RIEMANNIAN EMBEDDED TRANSFORMERS

Hilal Kılıç
Department of Physics
Middle East Technical University
Ankara, Türkiye
hilal.kilic@metu.edu.tr

February 2, 2026

ABSTRACT

The capacity of Euclidean Neural Networks to represent complex hierarchical data is fundamentally limited by the polynomial growth of Euclidean space. In this work, we present a robust implementation of a Transformer architecture grounded in the Poincaré Ball model of Hyperbolic geometry. By defining the manifold's structural integrity through its metric tensor $g = \lambda^2\delta$, we derive and implement a suite of Riemannian operations, including the Möbius addition, exponential mapping, and geodesic distance functions. Unlike standard architectures that collapse hierarchical relationships into flat embeddings, our Riemannian Embedding layer utilizes the exponential map to project tangent vectors into a curved manifold where space grows exponentially with the radius. We provide a rigorous translation of the geodesic equations into computationally efficient PyTorch operations, ensuring numerical stability through precision clamping and conformal scaling. This framework lays the foundation for Hyperbolic Attention mechanisms, offering a geometrically principled approach to modeling the latent hierarchies inherent in natural language and relational datasets.

Mathematical Framework for the Poincaré Manifold

The Poincaré Ball model $(\mathbb{B}^n, g_x)$ is defined on the open unit ball $\mathbb{B}^n = \{x \in \mathbb{R}^n : \sqrt{c}\|x\| < 1\}$ with curvature $c > 0$.

1. The Riemannian Metric

The metric tensor g_x is a conformal scaling of the Euclidean metric δ:

$$g_x = \lambda_x^2 \delta, \quad where \quad \lambda_x = \frac{2}{1 - c\|x\|^2} \tag{1}$$

The conformal factor λ_x represents the local "stretching" of the space as one moves toward the boundary.

2. Möbius Addition

The operation used to perform translations within the manifold, ensuring the result remains in $\mathbb{B}^n$:

$$x \oplus_c y = \frac{(1 + 2c\langle x, y\rangle + c\|y\|^2)x + (1 - c\|x\|^2)y}{1 + 2c\langle x, y\rangle + c^2\|x\|^2\|y\|^2} \tag{2}$$

3. Exponential Map

The map from the tangent space $T_p\mathbb{B}^n$ to the manifold $\mathbb{B}^n$. For a vector v at the origin ($p = 0$):

$$exp_0(v) = \tanh(\sqrt{c}\|v\|)\frac{v}{\sqrt{c}\|v\|} \tag{3}$$

For a general point p, the map is given by:

$$exp_p(v) = p \oplus_c \left(\tanh\left(\frac{\sqrt{c}\lambda_p \|v\|}{2} \right) \frac{v}{\sqrt{c}\|v\|} \right) \tag{4}$$

4. Geodesic Distance

The distance between two points $x, y \in \mathbb{B}^n$ along the manifold's shortest curved path:

$$d_c(x, y) = \frac{1}{\sqrt{c}} acosh \left(1 + 2c \frac{\|x - y\|^2}{(1 - c\|x\|^2)(1 - c\|y\|^2)} \right) \tag{5}$$