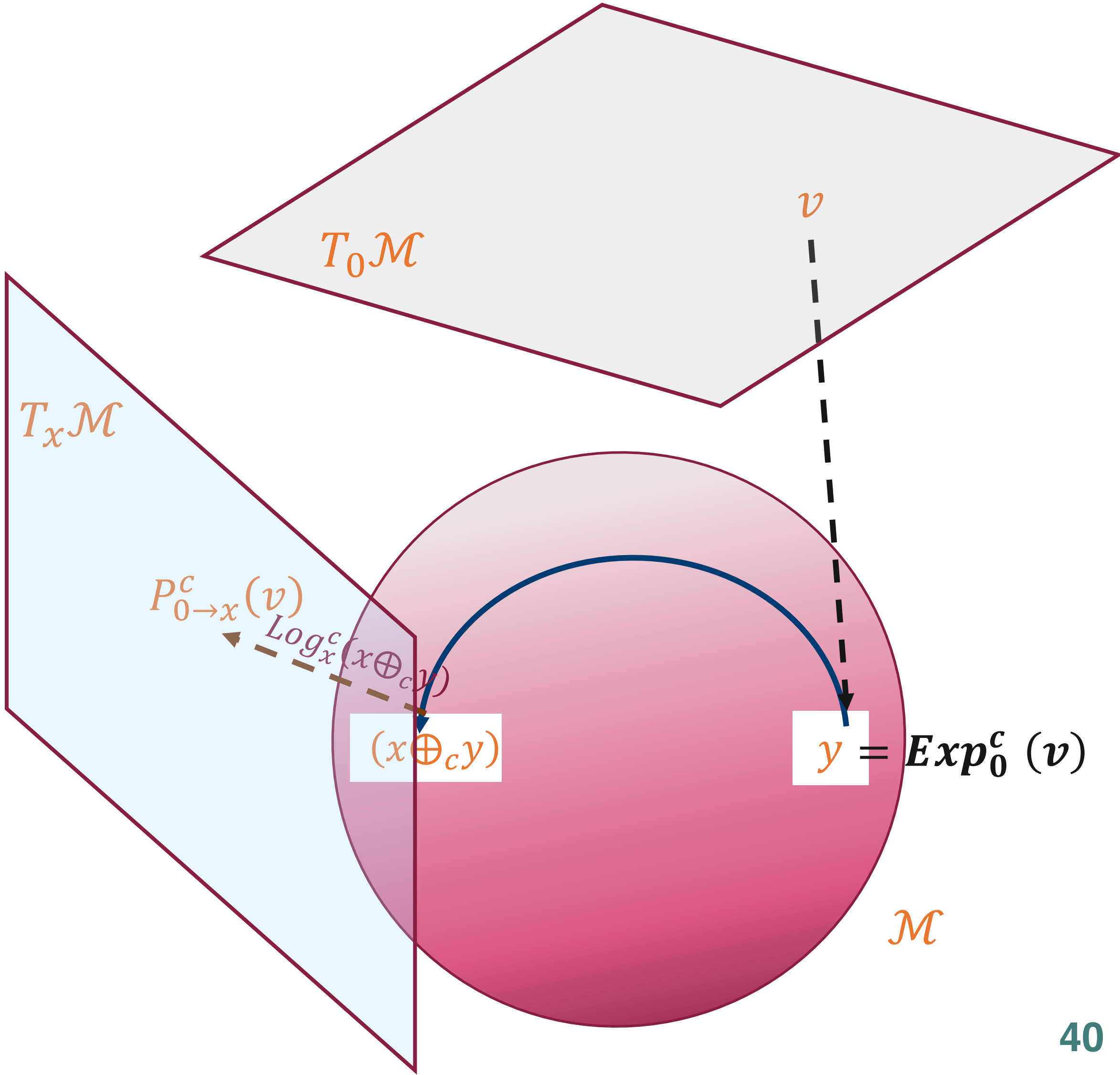
Parallel Transport is the movement of a tangent vector at y to a tangent vector at x along the geodesic between y and x.

$$P^c_{0\to x}(v) = Log^c_x\left(x \oplus_c Exp^c_0(v)\right) = \frac{\lambda^c_0}{\lambda^c_x} v$$

In Euclidean space, parallel transport is identity transformation because the space is flat.

The common application of PT is for *hyperbolic* bias addition.

$$x \oplus b = Exp_x\left(P_{0\to x}\left(Log_0\ (b)\right)\right)$$



Poincaré Ball Manifold

Multiclass Logistic Regression

Euclidean MLR: $P(y = k|x) \propto exp(\langle a_k, x\rangle - b_k)$

Let us say for a class k, the Euclidean separator hyperplane is defined as $\mathbb{E}_k: \langle a_k, x\rangle - b_k = 0$

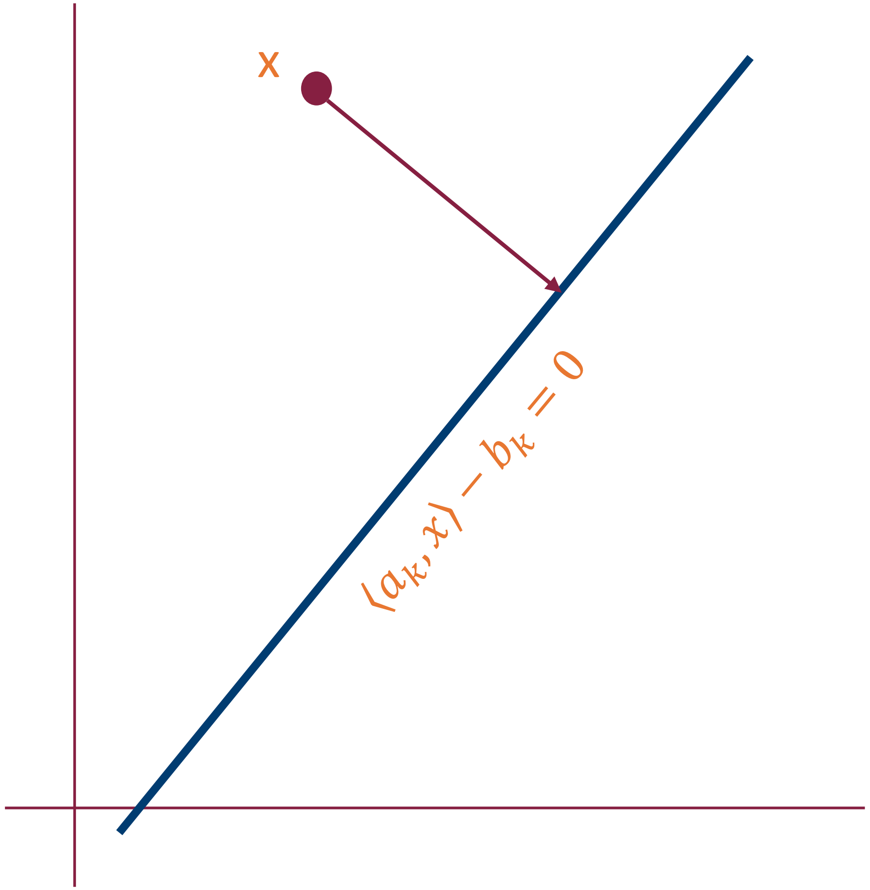
Consequently,

$$P(y = k|x) \propto \exp\left(sign(\langle a_k, x\rangle - b_k)\|a_k\|d_{\mathbb{R}}(x, \mathbb{E}_k)\right)$$

Half space(+/-)

Perpendicular Magnitude

Distance from Hyperplane



Poincaré Ball Manifold

Multiclass Logistic Regression

Euclidean MLR:

Replacing $b_k = a_k p_k$; $\mathbb{E}_k : \langle a_k, -p_k + x \rangle = 0$

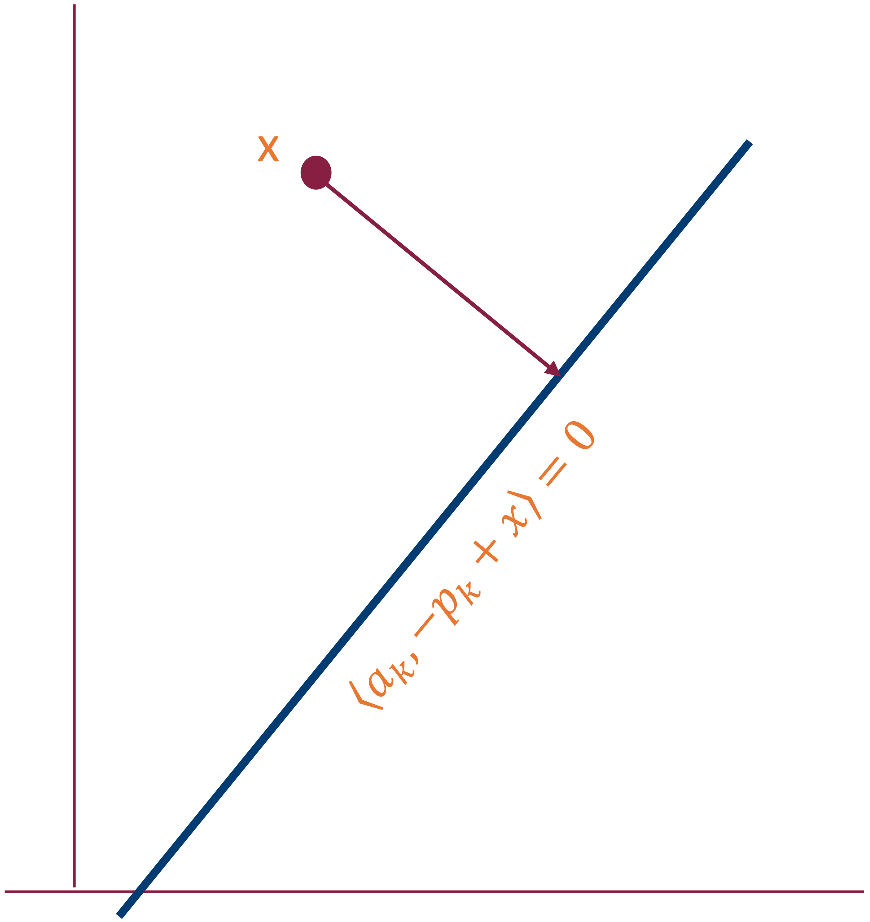
Consequently,

$$P(y = k|x) \propto exp\big(sign(\langle a_k, -p_k + x \rangle)\|a_k\| d_{\mathbb{R}}(x, \mathbb{E}_k)\big)$$

Half space (+/-)

Perpendicular Magnitude

Distance from Hyperplane



MLR in Euclidean space is defined by hyperplanes.
Probability depends on distance from the hyperplane and side.

Poincaré Ball Manifold

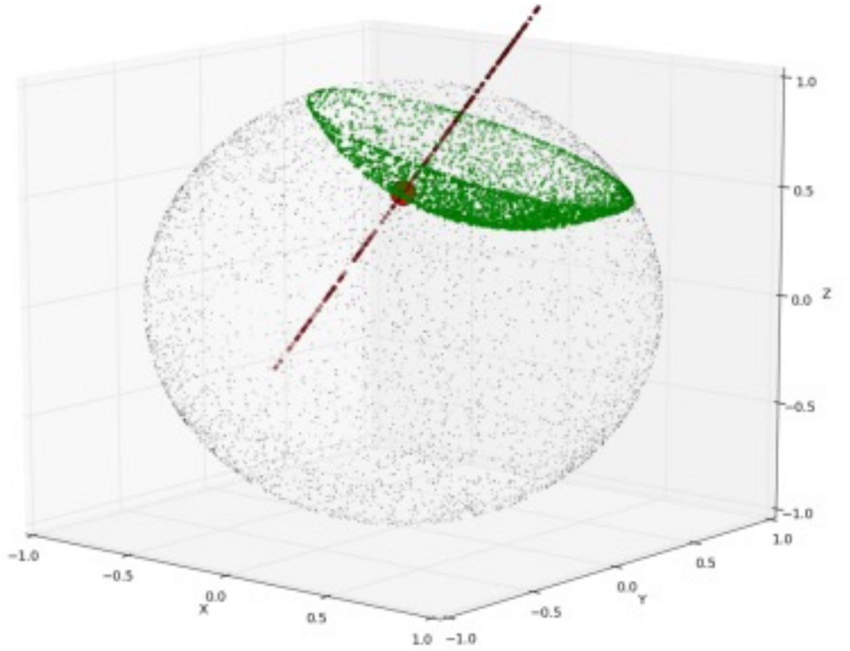
Multiclass Logistic Regression

Hyperbolic MLR:

The hyperbolic equivalent of the hyperplane; $\mathbb{H}_k^c : \langle -p_k \oplus_c x, a_k \rangle = 0$
Consequently,

$$P(y = k|x) \propto exp\left(sign(\langle -p_k \oplus_c x, a_k \rangle) \sqrt{g_{p_k}^c(a_k, a_k)} d_{\mathbb{H}}(x, \mathbb{H}_k^c) \right)$$

$$P(y = k|x) \propto exp\left(\frac{\lambda_{p_k}^c \|a_k\|}{\sqrt{c}} \sinh^{-1}\left(\frac{2\sqrt{c}\langle -p_k \oplus_c x, a_k \rangle}{(1 - c\|-p_k \oplus_c x\|^2)\|a_k\|} \right) \right)$$

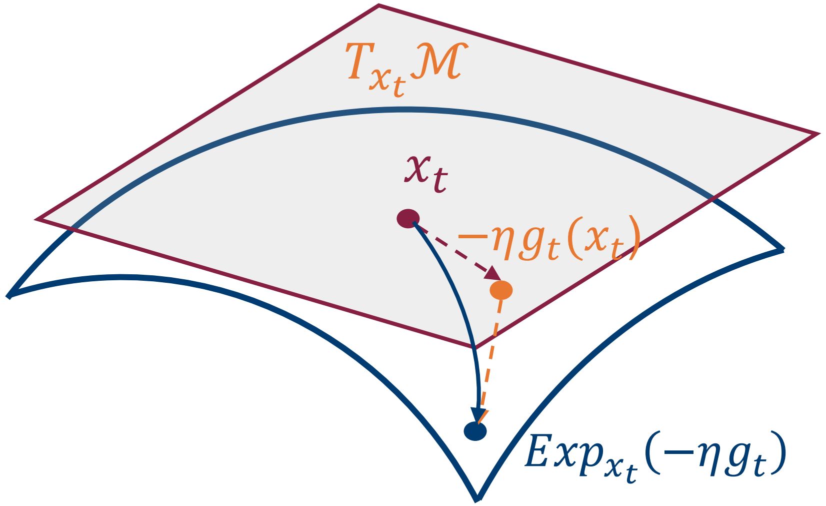


Green point cloud denotes hyperplane in the Hyperbolic Plane.
The red point is p.
The red normal axis to the hyperplane through p is parallel to a.

Image Source: Ganea, Octavian, Gary Bécigneul, and Thomas Hofmann. "Hyperbolic neural networks." Advances in neural information processing systems 31 (2018).

Riemannian Optimization

$$\text{SGD: } x_{t+1} \leftarrow x_t - \alpha g_t$$



RSGD consists of the following three steps at x_t.

1. Evaluate the gradient g_t
2. Project g_t to the tangent space
3. Perform gradient step in the negative direction of the tangent vector

$$\text{RSGD: } x_{t+1} \leftarrow Exp_{x_t}(-\alpha g_t)$$

Approximation: For computational efficiency, we use *retraction map* instead of the *exponential map*.

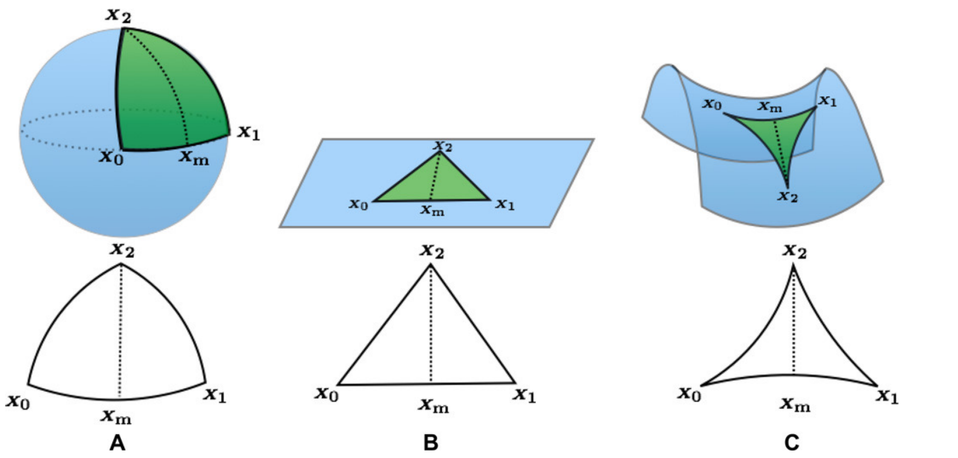
$$x_{t+1} \leftarrow Proj_{\mathcal{M}}(x_t - \alpha g_t)$$

Riemannian AdaGrad: Decompose manifold, $\mathcal{M}$ into cartesian product of n sub-manifolds $\mathcal{M}_k$ and updates are manifold-wise and not coordinate-wise.

$$x_{t+1}^i \leftarrow \mathrm{Exp}_{x_t^i}^i\left(-\alpha g_t^i / \sqrt{\sum_{k=1}^{t} \|g_k^i\|_{x_k^i}^2}\right)$$

Hyperbolicity of Network Data and Clustering

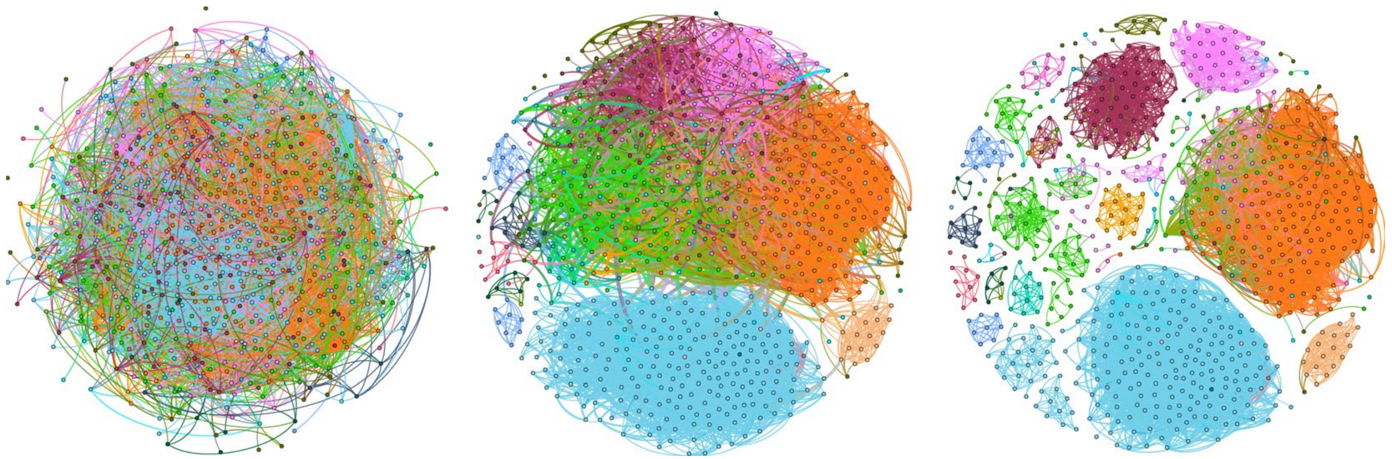
Can we measure hyperbolicity of network data?



From https://web.math.princeton.edu/~ruobingz/research.html

A geodesic metric space (X,d) is called δ-hyperbolic if every geodesic triangle δ-slim, i.e. any point on any of the sides of the triangle is within distance δ of the other two sides. Finally, a metric space is Gromov hyperbolic if it is δ-hyperbolic for some δ ≥ 0.

Ricci tensor is that it describes how much a volume element would differ in curved space compared to Euclidean or flat space.



(a) (b) (c)

(a) A Facebook ego network of one user with 792 friends and 14025 edges generated by Gephi's Fruchterman Reingold layout59. The colors represent 24 different friend circles (communities) hand labeled by the user.
(b) By the Ricci flow process of 20 iterations, the weights of inter-community edges are increased (thick edges in the figure) while the weights of intra-community edges gradually shrink to 0 (thin edges in the figure).
(c) By removing the inter-community edges with high weights, the communities are clearly detected.

Poincaré Ball Manifold

Summary

Attributes	Euclidean	Poincaré Ball
Metric	$g^E = I$(flat)	$\lambda_x^2 g^E$(curved)
Geodesic	Straight Line	Line (or) Arc/Circle
Algebraic structure	Vector Space	Gyrovector Space
Operations	Addition, Subtraction, scalar product, matrix-vector product	Möbius versions
Activation	$f(x)$	$Exp_0^c(f(Log_0^c(x)))$
Distance	$L2 - norm$	$\frac{2}{\sqrt{c}}\tanh^{-1}(\sqrt{c} \parallel -x \oplus_c y \parallel)$
Multiclass Logistic Regression	$exp\left(sign(\langle a_k, -p_k + x\rangle)\|a_k\|d_{\mathbb{R}}(x, \mathbb{E}_k)\right)$	$exp\left(sign(\langle -p_k \oplus_c x, a_k\rangle)\sqrt{g_{p_k}^c(a_k, a_k)}d_{\mathbb{H}}(x, \mathbb{H}_k^c)\right)$
Optimization	$x_{t+1} \leftarrow x_t - \alpha g_t$ (SGD)	$x_{t+1} \leftarrow Exp_{x_t}(-\alpha g_t)$ (RSGD)

Q&A

Please provide your feedback at:

https://bit.ly/feedback-hyperbolic-tutorial