Hyperbolic Deep Learning

Pascal Mettes
University of Amsterdam
What to expect

First hour
Origins, fundamentals, and initial work in hyperbolic deep learning

Second hour
Network layers, computer vision, and more with hyperboles
Origins: Euclid’s postulates

Five postulates as foundation of geometry.

1. Each pair of points can be joined by one and only one straight line segment.
2. Any straight line segment can be indefinitely extended in either direction.
3. There is exactly one circle of any given radius with any given center.
4. All right angles are congruent to one another.
5. If a straight line falling on two straight lines makes the interior angles on the same side less than two right angles, the two straight lines, if extended indefinitely, meet on that side on which the angles are less than two right angles.
Breaking the 5th postulate

Also known as the parallel postulate, but is it needed?

For centuries, mathematicians have tried to distill it from the first four.
Making a new 5th postulate

For any given line R and point P not on R, in the plane containing both line R and point P there are at least two distinct lines through P that do not intersect R.
Spherical and hyperbolic geometry

**Flat**
- Triangle: sum of angles is 180°
- Straightest Possible Path: is a straight line
- Circle: $C = 2\pi$
- Parallel Lines: remain parallel

**Spherical**
- Triangle: sum of angles is greater than 180°
- Straightest Possible Path: is a piece of a great circle
- Circle: $C < 2\pi$
- Parallel Lines: eventually converge

**Hyperbolic**
- Triangle: sum of angles is less than 180°
- Straightest Possible Path: is a piece of a hyperbola
- Circle: $C > 2\pi$
- Parallel Lines: eventually diverge
Geometry and special relativity

Places space and time on equal footing, with velocity linked to hyperbolic angles.

\[ ds^2 = c^2dt - dx^2 - dy^2 - dz^2 < 0 \]
Hyperbolic deep learning
= Efficient deep learning
The different models of hyperbolic geometry

Many models, most used in deep learning: Poincaré ball model and Lorentz model.
The Poincaré ball model

Points inside unit ball
\( \mathbb{D}^n = \{ x \in \mathbb{R}^n : \|x\| < 1 \} \)

Tensor metric
\[ g^\mathbb{D}_x = \lambda_x^2 g^E, \quad \text{where } \lambda_x := \frac{2}{1 - \|x\|^2} \]

Distance between two points:
\[ d_\mathbb{D}(x, y) = \cosh^{-1} \left( 1 + 2 \frac{\|x - y\|^2}{(1 - \|x\|^2)(1 - \|y\|^2)} \right) \]

Möbius addition:
\[ x \oplus_c y := \frac{(1 + 2c(x, y) + c\|y\|^2)x + (1 - c\|x\|^2)y}{1 + 2c(x, y) + c^2\|x\|^2\|y\|^2} \]
From Poincaré to Euclidean and back

\[ \exp_c(v) = \tanh(\sqrt{c\|v\|}) \frac{v}{\sqrt{c\|v\|}} \quad \text{and} \quad \log_c(y) = \tanh^{-1}(\sqrt{c\|y\|}) \frac{y}{\sqrt{c\|y\|}} \]
Initial success: Poincaré embeddings [Nickel and Kiela, NeurIPS 2017]

Input

\( S = \{x_i\}_{i=1}^n \)

Set of symbols

\( \mathcal{D} = \{(u, v)\} \)

Parent-child relations of the symbols

\( \mathcal{N}(u) = \{v' \mid (u, v') \notin \mathcal{D} \} \cup \{v\} \)

Non parent-child relations of the symbols

Output

\( \Theta = \{\theta_i\}_{i=1}^n \)

Hyperbolic embedding of symbols
Poincaré embedding loss

\[ \Theta' \leftarrow \arg \min_{\Theta} \mathcal{L}(\Theta) \]

\[ \text{s.t. } \forall \theta_i \in \Theta : \|\theta_i\| < 1 \]

Find embeddings that minimize some loss

Embeddings should be inside the ball

\[
\mathcal{L}(\Theta) = \sum_{(u,v) \in \mathcal{D}} \log \frac{e^{-d(u,v)}}{\sum_{v' \in \mathcal{N}(u)} e^{-d(u,v')}}
\]

Pull parent-child nodes together, push others away

\[
d(u, v) = \text{arcosh} \left( 1 + 2 \frac{\|u - v\|^2}{(1 - \|u\|^2)(1 - \|v\|^2)} \right)
\]

Hyperbolic distance function
Poincaré embedding optimization

\[ \theta_{t+1} = R_{\theta_t} ( -\eta_t \nabla_R \mathcal{L}(\theta_t) ) \]

Riemannian gradient descent

\[ \theta_{t+1} \leftarrow \text{proj} \left( \theta_t - \eta_t \frac{1 - \|\theta_t\|^2}{4} \nabla_E \right) \]

Descent rule: Euclidean gradient + scaling + projection
Poincaré embeddings: Embedding WordNet

(a) Intermediate embedding after 20 epochs

(b) Embedding after convergence
Why are hyperboles and hierarchies such a good match?
From hierarchies to graphs

Hierarchies are basically graphs, what if we want more than just embedding?

Can we bring hyperbolic geometry to graph networks?

[Kipf and Welling, ICLR 2017]
Intuition behind hyperbolic graph networks

Project hyperbolic nodes to Euclidean space, do normal graph layer, project back

[Chami et al., NeurIPS 2019, Liu et al., NeurIPS 2019]
Hyperbolic Graph Networks

Standard graph layer

\[ h^{k+1}_u = \sigma \left( \sum_{v \in I(u)} \tilde{A}_{uv} W^k h^k_v \right) \]

\[ \tilde{A} = D^{-\frac{1}{2}} (A + I) D^{-\frac{1}{2}} \]

Normalized affinity matrix

Weights to learn

Current node embeddings

Hyperbolic graph layer

\[ h^{k+1}_u = \sigma \left( \exp_x' \left( \sum_{v \in I(u)} \tilde{A}_{uv} W^k \log_x'(h^k_v) \right) \right) \]

Projection steps

\[ \exp_x(v) = x + v \]
\[ \log_x(y) = y - x \]
Hyperbolic graph network examples

(a) GCN layers.  (b) HGCN layers.  (c) GCN (left), HGCN (right).

Improved node and link prediction, especially when using few embedding dimensions.
Hyperbolic graphs without tangents  [Dai et al., CVPR 2021]

Latest attempts avoid the tangent space (but need to switch between hyperbolic models).
Recap

Hyperbolic geometry is a not so old subfield in mathematics.

Hyperboles are curved spaces, especially close to the border.

A lot of initial promise on graph-like data.

After the break: hyperbolic deep learning for vision, text, and more.
Break
Network layers in hyperbolic space

The field is in an ongoing effort to make all network layers hyperbolic.

Hyperbolic variants have been proposed for:
- Convolutional layers
- Recurrent layers
- Activation functions
- Batch normalization
- Classification layers
- ...

Still an ongoing effort: there is no unified solution and library.
Logistic regression in hyperbolic space

Gyroplane with offset and orientation

\[ H^c = \{ z_{ij} \in D^n_c, \langle -p \oplus_c z_{ij}, w \rangle = 0 \} \]

Hyperbolic distance to the gyroplane

\[
d_c(z_{ij}, H_y^c) = \frac{1}{\sqrt{c}} \sinh^{-1} \left( \frac{2\sqrt{c}\langle -p_y \oplus_c z_{ij}, w_y \rangle}{(1-c|| - p_y \oplus_c z_{ij} ||^2)||w_y||} \right)
\]
Hyperbolic network layers give us a hammer for all sorts of machine learning problems, let’s use it!
Hyperbolic embeddings for computer vision

Visual dataset are highly hierarchical and a good match for hyperbolic spaces.

delta-hyperbolicity: 0 is fully hyperbolic, 1 = fully Euclidean

<table>
<thead>
<tr>
<th>Encoder</th>
<th>Dataset</th>
<th>CIFAR10</th>
<th>CIFAR100</th>
<th>CUB</th>
<th>MiniImageNet</th>
</tr>
</thead>
<tbody>
<tr>
<td>Inception v3</td>
<td>CIFAR10</td>
<td>0.25</td>
<td>0.23</td>
<td>0.23</td>
<td>0.21</td>
</tr>
<tr>
<td>ResNet34</td>
<td>CIFAR100</td>
<td>0.26</td>
<td>0.25</td>
<td>0.25</td>
<td>0.21</td>
</tr>
<tr>
<td>VGG19</td>
<td>CUB</td>
<td>0.23</td>
<td>0.22</td>
<td>0.23</td>
<td>0.17</td>
</tr>
<tr>
<td></td>
<td>MiniImageNet</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

[Khrulkov et al. CVPR 2020]
Hyperbolic Image Embeddings  [Khrulkov et al. CVPR 2020]

1. Feed input through Euclidean network and project to Poincaré ball.

2. Average examples from same class on the hyperbole.

3. For new few-shot test example, Select class whose average is closest.
Main idea: what if we *a priori* embed class hierarchies on the hyperbole and afterwards learn to project inputs to the embedded hierarchies?

[Long et al. CVPR 2020]
Searching for actions on the hyperbola [Long et al. CVPR 2020]
Searching for actions on the hyperbole [Long et al. CVPR 2020]

(a) One-hot.  (b) Hyperbolic (ours).
Hierarchies and hyperboles allow us to transfer knowledge from seen to unseen classes.
Can we also benefit from hyperbolic embeddings when hierarchies are not known?

\[ z = \exp_\nu(F(x; \theta)) \]
\[ \exp_\nu(x) = \tanh(||x||/2) \frac{x}{||x||} \]

[Ghadimi et al. NeurIPS 2021]
Hyperbolic Busemann Learning [Ghadimi et al. NeurIPS 2021]

\[
z = \exp_v(\mathcal{F}(x; \theta))
\]

\[
\exp_0(x) = \tanh(||x||/2)^{x/||x||}
\]

Ideal prototype

\[
\ell(z, p) = b_p(z) - \phi(d) \cdot \log(1 - ||z||^2)
\]

Busemann distance function

\[
b_p(z) = \log \frac{||p - z||^2}{(1 - ||z||^2)}
\]

Scaling factor

\[
\phi(d; s) = s \cdot d
\]

Regularization term

This loss is directly linked to logistic regression.
Hyperbolic Busemann Learning [Ghadimi et al. NeurIPS 2021]

The slope balances confidence and inter-class confusion.
Distance to the origin provides a natural measure of uncertainty.
Hyperbolic embeddings and the future [Suris et al. CVPR 2021]

Can you tell what will happen next?
Hyperbolic embeddings and the future [Suris et al. CVPR 2021]

Hyperboles allow us to quantify prediction certainty as more information comes in.
Hyperbolic Image Segmentation [under submission]

Perform per-pixel classification in hyperbolic embedding space.
Hyperbolic Image Segmentation

(a) Hyperbolic uncertainty correlates with boundary distance.
Hyperbolic deep learning for text

[Di et al. 2020]
Hyperbolic deep learning for text

- the national cancer institute ban smoking
- the national cancer institute warns citizens to avoid smoking cigarette
- the national cancer institute claims that smoking cigarette too often would increase the chance of getting lung cancer
- the national cancer institute study the effect of chemical in cigarette on different group of workers.
- the national cancer institute also projected that overall u.s. mortality rates from lung cancer should begin to drop in several years if cigarette smoking continues to decrease
- the national cancer institute report a form of asbestos once used to make cigarette filters has caused a high percentage of cancer deaths among a group of workers exposed to it

[Dai et al. 2020]
Hyperbolic deep learning for text

[Chen et al. ACL 2020]
Poincaré embeddings for modelling hierarchies in cells.
Hyperbolic embeddings benefit recommender systems as well.
Even Generative Adversarial Networks have been made hyperbolic.
The default space in deep learning is Euclidean, but should it be?

Hyperbolic geometry often matches (latent) data structures better.

A lot of recent work to make hyperbolic geometry work with deep learning.

Perhaps your research problem can also benefit from hyperboles?
Thank you