Convergence Properties of Hyperbolic Neural Networks

N. Alvarado[1,2] S. Burgos[3]

[1]National Center of Artificial Intelligence [2]Pontificia Universidad Católica de Chile [3]Pennsylvania State University

Motivation

- Hyperbolic neural networks (HNNs) are a emerging field AI that leverage hyperbolic geometry to enhance neural network performance.
- Theoretical foundations of HNNs are still not fully understood.
- Applying concepts from dynamical systems and ergodic theory to the convergence of neural networks can lead to significant improvements.
- Ergodic theory also helps mitigate chaotic behavior during training, leading to more stable and predictable training dynamics.

Hyperbolic Geometry

- Understanding how hyperbolic space is represented and visualized using models is crucial.

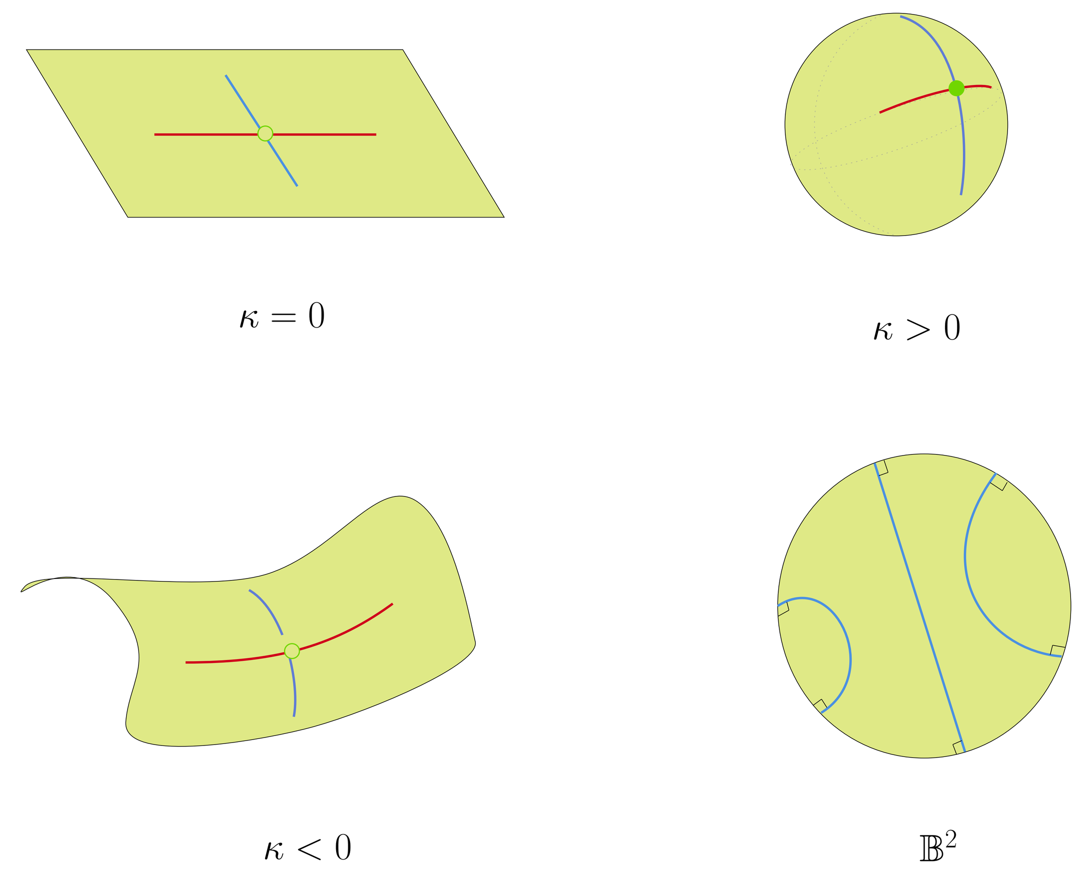

Figure 1:Different curvatures in geometry. At the bottom we have hyperbolic spaces (left) and the Poincaré ball model (right).

- We consider the set $\mathbb{L}^n := \{x \in \mathbb{R}^{n+1} : -x_0^2 + \Sigma_{i=1}^n x_i^2 = -1,\ x_0 > 0\}$ and we fix its origin $y = (1, 0, \ldots, 0) \in \mathbb{L}^n$.
- In this setting the exponential map $\exp_y : T_y\mathbb{L}^n \to \mathbb{L}^n$ is invertible and its inverse is denoted by $\log_y : \mathbb{L}^n \to T_y\mathbb{L}^n$.

Hyperbolic Neural Networks

- For $a, b, x \in \mathbb{L}^n$ and $\alpha \in \mathbb{R}$, we define $\oplus$ and $\otimes$ by $a \oplus b = \exp_y(\log_y a + \log_y b)$ and $\alpha \otimes x = \exp_y(\alpha \log_y(x))$.
- We define a Hyperbolic Neural Network as

$$f(x) = f_1 \circ f_2 \circ \cdots \circ f_k(x)$$
$$f_i(x) = \sigma_i^{\otimes}(W_i^{\otimes} x \oplus b_i), \quad 1 \le i \le k.$$

where $W_i \in \mathbb{R}^{n \times n}$, $b_i \in \mathbb{L}^n$ and σ is the activation function. Recall that we are always identifying $T_y\mathbb{L}^n \simeq \mathbb{R}^n$.

Ergodic Theory Basics

- Let $(M, \mathcal{B}, \mu, T)$ be an ergodic dynamical system. A *subadditive cocycle* over T is a measurable function $\phi : M \times \mathbb{N}_0 \to \mathbb{R}$ satisfying

$$\phi(\omega, n+m) \le \phi(\omega, n) + \phi(T^n\omega, m) \quad \text{for all } \omega \in M \text{ and } n, m > 0.$$

- Let $X \subset \mathbb{L}^n$ be a cone. A map $f : X \to X$ is called subhomogeneous if for every $x \in X$ and $\lambda \in (0, 1)$ we have $f(\lambda \otimes x) \le \lambda \otimes f(x)$, whenever the order is possible.
- Let $f : T_y\mathbb{L}^n \to T_y\mathbb{L}^n$ be subhomogeneous. Then, the induced map on the hyperboloid $f^{\otimes} : \mathbb{L}^n \to \mathbb{L}^n$ is also subhomogeneous.

Main Results

- Let $Y = \exp_y(X)$, where X is the positive cone in $\mathbb{R}^n$. Let $f_i : Y \to Y$ be a sequence of order preserving and subhomogeneous maps such that $T_m := \log_y \circ f_m \circ \exp_y$ is a stationary sequence of maps in X. Let $z_m = f_1 f_2 \cdots f_m(z_0)$ for a fixed $z_0 \in Y$. Then, we have

$$\lim_{m\to\infty} \sup_{1 \le i \le n} \left(\frac{\sqrt{2}\,\mathrm{arccosh}(z_m(0))}{\sqrt{\|z_m\|^2 - 1}} z_m(i) \right)^{1/m} = e^{\lambda}.$$

- Let (Ω, d_0) be a compact metric space and consider a stationary sequence of homeomorphisms $T_m : \Omega \to \Omega$. Then, almost surely there is a number λ such that

$$\lim_{m\to\infty} \left(\sup_{x \ne y} \frac{d_0(T_m T_{m-1} \cdots T_1 x, T_m T_{m-1} \cdots T_1 y)}{d_0(x, y)} \right)^{1/m} = e^{\lambda}.$$

Main Results

- Let (M, g) be a Riemannian manifold. Fix $y \in M$ and $r > 0$ such that $\varphi := \exp_y : B_r(0) \subset T_yM \to V := \exp_y(B_r(0))$ is a diffeomorphism. Consider a sequence $f_n : V \to V$ consisting of maps of the form $f(x) = \varphi(W^{\top}\sigma(W\varphi^{-1}(x) + b))$, where $\|W\| \le 1$, σ is 1-Lipschitz componentwise and $b \in T_yM$ satisfy $f_n(V) \subset V$, and such that $\tilde{f}_n(v) = W_n^{\top}\sigma(W_n v + b_n))$ is a stationary sequence of layer maps in $\mathbb{R}^n$. Then, as $m \to \infty$, almost surely there exist $z \in V$ such that

$$\frac{1}{m} \otimes f_1 f_2 \cdots f_m(z_0) \to z.$$

- An immediate application of this result is use it in the hyperboloid model (or in any isometric model, e.g. the Poincaré ball model).

Conclusions and future work

- In this work, we extend neural network convergence results from Euclidean spaces to Riemannian manifolds.
- We proved the convergence of HNNs under certain conditions, particularly in the Lorentz model, ensuring their stability and predictability.
- This work suggests that understanding parameter trajectories can lead to new regularization methods that prevent overfitting and enhance the generalization of neural networks.
- Empirical validation of the theorems is necessary to confirm their practical applicability and effectiveness in real-world scenarios.
- By using the exponential map and its inverse (when defined), it would be interesting to study neural networks in specific manifolds, e.g. the sphere, the torus, etc.

References