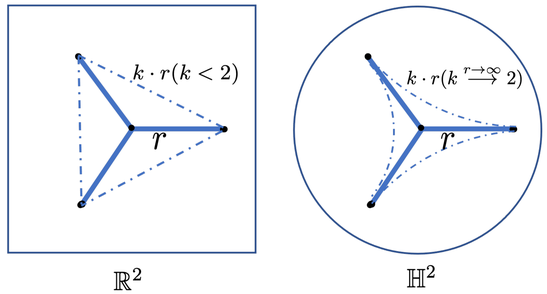
A Numerical Comparison between Lorentz and Poincaré Model for Representation Learning

Gal Mishne[1], Zhengchao Wan[1], Yusu Wang[1], Sheng Yang[2]
[1]University of California - San Diego, [2]Harvard University

Introduction

- Distances grow exponentially in hyperbolic spaces $\mathbb{H}^n$
- Finite trees can be embeded into $\mathbb{H}^n$ with arbitrarily low distortion[1]
- Motivates hyperbolic embeddings for hierarchical datasets such as words and images
- **Concern**: highly numerically unstable; mysteriously better performance of Lorentz than Poincaré despite being isometric.

Hyperbolic Distances are more Tree-like



Main Contribution

- Identify the source of numerical instability in two popular isometric Hyperbolic models, Lorentz $\mathbb{L}^n$and Poincare $\mathbb{B}^n$
- Clarify comparative advantage of $\mathbb{L}^n$ and $\mathbb{B}^n$
  - In **representation**, $\mathbb{B}^n$ has a larger diameter than $\mathbb{L}^n$
  - In **optimization**, $\mathbb{L}^n$ has a larger correct update region
- Demonstrate empirical **superiority of $\mathbb{L}^n$ in tree-embedding related tasks**

Visualization of Poincaré and Lorentz model

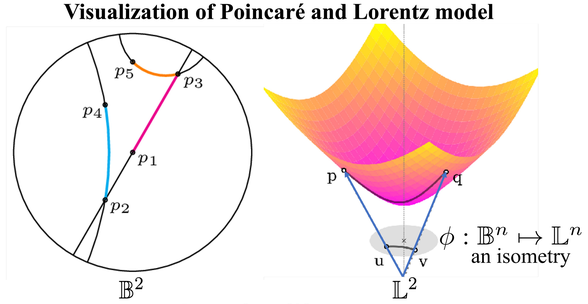


Image from [2]

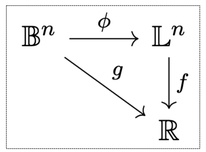
Key Perspective

- By IEEE float64 standard, the addition of two numbers that differ by a magnitude of 10^{16} gives just the large number.
- **Thm 1**: *For a point* $\tilde{x} \in \mathbb{B}^n$ *and* $x = \phi(\tilde{x}) \in \mathbb{L}^n$ *, then*

$$||x|| = \Omega(\frac{1}{1-||\tilde{x}||})$$

- **Thm 2***: Consider a smooth function* $f : \mathbb{L}^n \mapsto \mathbb{R}$ *and* $g := f \circ \phi$ *, then for any* $x \in \mathbb{B}^n : ||x|| = 1 - \delta$ *for small* δ *,*

$$\|\nabla_{\mathbb{B}^n} f(x)\| = \Omega\left(\delta^2 \|\nabla f(x)\|\right),$$
$$\|\nabla_{\mathbb{L}^n} g(\tilde{x})\| = O\left(\|\nabla f(x)\|\right).$$

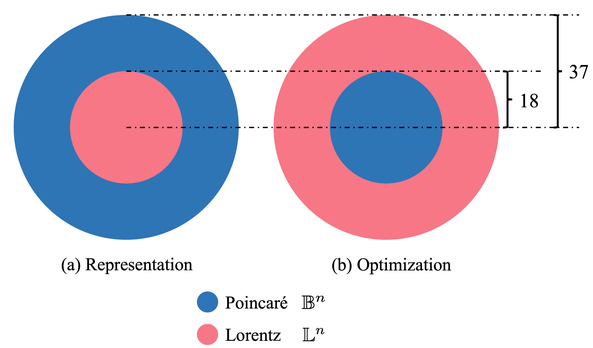


Relationship of definitions

Interpretation

- **Thm 1** implies that $\mathbb{B}^n$ and $\mathbb{L}^n$ use similar number of bits to represent points
- A more refined analysis shows that $\mathbb{B}^n$ has twice the diameter than $\mathbb{L}^n$ under the IEEE constraint.
- **Thm 2** suggests that $\mathbb{B}^n$ suffers more from a gradient vanishing problem when the magnitude is large due to its small Euclidean magnitude.

Regions of Accurate Representation/Optimization



Demonstration: Tree Embedding Optimization

- Create synthetic trees in $\mathbb{R}^2$ as embeddings
- Optimize using a proxy measure of distortion: d_E is thembedding distance and d_R the tree distances

$$\mathcal{L}(\theta) = \frac{1}{|S|(|S|-1)} \sum_{x \neq y \in S} (\alpha \frac{d_E(x,y)}{d_R(x,y)} - 1)^2$$

- We use Riemannian Adam with fixed learning rate and number of epochs
- Evaluate embedding using the average distortion

Results[3]

- A good embeding comes with a large diameter d and small distortion δ
- In all tested trees, $\mathbb{L}^n$ has larger diameter (more spread out) and smaller distortion
- The hierarchical structure is clearer in $\mathbb{L}^n$ embeddings

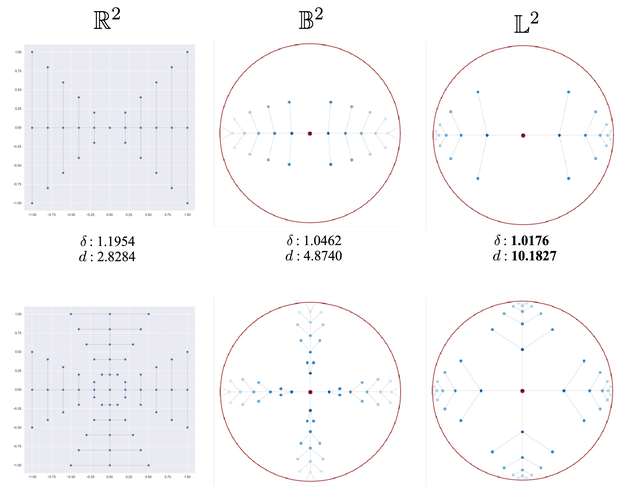


Table 1: Average distortion δ and diameter d by Dataset and Model

tree	manifold	δ	d	tree	manifold	δ	d
1	raw	1.1954	2.8284	5	raw	1.3349	1.4142
	$\mathbb{D}^2$	1.0462	4.8740		$\mathbb{D}^2$	1.0176	4.2523
	$\mathbb{L}^2$	**1.0176**	**10.1827**		$\mathbb{L}^2$	**1.0029**	**7.3850**
2	raw	1.1186	2.2361	6	raw	1.5080	1.4142
	$\mathbb{D}^2$	1.0589	4.9757		$\mathbb{D}^2$	1.1243	4.3168
	$\mathbb{L}^2$	**1.0180**	**10.6719**		$\mathbb{L}^2$	**1.0392**	**8.9204**
3	raw	1.0511	2.0396	7	raw	1.2108	0.8321
	$\mathbb{D}^2$	1.0321	6.3564		$\mathbb{D}^2$	1.2145	5.0657
	$\mathbb{L}^2$	**1.0190**	**9.6830**		$\mathbb{L}^2$	**1.0166**	**19.1703**
4	raw	1.1539	2.8284	8	raw	1.1421	0.9220
	$\mathbb{D}^2$	1.0754	5.0453		$\mathbb{D}^2$	1.0324	4.5670
	$\mathbb{L}^2$	**1.0421**	**10.4028**		$\mathbb{L}^2$	**1.0157**	**8.8324**

Acknowledgement

This work is partially supported by NSF under grants CCF-2112665, CCF- 2217058, and by NIH under grant RF1MH125317.

References

1. Sarkar, Rik. "Low distortion delaunay embedding of trees in hyperbolic plane." *International Symposium on Graph Drawing*. Springer, Berlin, Heidelberg, 2011.
2. Nickel, Maximillian, and Douwe Kiela. "Learning continuous hierarchies in the lorentz model of hyperbolic geometry." *ICML*. PMLR, 2018.
3. **Mishne, Gal, et al. "The Numerical Stability of Hyperbolic Representation Learning." *arXiv preprint arXiv:2211.00181* (2022).**