Proceedings Track

Klein Model for Hyperbolic Neural Networks

Editors: List of editors' names

Abstract

Hyperbolic neural networks (HNNs) have been proved effective in modeling complex data structures. However, previous works mainly focused on the Poincaré ball model and the hyperboloid model as coordinate representations of the hyperbolic space, often neglecting the Klein model. Despite this, the Klein model offers its distinct advantages thanks to its straight-line geodesics, which facilitates the well-known Einstein midpoint construction, previously leveraged to accompany HNNs in other models. In this work, we introduce a framework for hyperbolic neural networks based on the Klein model. We provide detailed formulation for representing useful operations using the Klein model. We further study the Klein linear layer and prove that the "tangent space construction" of the scalar multiplication and parallel transport are exactly the Einstein scalar multiplication and the Einstein addition, analogous to the Möbius operations used in the Poincaré ball model. We show numerically that the Klein HNN performs on par with the Poincaré ball model, providing a third option for HNN that works as a building block for more complicated architectures.

Keywords: Hyperbolic neural network, Klein model, Einstein gyrovector space

1. Introduction

Hyperbolic spaces have shown considerable promise in embedding complex networks (Krioukov et al., 2010), trees (Wilson et al., 2014; Sonthalia and Gilbert, 2020) and hierarchical datasets (Nickel and Kiela, 2018). To leverage the inherent geometric structures within these data types in learning neural representations, hyperbolic neural networks (HNNs) were introduced, initially by Ganea et al. (2018a) and later expanded upon by many recent works (Peng et al., 2021). To define neural operations in hyperbolic spaces, HNNs typically utilize a model of hyperbolic geometry, where points are represented by Euclidean coordinates. Most of the literature focuses on either the Poincaré ball model or the hyperboloid model. Indeed, Ganea et al. (2018a); Shimizu et al. (2021) presented a set of operations for implementing HNNs using the Poincaré ball model, which often enjoys a simple mathematical description. On the other hand, the hyperboloid model proves to be numerically more stable than the Poincaré ball model (Nickel and Kiela, 2018; Qu and Zou, 2022; Mishne et al., 2023) and facilitates hyperbolic "linear" layers without using tangent spaces (Chen et al., 2022).

Despite the success of HNNs using the Poincaré ball and the hyperboloid models, HNNs using other models remain largely unexplored. This raises problems when important operations have to be done using a certain model. Indeed, in many works (Gulcehre et al., 2019; Zhu et al., 2020; Khrulkov et al., 2020; Zhang and Gao, 2021; Tai et al., 2021; Song et al., 2022; Fu et al., 2024; Skrodzki et al., 2024; Li et al., 2024a), in order to use the Einstein midpoint to perform aggregation, a mapping between the Poincaré ball/hyperboloid model and the Klein model has to be implemented whenever aggregation is needed. This causes unnecessary computational complexity. The Klein model also enjoys other crucial properties

such as straight-line geodesics. However, to the best of our knowledge, no prior works have discussed compact neural operations using the Klein model, that is, "pure Klein" HNNs. Moreover, recently developed software packages for HNNs lack implementation in the Klein model (van Spengler et al., 2023).

In this paper, we derive compact formulas for key operations in the Klein model. Analogous to the Möbius operations in the Poincaré ball model, we show that the Klein model facilitates a set of operations known as Einstein scalar multiplication and Einstein addition (Ungar, 2009, 2012). For the Poincaré ball model, Ganea et al. (2018a) demonstrated that the Möbius operations are exactly the exponential maps of "linear" operations (as used in multilayer perceptrons for obtaining pre-activations) in tangent spaces. However, what remains unclear is the precise relationship between Einstein operations and similar "tangent space constructions" in the Klein model. In our paper, we prove that they are equivalent operations. Thanks to this interpretation, we build Klein HNNs based on these Einstein operations. Through experiments on well-known heterophilic graph datasets, we show that the performance of Klein HNNs is on par with both Poincaré HNNs and hyperboloid HNNs. Furthermore, Klein HNNs are efficient to implement.

2. Related Works

Hyperbolic neural networks The first HNN was introduced relatively recently by Ganea et al. (2018a), who utilized the Poincaré ball model to develop fundamental neural operations including linear layers and recurrent layers. Importantly, they demonstrated that the Möbius operations correspond to the tangent space operations that one would intuitively define for neural operations in the Poincaré ball model. Later, Shimizu et al. (2021) extended upon their work and introduced more operations. Another line of research employed the hyperboloid model, favored for its numerical stability (Gulcehre et al., 2019; Chen et al., 2022). More complex hyperbolic neural operations have since been designed for a variety of applications, particularly for graph data (Chami et al., 2019; Liu et al., 2019; Mathieu et al., 2019; Dai et al., 2021; Zhang et al., 2021; Yang et al., 2022, 2024b) and image data (Khrulkov et al., 2020; Atigh et al., 2022; Ermolov et al., 2022; Desai et al., 2023; Yang et al., 2024c). Additionally, there have been studies that focused on the numerical stability (Mishne et al., 2023) and robustness (Li et al., 2024b) of HNNs. Despite these advances, most models continue to rely on either the Poincaré ball or hyperboloid models, with other hyperbolic models largely overlooked.

The Klein model in use Recently, the Klein model has been proved effective in representing various types of data, including hierarchical graphs (McDonald and He, 2020; Yang et al., 2024a), protein sequences (Ali et al., 2024), minimal spanning trees (García-Castellanos et al., 2024), and scene images (Bi et al., 2017). A key advantage of the Klein model is that its geodesics are represented as straight lines, which facilitates intuitive constructions of Voronoi diagrams (Nielsen and Nock, 2010), Delaunay graphs (Medbouhi et al., 2024) and SVM decision boundaries (Cho et al., 2019). Celińska-Kopczyńska and Kopczyński (2024) studied numerical precision of hyperbolic models including the Klein model. Moreover, properties of the Klein model are also used to assist in proofs for other hyperbolic models (He et al., 2024). However, the above works did not explore neural operations in the Klein model.

Although early works (Taherian, 2010; Rostamzadeh and Taherian, 2014) studied the algebraic structures and trigonometry of the Klein model, they did not address constructing neural network layers. While Ungar (2009, 2012); Kim and Lawson (2013) summarized the gyrogroup properties of the Einstein operations in connection with the Klein model, they did not mention the properties that we introduce in this paper.

3. The Klein Model of Hyperbolic Geometry

The n-dimensional hyperbolic space, denoted as $\mathbb{H}^n$, is a simply connected n-dimensional Riemannian manifold with constant negative curvature. To represent points in $\mathbb{H}^n$, there are several isometric models including the Poincaré ball model $\mathbb{B}^n$ and the hyperboloid (Lorentz) model $\mathbb{L}^n$, both of which are commonly used in the HNN literature. The Klein model (also known as Beltrami-Klein), denoted as $\mathbb{K}^n$, is another notable representation. In our paper, we consider $\mathbb{H}^n$ with a curvature -1 since this is used in most of the HNN literature.

Recent works have derived compact formulas for basic operations in $\mathbb{B}^n$ and $\mathbb{L}^n$ (Ganea et al., 2018a; Chami et al., 2019). Our derivation will use these formulas to avoid complicated calculations of, e.g., connections. In this section, we introduce the Klein model and describe the isometric mappings between the Klein model and the other two models. A detailed review of $\mathbb{B}^n$ and $\mathbb{L}^n$ is provided in Appendix A.

The n-dimensional Klein model with a curvature -1 is represented as $\mathbb{K}^n = \{\mathbf{x} \in \mathbb{R}^n, \|\mathbf{x}\| < 1\}$, with the Riemannian metric tensor $g^{\mathcal{K}}_{\mathbf{x}} = g^{\mathcal{K}}(\mathbf{x})$ expressed as:

$$g^{\mathcal{K}}_{ij}(\mathbf{x}) = \frac{\delta_{ij}}{1 - \|\mathbf{x}\|^2} + \frac{x_i x_j}{(1 - \|\mathbf{x}\|^2)^2}, \tag{1}$$

where $\|\cdot\|$ is the Euclidean norm. Denoting points in $\mathbb{R}^{n+1}$ as $[x_0, x_1, \cdots, x_n]^\top$, the Klein ball can be obtained by mapping $\mathbf{x} \in \mathbb{L}^n$ to the hyperplane $x_0 = 1$, using a gnomonic (central) projection with rays emanating from the origin. Specifically, we review the following formulas for isometric mappings between hyperbolic models. We say two points are "corresponding" if they are related by such an isometric mapping.

Mappings between models We use $\pi_{\mathcal{L}\to\mathcal{K}}$ to denote the isometric mapping from $\mathbb{L}^n$ to $\mathbb{K}^n$ and $\pi_{\mathcal{K}\to\mathcal{L}}$ the isometric mapping from $\mathbb{K}^n$ to $\mathbb{L}^n$. Namely, for $\mathbf{x}^{\mathcal{L}} = [x^{\mathcal{L}}_0, x^{\mathcal{L}}_1, \cdots, x^{\mathcal{L}}_n]^\top \in \mathbb{L}^n$ and $\mathbf{x}^{\mathcal{K}} = [x^{\mathcal{K}}_1, \cdots, x^{\mathcal{K}}_n]^\top \in \mathbb{K}^n$,

$$\pi_{\mathcal{L}\to\mathcal{K}}(\mathbf{x}^{\mathcal{L}}) = \left[\frac{x^{\mathcal{L}}_1}{x^{\mathcal{L}}_0}, \cdots, \frac{x^{\mathcal{L}}_n}{x^{\mathcal{L}}_0}\right]^\top \in \mathbb{K}^n, \quad \pi_{\mathcal{K}\to\mathcal{L}}(\mathbf{x}^{\mathcal{K}}) = \frac{1}{\sqrt{1 - \|\mathbf{x}^{\mathcal{K}}\|^2}}\left[1, \mathbf{x}^\top_{\mathcal{K}}\right]^\top \in \mathbb{L}^n. \tag{2}$$

We use $\pi_{\mathcal{B}\to\mathcal{K}}$ to denote the isometric mapping from $\mathbb{B}^n$ to $\mathbb{K}^n$ and $\pi_{\mathcal{K}\to\mathcal{B}}$ the isometric mapping from $\mathbb{K}^n$ to $\mathbb{B}^n$. Namely, for $\mathbf{x}^{\mathcal{B}} \in \mathbb{B}^n$ and $\mathbf{x}^{\mathcal{K}} \in \mathbb{K}^n$,

$$\pi_{\mathcal{B}\to\mathcal{K}}(\mathbf{x}^{\mathcal{B}}) = \frac{2}{1 + \|\mathbf{x}^{\mathcal{B}}\|^2}\mathbf{x}^{\mathcal{B}} \in \mathbb{K}^n, \quad \pi_{\mathcal{K}\to\mathcal{B}}(\mathbf{x}^{\mathcal{K}}) = \frac{1}{1 + \sqrt{1 - \|\mathbf{x}^{\mathcal{K}}\|^2}}\mathbf{x}^{\mathcal{K}} \in \mathbb{B}^n. \tag{3}$$

In $\mathbb{K}^n$, the tangent space $\mathcal{T}_{\mathbf{x}^{\mathcal{K}}}\mathbb{K}^n$ at $\mathbf{x}^{\mathcal{K}}$ is represented as $\mathbb{R}^n$ where $\frac{\partial}{\partial x^{\mathcal{K}}_i}$ is represented as $\mathbf{e}_i$, the i-th canonical basis. Similar conventions are adopted for $\mathbb{B}^n$ and $\mathbb{L}^n$. The tangent

vectors are related using pushforward maps associated with the above isometric mappings. We say that two vectors are "corresponding" if they are related by such a pushforward map. We recall that a pushforward map of an isometry preserves inner products and parallel transports.

The following lemma presents formulas for obtaining corresponding tangent vectors between $\mathbb{K}^n$ and other models. The proofs of all the results are presented in Appendix B.

Lemma 1 *Let $\mathbf{x}^{\mathcal{K}} \in \mathbb{K}^n$ and $\mathbf{v}^{\mathcal{K}} \in \mathcal{T}_{\mathbf{x}^{\mathcal{K}}}\mathbb{K}^n$ be a tangent vector at $\mathbf{x}^{\mathcal{K}}$.*

1. *Let $\mathbf{x}^{\mathcal{B}} \in \mathbb{B}^n$ be the corresponding point of $\mathbf{x}^{\mathcal{K}}$ and $\mathbf{v}^{\mathcal{B}} \in \mathcal{T}_{\mathbf{x}^{\mathcal{B}}}\mathbb{B}^n$ be the corresponding tangent vector of $\mathbf{x}^{\mathcal{K}}$. Then*

$$\mathbf{v}^{\mathcal{K}} = \frac{\partial \pi_{\mathcal{B}\to\mathcal{K}}(\mathbf{x}^{\mathcal{B}})}{\partial \mathbf{x}^{\mathcal{B}}}\mathbf{v}^{\mathcal{B}} = \frac{2}{1+\|\mathbf{x}^{\mathcal{B}}\|^2}\mathbf{v}^{\mathcal{B}} - \frac{4\mathbf{x}^{\mathcal{B}}\cdot\mathbf{v}^{\mathcal{B}}}{(1+\|\mathbf{x}^{\mathcal{B}}\|^2)^2}\mathbf{x}^{\mathcal{B}}, \tag{4}$$

$$\begin{aligned}\mathbf{v}^{\mathcal{B}} &= \frac{\partial \pi_{\mathcal{K}\to\mathcal{B}}(\mathbf{x}^{\mathcal{K}})}{\partial \mathbf{x}^{\mathcal{K}}}\mathbf{v}^{\mathcal{K}} \\ &= \frac{1}{1+\sqrt{1-\|\mathbf{x}^{\mathcal{K}}\|^2}}\mathbf{v}^{\mathcal{K}} + \frac{\mathbf{x}^{\mathcal{K}}\cdot\mathbf{v}^{\mathcal{K}}}{\sqrt{1-\|\mathbf{x}^{\mathcal{K}}\|^2}(1+\sqrt{1-\|\mathbf{x}^{\mathcal{K}}\|^2})^2}\mathbf{x}^{\mathcal{K}}.\end{aligned} \tag{5}$$

2. *Let $\mathbf{x}^{\mathcal{L}} \in \mathbb{L}^n$ be the corresponding point of $\mathbf{x}^{\mathcal{K}}$ and $\mathbf{v}^{\mathcal{L}} \in \mathcal{T}_{\mathbf{x}^{\mathcal{L}}}\mathbb{L}^n$ be the corresponding tangent vector of $\mathbf{x}^{\mathcal{K}}$. Then*

$$\mathbf{v}^{\mathcal{K}} = \frac{\partial \pi_{\mathcal{L}\to\mathcal{K}}(\mathbf{x}^{\mathcal{L}})}{\partial \mathbf{x}^{\mathcal{L}}}\mathbf{v}^{\mathcal{L}} = -\frac{v_t}{x_t^2}\mathbf{x}_s + \frac{1}{x_t}\mathbf{v}_s, \tag{6}$$

$$\mathbf{v}^{\mathcal{L}} = \frac{\partial \pi_{\mathcal{K}\to\mathcal{L}}(\mathbf{x}^{\mathcal{K}})}{\partial \mathbf{x}^{\mathcal{K}}}\mathbf{v}^{\mathcal{K}} = \begin{bmatrix} \frac{\mathbf{x}^{\mathcal{K}}\cdot\mathbf{v}^{\mathcal{K}}}{(1-\|\mathbf{x}^{\mathcal{K}}\|^2)^{\frac{3}{2}}} \\ \frac{1}{\sqrt{1-\|\mathbf{x}^{\mathcal{K}}\|^2}}\mathbf{v}^{\mathcal{K}} + \frac{\mathbf{x}^{\mathcal{K}}\cdot\mathbf{v}^{\mathcal{K}}}{(1-\|\mathbf{x}^{\mathcal{K}}\|^2)^{\frac{3}{2}}}\mathbf{x}^{\mathcal{K}} \end{bmatrix}. \tag{7}$$

Distances The induced distance function on the Klein model can be inferred from the one on the hyperboloid model, which we present as the following lemma.

Lemma 2 *Let $\mathbf{x}^{\mathcal{K}}, \mathbf{y}^{\mathcal{K}} \in \mathbb{K}^n$. Their geodesic distance is given by*

$$d_{\mathcal{K}}(\mathbf{x}^{\mathcal{K}}, \mathbf{y}^{\mathcal{K}}) = d_{\mathcal{L}}(\pi_{\mathcal{K}\to\mathcal{L}}(\mathbf{x}^{\mathcal{K}}), \pi_{\mathcal{K}\to\mathcal{L}}(\mathbf{y}^{\mathcal{K}})) = \cosh^{-1}\left(\frac{1-\mathbf{x}^{\mathcal{K}}\cdot\mathbf{y}^{\mathcal{K}}}{\sqrt{1-\|\mathbf{x}^{\mathcal{K}}\|^2}\sqrt{1-\|\mathbf{y}^{\mathcal{K}}\|^2}}\right). \tag{8}$$

Unit-speed geodesics The parametric expression of unit-speed geodesics in the Klein model is given in the following lemma, which is derived based on the fact that an isometric mapping maps geodesics in $\mathbb{L}^n$ to geodesics in $\mathbb{K}^n$.

Lemma 3 *Let $\mathbf{x}^{\mathcal{K}} \in \mathbb{K}^n$ and $\mathbf{v}^{\mathcal{K}} \in \mathcal{T}_{\mathbf{x}}\mathbb{K}^n$ with $g^{\mathcal{K}}_{\mathbf{x}^{\mathcal{K}}}(\mathbf{v}^{\mathcal{K}}, \mathbf{v}^{\mathcal{K}}) = 1$. Let $\gamma_{\mathbf{x}^{\mathcal{K}},\mathbf{v}^{\mathcal{K}}}(t)$ denote the unit-speed geodesic in $\mathbb{K}^n$ with $\gamma_{\mathbf{x}^{\mathcal{K}},\mathbf{v}^{\mathcal{K}}}(0) = \mathbf{x}^{\mathcal{K}}$ and $\dot{\gamma}_{\mathbf{x}^{\mathcal{K}},\mathbf{v}^{\mathcal{K}}}(0) = \mathbf{v}^{\mathcal{K}}$, then*

$$\gamma_{\mathbf{x}^{\mathcal{K}},\mathbf{v}^{\mathcal{K}}}(t) = \mathbf{x}^{\mathcal{K}} + \frac{\sinh(t)\mathbf{v}^{\mathcal{K}}}{\cosh(t) + \lambda_{\mathbf{x}^{\mathcal{K}}}(\mathbf{x}^{\mathcal{K}}\cdot\mathbf{v}^{\mathcal{K}})\sinh(t)}, \tag{9}$$

where $\lambda_{\mathbf{x}^{\mathcal{K}}} := 1/\sqrt{1-\|\mathbf{x}^{\mathcal{K}}\|^2}$ is used globally in this paper.

Accordingly, we can write out the exponential maps and logarithmic maps for the Klein model, summarized as the following corollary.

Corollary 4 *Given* $\mathbf{x}^{\mathcal{K}} \in \mathbb{K}^n$ *and* $\mathbf{v}^{\mathcal{K}} \in \mathcal{T}_{\mathbf{x}^{\mathcal{K}}}\mathbb{K}^n$*, denote* $\mathbf{u}^{\mathcal{K}} = \mathbf{v}^{\mathcal{K}}/\|\mathbf{v}^{\mathcal{K}}\|_{\mathcal{K}}$ *with* $\|\cdot\|_{\mathcal{K}}^2 = g_{\mathbf{x}^{\mathcal{K}}}^{\mathcal{K}}(\cdot,\cdot)$*. Let* $\gamma_{\mathbf{x}^{\mathcal{K}},\mathbf{u}^{\mathcal{K}}}(t)$ *be the unit-speed geodesic in* $\mathbb{K}^n$ *with* $\gamma_{\mathbf{x}^{\mathcal{K}},\mathbf{u}^{\mathcal{K}}}(0) = \mathbf{u}^{\mathcal{K}}$ *and* $\dot{\gamma}_{\mathbf{x}^{\mathcal{K}},\mathbf{u}^{\mathcal{K}}}(0) = \mathbf{u}^{\mathcal{K}}$*, the exponential map* $\exp_{\mathbf{x}^{\mathcal{K}}}^{\mathcal{K}} : \mathcal{T}_{\mathbf{x}^{\mathcal{K}}}\mathbb{K}^n \to \mathbb{K}^n$ *is given by*

$$\exp_{\mathbf{x}^{\mathcal{K}}}^{\mathcal{K}}(\mathbf{v}^{\mathcal{K}}) = \gamma_{\mathbf{x}^{\mathcal{K}},\mathbf{u}^{\mathcal{K}}}(\|\mathbf{v}^{\mathcal{K}}\|_{\mathcal{K}}) = \mathbf{x}^{\mathcal{K}} + \frac{\sinh(\|\mathbf{v}^{\mathcal{K}}\|_{\mathcal{K}})\frac{\mathbf{v}^{\mathcal{K}}}{\|\mathbf{v}^{\mathcal{K}}\|_{\mathcal{K}}}}{\cosh(\|\mathbf{v}^{\mathcal{K}}\|_{\mathcal{K}}) + \frac{\lambda_{\mathbf{x}^{\mathcal{K}}}^2}{\|\mathbf{v}^{\mathcal{K}}\|_{\mathcal{K}}}(\mathbf{x}^{\mathcal{K}} \cdot \mathbf{v}^{\mathcal{K}})\sinh(\|\mathbf{v}^{\mathcal{K}}\|_{\mathcal{K}})}. \tag{10}$$

The logarithmic map $\log_{\mathbf{x}^{\mathcal{K}}}^{\mathcal{K}} : \mathbb{K}^n \to \mathcal{T}_{\mathbf{x}^{\mathcal{K}}}\mathbb{K}^n$ *is given by*

$$\log_{\mathbf{x}^{\mathcal{K}}}^{\mathcal{K}}(\mathbf{y}^{\mathcal{K}}) = d_{\mathcal{K}}(\mathbf{x}^{\mathcal{K}},\mathbf{y}^{\mathcal{K}})\frac{\mathbf{y}^{\mathcal{K}} - \mathbf{x}^{\mathcal{K}}}{\|\mathbf{y}^{\mathcal{K}} - \mathbf{x}^{\mathcal{K}}\|_{\mathcal{K}}} = \cosh^{-1}\left(\frac{1 - \mathbf{x}^{\mathcal{K}} \cdot \mathbf{y}^{\mathcal{K}}}{\sqrt{1 - \|\mathbf{x}^{\mathcal{K}}\|^2}\sqrt{1 - \|\mathbf{y}^{\mathcal{K}}\|^2}}\right)\frac{\mathbf{y}^{\mathcal{K}} - \mathbf{x}^{\mathcal{K}}}{\|\mathbf{y}^{\mathcal{K}} - \mathbf{x}^{\mathcal{K}}\|_{\mathcal{K}}}, \tag{11}$$

In particular, it follows that

$$\exp_{\mathbf{o}^{\mathcal{K}}}^{\mathcal{K}}(\mathbf{v}^{\mathcal{K}}) = \tanh(\|\mathbf{v}^{\mathcal{K}}\|)\frac{\mathbf{v}^{\mathcal{K}}}{\|\mathbf{v}^{\mathcal{K}}\|}, \quad \log_{\mathbf{o}^{\mathcal{K}}}^{\mathcal{K}}(\mathbf{y}^{\mathcal{K}}) = \cosh^{-1}(\lambda_{\mathbf{y}^{\mathcal{K}}})\frac{\mathbf{y}^{\mathcal{K}}}{\|\mathbf{y}^{\mathcal{K}}\|}. \tag{12}$$

Parallel transport The parallel transport $P_{\mathbf{x}\to\mathbf{y}}$ defines a linear isometry moving tangent vectors along the geodesic from $\mathbf{x}$ to $\mathbf{y}$. Unfortunately, the formula for parallel transport between general points would be complicated in the Klein model. We only present the case where the starting point is $\mathbf{o}^{\mathcal{K}}$, the hyperbolic origin in $\mathbb{K}^n$, which is sufficient for the use case of hyperbolic neural networks.

Proposition 5 *The parallel transport of a tangent vector* $\mathbf{v}^{\mathcal{K}} \in \mathcal{T}_{\mathbf{o}^{\mathcal{K}}}\mathbb{K}^n$ *to the tangent space* $\mathcal{T}_{\mathbf{x}^{\mathcal{K}}}\mathbb{K}^n$ *at an arbitrary point* $\mathbf{x}^{\mathcal{K}}$ *in the Klein model is*

$$P_{\mathbf{o}^{\mathcal{K}}\to\mathbf{x}^{\mathcal{K}}}(\mathbf{v}^{\mathcal{K}}) = \frac{(\mathbf{x}^{\mathcal{K}} \cdot \mathbf{v}^{\mathcal{K}})(\sqrt{1 - \|\mathbf{x}^{\mathcal{K}}\|^2} - 2)}{1 - \sqrt{1 - \|\mathbf{x}^{\mathcal{K}}\|^2}}\mathbf{x}^{\mathcal{K}} + \sqrt{1 - \|\mathbf{x}^{\mathcal{K}}\|^2}\mathbf{v}^{\mathcal{K}} \in \mathcal{T}_{\mathbf{x}^{\mathcal{K}}}\mathbb{K}^n. \tag{13}$$

4. Klein Model for Hyperbolic Neural Networks

To build HNNs in hyperbolic space, we need to define operations including weight matrix transformation and bias translation in a way analogous to "$\mathbf{W}\mathbf{x} + \mathbf{b}$". A natural approach is as follows. First, apply the logarithmic map to project the input to the tangent space at the hyperbolic origin. Then for weight matrix transformation, perform the matrix-vector multiplication in the tangent space; for bias translation, apply the parallel transport to the hyperbolic bias vector from the origin to the input point. Finally, use the exponential map to bring the output back to hyperbolic space. For the Poincaré ball model, Ganea et al. (2018a) proved that these "tangent space operations" are equivalent to basic operations in the Möbius gyrovector spaces. In the context of the Klein model, the corresponding gyrovector space is the Einstein gyrovector space. This raises a natural question: are the tangent space operations in the Klein model also equivalent to those in the Einstein gyrovector space?

In Section 4.1, we review the definitions of the Einstein gyrovector spaces. We prove the equivalence of tangent space operations and Einstein gyrovector space operations in Section 4.2. Finally, we derive HNNss in the Klein model in Section 4.3.

4.1. Einstein Gryogroups and Gyrovector Spaces

Einstein gyrogroup Let $\mathbb{V}$ be a real inner product space and let $\mathbb{V}_c = \{\mathbf{x} \in \mathbb{V} : \|\mathbf{x}\| < c\}$ be an open c-ball of $\mathbb{V}$. Then the Einstein addition of $\mathbf{x}, \mathbf{y} \in \mathbb{V}_c$ is defined as

$$\mathbf{x} \oplus_{\mathrm{E}} \mathbf{y} = \frac{1}{1 + \frac{\mathbf{x}\cdot\mathbf{y}}{c^2}} \left(\mathbf{x} + \frac{1}{\gamma_{\mathbf{x}}}\mathbf{y} + \frac{1}{c^2}\frac{\gamma_{\mathbf{x}}}{1+\gamma_{\mathbf{x}}}(\mathbf{x}\cdot\mathbf{y})\mathbf{x} \right), \tag{14}$$

where $\gamma_{\mathbf{x}} = (1 - \frac{\|\mathbf{x}\|^2}{c^2})^{-\frac{1}{2}}$ is the Lorentz factor. These designations originate from Einstein's theory of special relativity: if $\mathbb{V} = \mathbb{R}^3$, Equation (14) represents the relativistic addition of velocities $\mathbf{x}, \mathbf{y}$ in space rather than spacetime, where c is the speed of light.

In the general case where $\mathbf{x}, \mathbf{y}$ are not parallel, $\oplus_{\mathrm{E}}$ is neither commutative nor associative, rendering the corresponding algebraic structure merely a groupoid. As observed by Ungar, however, this groupoid has the interesting property whereby an automorphism $\mathrm{gyr}[\mathbf{x}, \mathbf{y}] : \mathbb{V}_c \to \mathbb{V}_c$ respecting $\oplus_E$, called gyration, defines the obstruction to commutativity through $\mathbf{x} \oplus_{\mathrm{E}} \mathbf{y} = \mathrm{gyr}[\mathbf{x}, \mathbf{y}]\mathbf{y} \oplus_{\mathrm{E}} \mathbf{x}$ as well as to associativity through $\mathbf{x} \oplus_{\mathrm{E}} (\mathbf{y} \oplus_{\mathrm{E}} \mathbf{z}) = (\mathbf{x} \oplus_{\mathrm{E}} \mathbf{y}) \oplus_{\mathrm{E}} \mathrm{gyr}[\mathbf{x}, \mathbf{y}]\mathbf{z}$. This groupoid structure $(\mathbb{V}_c, \oplus_{\mathrm{E}})$ is called a gyrocommutative gyrogroup, and describes Thomas precession in the example of special relativity. We refer to $(\mathbb{V}_c, \oplus_{\mathrm{E}})$ as Einstein gyrogroup for short, and more details on its properties can be found, e.g., in Ungar (2009, sec. 1.2, 2.3).

Einstein gyrovector space Given an Einstein gyrogroup, we can endow it with another operation called Einstein scalar multiplication: letting $r \in \mathbb{R}, \mathbf{x} \in \mathbb{V}_c$, and $\mathbf{x} \neq \mathbf{0}$, it is

$$r \otimes_{\mathrm{E}} \mathbf{x} = c \tanh\left(r \tanh^{-1}\left(\frac{\|\mathbf{x}\|}{c} \right) \right) \frac{\mathbf{x}}{\|\mathbf{x}\|}. \tag{15}$$

This turns out to have associative and distributive properties, and we apply the notation $r \otimes_{\mathrm{E}} \mathbf{x} = \mathbf{x} \otimes_{\mathrm{E}} r$. The resulting structure $(\mathbb{V}_c, \oplus_{\mathrm{E}}, \otimes_{\mathrm{E}})$ is called Einstein gyrovector space, and see, e.g., Ungar (2009, sec. 3.1, 3.8) for relevant properties. Notably, its inherited inner product is invariant under gyrations, that is, $\mathrm{gyr}[\mathbf{x}, \mathbf{y}]\mathbf{u} \cdot \mathrm{gyr}[\mathbf{x}, \mathbf{y}]\mathbf{v} = \mathbf{u} \cdot \mathbf{v}$ for all $\mathbf{u}, \mathbf{v} \in$. Einstein gyrovector space turns out to be a useful framework for hyperbolic geometry, since we recover the Klein model of curvature -1 for $\mathbb{V} = \mathbb{R}^n$ and $c = 1$.

4.2. Connecting the Klein Model with Einstein Gyrovector Spaces

Next, we show how Einstein gyrovector spaces facilitate the tangent space construction of scalar multiplication and vector addition in the Klein model. We start with the geodesics in the Klein model and their formulation using operations in the Einstein gyrovector spaces.

Geodesics The geodesic $\gamma_{\mathbf{x}^{\mathcal{K}} \to \mathbf{y}^{\mathcal{K}}} : \mathbb{R} \to \mathbb{K}^n$ connecting points $\mathbf{x}^{\mathcal{K}}, \mathbf{y}^{\mathcal{K}} \in \mathbb{K}^n$, such that $\gamma_{\mathbf{x}^{\mathcal{K}} \to \mathbf{y}^{\mathcal{K}}}(0) = \mathbf{x}^{\mathcal{K}}$ and $\gamma_{\mathbf{x}^{\mathcal{K}} \to \mathbf{y}^{\mathcal{K}}}(1) = \mathbf{y}^{\mathcal{K}}$, is shown by Ungar (2009, sec. 3.9) to be:

$$\gamma_{\mathbf{x}^{\mathcal{K}} \to \mathbf{y}^{\mathcal{K}}}(t) = \mathbf{x}^{\mathcal{K}} \oplus_{\mathrm{E}} \left(\ominus_{\mathrm{E}} \mathbf{x}^{\mathcal{K}} \oplus_{\mathrm{E}} \mathbf{y}^{\mathcal{K}} \right) \otimes_{\mathrm{E}} t, \tag{16}$$

where $\ominus_{\mathrm{E}} \mathbf{x}^{\mathcal{K}}$ is the inverse of $\mathbf{x}^{\mathcal{K}}$ in the Einstein gyrogroup: $\ominus_{\mathrm{E}} \mathbf{x}^{\mathcal{K}} \oplus_{\mathrm{E}} \mathbf{x}^{\mathcal{K}} = \mathbf{o}^{\mathcal{K}}$.

Scaling and Einstein scalar multiplication We study the tangent space construction of performing scalar multiplication in the Klein model and build the identity with the Einstein addition. The following theorem summarizes our finding:

Theorem 6 *For $\forall r \in \mathbb{R}, \mathbf{x}^{\mathcal{K}} \in \mathbb{K}^n$, performing scalar multiplication in the Klein model, which first uses the logarithmic map to project $\mathbf{x}^{\mathcal{K}}$ to $\mathcal{T}_{\mathbf{o}^{\mathcal{K}}}\mathbb{K}^n$, then multiplies this projection by a scalar in the tangent space and projects it back to the manifold with the exponential map, can be achieved by directly applying the Einstein scalar multiplication, namely,*

$$r \otimes_E \mathbf{x}^{\mathcal{K}} = \exp^{\mathcal{K}}_{\mathbf{o}^{\mathcal{K}}}\left(r \log^{\mathcal{K}}_{\mathbf{o}^{\mathcal{K}}}\left(\mathbf{x}^{\mathcal{K}}\right)\right). \tag{17}$$

Parallel transport and Einstein addition We also connect the parallel transport with the Einstein addition using exponential and logarithmic maps in the following theorem:

Theorem 7 *The parallel transport w.r.t. the Levi-Civita connection of a tangent vector $\mathbf{v}^{\mathcal{K}} \in \mathcal{T}_{\mathbf{o}^{\mathcal{K}}}\mathbb{K}^n$ to another tangent space $\mathcal{T}_{\mathbf{x}^{\mathcal{K}}}\mathbb{K}^n$ in the Klein model is given by*

$$P_{\mathbf{o}^{\mathcal{K}} \to \mathbf{x}^{\mathcal{K}}}(\mathbf{v}^{\mathcal{K}}) = \log^{\mathcal{K}}_{\mathbf{x}^{\mathcal{K}}}\left(\mathbf{x}^{\mathcal{K}} \oplus_E \exp^{\mathcal{K}}_{\mathbf{o}^{\mathcal{K}}}(\mathbf{v}^{\mathcal{K}})\right). \tag{18}$$

4.3. Hyperbolic Neural Networks in the Klein Model

Theorems 6 and 7 provide compact formulas for defining neural operations for HNNs using the Klein model.

Weight matrix transformation and nonlinear activation First, thanks to Theorem 6, we define the "Einstein version" of Euclidean functions that preserve the origin.

Definition 8 (Einstein version) *Given $f : \mathbb{R}^n \to \mathbb{R}^m$ such that $f(\mathbf{0}) = (\mathbf{0})$, the Einstein version of f, denoted as $f^{\otimes_E} : \mathbb{K}^n \to \mathbb{K}^m$, is defined by*

$$f^{\otimes_E}(\mathbf{x}^{\mathcal{K}}) := \exp^{\mathcal{K}}_{\mathbf{o}^{\mathcal{K}}}\left(f\left(\log^{\mathcal{K}}_{\mathbf{o}^{\mathcal{K}}}(\mathbf{x}^{\mathcal{K}})\right)\right). \tag{19}$$

For instance, if $\sigma : \mathbb{R}^n \to \mathbb{R}^n$ is a non-linear activation function, then its Einstein version $\sigma^{\otimes_{\mathrm{E}}}$ can be applied to points in the Klein model. Namely,

$$\sigma^{\otimes_{\mathrm{E}}}(\mathbf{x}^{\mathcal{K}}) := \exp^{\mathcal{K}}_{\mathbf{o}^{\mathcal{K}}}\left(\sigma\left(\log^{\mathcal{K}}_{\mathbf{o}^{\mathcal{K}}}(\mathbf{x}^{\mathcal{K}})\right)\right). \tag{20}$$

Moreover, we identify matrix-vector multiplication as an Einstein version of linear maps. Then the following result is straightforward.

Theorem 9 (Einstein matrix-vector multiplication) *If $\mathbf{M} : \mathbb{R}^n \to \mathbb{R}^m$ is a linear map, which we identify with its matrix representation, then for $\mathbf{x}^{\mathcal{K}} \in \mathbb{K}^n$ s.t. $\mathbf{M}\mathbf{x}^{\mathcal{K}} \neq \mathbf{0}$,*

$$\begin{aligned}\mathbf{M}^{\otimes_E}(\mathbf{x}^{\mathcal{K}}) &= \exp^{\mathcal{K}}_{\mathbf{o}^{\mathcal{K}}}(\mathbf{M} \log^{\mathcal{K}}_{\mathbf{o}^{\mathcal{K}}}(\mathbf{x}^{\mathcal{K}})) \\ &= \tanh\left(\frac{2\|\mathbf{M}\mathbf{x}^{\mathcal{K}}\|}{\|\mathbf{x}^{\mathcal{K}}\|}\tanh^{-1}\left(\frac{\|\mathbf{x}^{\mathcal{K}}\|}{1+\sqrt{1-\|\mathbf{x}^{\mathcal{K}}\|^2}}\right)\right)\frac{\mathbf{M}\mathbf{x}^{\mathcal{K}}}{\|\mathbf{M}\mathbf{x}^{\mathcal{K}}\|};\end{aligned} \tag{21}$$

and if $\mathbf{M}\mathbf{x}^{\mathcal{K}} = \mathbf{0}$, then $\mathbf{M}^{\otimes_E}(\mathbf{x}^{\mathcal{K}}) = \mathbf{0}$. Moreover, if we define the Einstein matrix-vector multiplication of $\mathbf{M} \in \mathcal{M}_{m,n}(\mathbb{R})$ and $\mathbf{x} \in \mathbb{K}^n$ by $\mathbf{M} \otimes_E \mathbf{x}^{\mathcal{K}} := \mathbf{M}^{\otimes_E}(\mathbf{x}^{\mathcal{K}})$, then it shares common properties with the matrix-vector multiplication in the Euclidean space.

1. *For $\mathbf{M} \in \mathcal{M}_{l,m}(\mathbb{R})$, $\mathbf{M}' \in \mathcal{M}_{m,n}(\mathbb{R})$, $(\mathbf{M}\mathbf{M}') \otimes_E \mathbf{x}^{\mathcal{K}} = \mathbf{M} \otimes_E (\mathbf{M}' \otimes_E \mathbf{x}^{\mathcal{K}})$;*
2. *For $r > 0$, $\mathbf{M} \in \mathcal{M}_{m,n}(\mathbb{R})$, $(r\mathbf{M}) \otimes_E \mathbf{x}^{\mathcal{K}} = r \otimes_E (\mathbf{M} \otimes_E \mathbf{x}^{\mathcal{K}})$;*
3. *For $\mathbf{M} \in \mathcal{O}_n(\mathbb{R})$, $\mathbf{M} \otimes_E \mathbf{x}^{\mathcal{K}} = \mathbf{M}\mathbf{x}^{\mathcal{K}}$.*

Bias translation Next, Theorem 7 provides a compact formula for bias translation. Specifically, we can use the Einstein addition in place of the bias translation of a point $\mathbf{x}^{\mathcal{K}} \in \mathbb{K}^n$ with a hyperbolic bias $\mathbf{b}^{\mathcal{K}} \in \mathbb{K}^n$,

$$\begin{aligned} \mathbf{x}^{\mathcal{K}} &\mapsto \exp^{\mathcal{K}}_{\mathbf{x}^{\mathcal{K}}}\left(P^{\mathcal{K}}_{\mathbf{o}^{\mathcal{K}}\to\mathbf{x}^{\mathcal{K}}}\left(\log^{\mathcal{K}}_{\mathbf{o}^{\mathcal{K}}}(\mathbf{b}^{\mathcal{K}})\right)\right) \\ &= \exp^{\mathcal{K}}_{\mathbf{x}^{\mathcal{K}}}\left(\log^{\mathcal{K}}_{\mathbf{x}^{\mathcal{K}}}\left(\mathbf{x}^{\mathcal{K}} \oplus_{\mathrm{E}} \exp^{\mathcal{K}}_{\mathbf{o}^{\mathcal{K}}}\left(\log^{\mathcal{K}}_{\mathbf{o}^{\mathcal{K}}}\left(\mathbf{b}^{\mathcal{K}}\right)\right)\right)\right) = \mathbf{x}^{\mathcal{K}} \oplus_{\mathrm{E}} \mathbf{b}^{\mathcal{K}}. \end{aligned} \tag{22}$$

In summary, to build Klein HNNs, Einstein operations are all we need.

5. Experiments

We perform node classification tasks on the well-known WebKB datasets including Texas, Wisconsin and Chameleon (Craven et al., 2000), Actor datasets (Tang et al., 2009) and citation datasets including Cora and Pubmed (Sen et al., 2008), summarized in Appendix C. These datasets enjoy high Gromov hyperbolicities (Jonckheere et al., 2008; Adcock et al., 2013). We compare HNNs using the three hyperbolic models we have discussed. We use HNNs containing a hyperbolic linear layer and a Euclidean classification layer. We use (hyperbolic version of) ReLU as the activation function. The HNNs are trained using the Riemannian Adam (Becigneul and Ganea, 2019) for 5000 epochs with early stopping.

Table 1 presents the test accuracies. The results indicate that Klein HNNs performs on par with both HNNs using the Poincaré ball model and the hyperboloid model.

Table 1: The test accuracy of two-layer HNNs (mean and std over 3 trials)

Dataset	Texas	Wisconsin	Chameleon	Actor	Cora	Pubmed
Poincaré	0.9697±0.0347	0.9506±0.0283	0.7442±0.0064	0.6436±0.0057	0.5960±0.0101	0.7270±0.0036
Hyperboloid	0.9697±0.0132	0.9444±0.0321	0.7418±0.0120	0.6342±0.0226	0.6067±0.0137	0.7293±0.0064
Klein	0.9697±0.0263	0.9568±0.0214	0.7375±0.0119	0.6509±0.0067	0.5957±0.0106	0.7230±0.0139

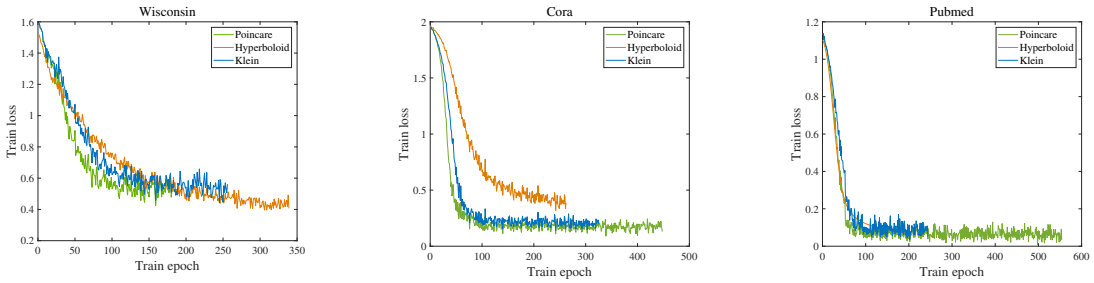
In Table 2, we report the average run time to train each epoch. Evidently, it takes a longer time to train one epoch using the hyperboloid HNN compared with its Poincaré ball and Klein counterparts. This is because no simple formulation exists for the weight matrix transformation and bias translation in the hyperboloid model and we have to perform the tangent space operations which involve a series of exponential and logarithmic maps, which cause computational complexity. Thanks to the connection between the Klein model and Einstein gyrovector spaces, we are able to implement the Klein linear layer in a simple and elegant way. Emperically, Klein HNNs are as efficient as Poincaré HNNs and sometimes the most efficient among all the three models.

Table 2: Average runtime for training one epoch (mean and std over 3 trials)

Dataset	Texas	Wisconsin	Chameleon	Actor	Cora	Pubmed
Poincaré	0.0222±0.0025	0.0205±0.0008	0.0248±0.0008	0.0263±0.0009	0.0198±0.0010	0.0220±0.0023
Hyperboloid	0.0396±0.0087	0.0341±0.0012	0.0224±0.0021	0.0420±0.0014	0.0362±0.0022	0.0351±0.0014
Klein	0.0219±0.0017	0.0208±0.0020	0.0234±0.0029	0.0238±0.0008	0.0214±0.0012	0.0222±0.0017

To study how training loss changes differently with training epochs using different models, we plot their relationships when training on some selected datasets. Figure 1 shows

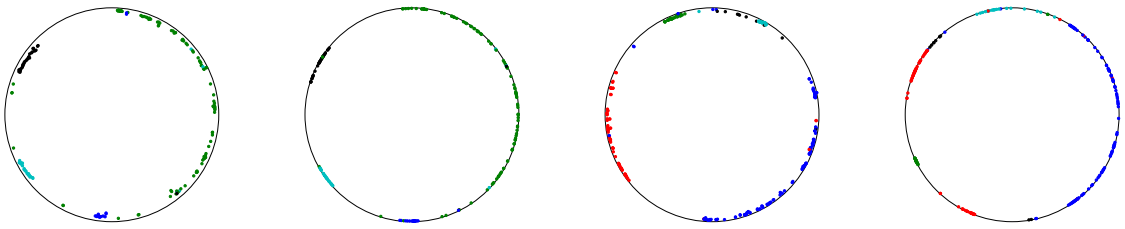
the result on the Wisconsin, Cora, and Pubmed datasets. We observe that the training dynamics of the HNNs are very similar.



(a) Wisconsin (b) Cora (c) Pubmed

Figure 1: Change of training loss with epochs using different models.

We also plot the hyperbolic features, i.e., output of the hyperbolic linear layer. For illustration, we project them into the 2D-disk using hyperbolic t-SNE. Figure 2 shows the results on the Texas and Wisconsin datasets for Poincaré and Klein HNNs. Features with different labels are indicated by different colors. We observe that the representations show similar patterns for the same dataset, indicating the effectiveness of our Klein HNNs.



(a) Texas ($\mathbb{B}$) (b) Texas ($\mathbb{K}$) (c) Wisconsin ($\mathbb{B}$) (d) Wisconsin ($\mathbb{K}$)

Figure 2: Features projected into the 2D-disks.

6. Conclusion

In this paper, we provide a detailed framework for Klein HNNs, addressing the gaps largely overlooked from previous works. We derive compact formulas for key operations in the Klein model. Subsequently, we connect the geometry of the Klein model with the Einstein gyrovector spaces, providing a simple and elegant formulation for the tangent space construction of the Klein linear layer. Our experiments show that Klein HNNs achieve comparable performance with Poincaré/hyperboloid HNNs while maintaining training efficiency. Furthermore, our framework supports additional operations, such as Einstein midpoints, all within the Klein model. Future works include extending the current operations to other common neural operations.

References

Aaron B Adcock, Blair D Sullivan, and Michael W Mahoney. Tree-like structure in large social and information networks. In *2013 IEEE 13th international conference on data mining*, pages 1–10. IEEE, 2013.

Sarwan Ali, Haris Mansoor, Prakash Chourasia, Yasir Ali, and Murray Patterson. Gaussian beltrami-klein model for protein sequence classification: A hyperbolic approach. In *International Symposium on Bioinformatics Research and Applications*, pages 52–62. Springer, 2024.

James W Anderson. *Hyperbolic geometry*. Springer Science & Business Media, 2006.

Mina Ghadimi Atigh, Julian Schoep, Erman Acar, Nanne van Noord, and Pascal Mettes. Hyperbolic image segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 4453–4462, June 2022.

Gary Becigneul and Octavian-Eugen Ganea. Riemannian adaptive optimization methods. In *International Conference on Learning Representations*, 2019. URL https://openreview.net/forum?id=r1eiqi09K7.

Yanhong Bi, Bin Fan, and Fuchao Wu. Multiple Cayley-Klein metric learning. *Plos one*, 12(9):e0184865, 2017.

Dorota Celińska-Kopczyńska and Eryk Kopczyński. Numerical aspects of hyperbolic geometry. In *International Conference on Computational Science*, pages 115–130. Springer, 2024.

Ines Chami, Zhitao Ying, Christopher Ré, and Jure Leskovec. Hyperbolic graph convolutional neural networks. *Advances in neural information processing systems*, 32:4868–4879, 2019.

Weize Chen, Xu Han, Yankai Lin, Hexu Zhao, Zhiyuan Liu, Peng Li, Maosong Sun, and Jie Zhou. Fully hyperbolic neural networks. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 5672–5686, 2022.

Hyunghoon Cho, Benjamin DeMeo, Jian Peng, and Bonnie Berger. Large-margin classification in hyperbolic space. In *The 22nd international conference on artificial intelligence and statistics*, pages 1832–1840. PMLR, 2019.

Mark Craven, Dan DiPasquo, Dayne Freitag, Andrew McCallum, Tom Mitchell, Kamal Nigam, and Seán Slattery. Learning to construct knowledge bases from the world wide web. *Artificial intelligence*, 118(1-2):69–113, 2000.

Jindou Dai, Yuwei Wu, Zhi Gao, and Yunde Jia. A hyperbolic-to-hyperbolic graph convolutional network. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 154–163, 2021.

Karan Desai, Maximilian Nickel, Tanmay Rajpurohit, Justin Johnson, and Shanmukha Ramakrishna Vedantam. Hyperbolic image-text representations. In *International Conference on Machine Learning*, pages 7694–7731. PMLR, 2023.

Aleksandr Ermolov, Leyla Mirvakhabova, Valentin Khrulkov, Nicu Sebe, and Ivan Oseledets. Hyperbolic vision transformers: Combining improvements in metric learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 7409–7419, 2022.

Xingcheng Fu, Yisen Gao, Yuecen Wei, Qingyun Sun, Hao Peng, Jianxin Li, and Xianxian Li. Hyperbolic geometric latent diffusion model for graph generation. In Ruslan Salakhutdinov, Zico Kolter, Katherine Heller, Adrian Weller, Nuria Oliver, Jonathan Scarlett, and Felix Berkenkamp, editors, *Proceedings of the 41st International Conference on Machine Learning*, volume 235 of *Proceedings of Machine Learning Research*, pages 14102–14124. PMLR, 21–27 Jul 2024. URL https://proceedings.mlr.press/v235/fu24c.html.

Octavian Ganea, Gary Bécigneul, and Thomas Hofmann. Hyperbolic neural networks. *Advances in neural information processing systems*, 31:5345–5355, 2018a.

Octavian Ganea, Gary Bécigneul, and Thomas Hofmann. Hyperbolic entailment cones for learning hierarchical embeddings. In *International conference on machine learning*, pages 1646–1655. PMLR, 2018b.

Alejandro García-Castellanos, Aniss Aiman Medbouhi, Giovanni Luca Marchetti, Erik J Bekkers, and Danica Kragic. Hypersteiner: Computing heuristic hyperbolic steiner minimal trees. *arXiv preprint arXiv:2409.05671*, 2024.

Caglar Gulcehre, Misha Denil, Mateusz Malinowski, Ali Razavi, Razvan Pascanu, Karl Moritz Hermann, Peter Battaglia, Victor Bapst, David Raposo, Adam Santoro, and Nando de Freitas. Hyperbolic attention networks. In *International Conference on Learning Representations*, 2019. URL https://openreview.net/forum?id=rJxHsjRqFQ.

Neil He, Menglin Yang, and Rex Ying. Lorentzian residual neural networks. In *ICML 2024 Workshop on Geometry-grounded Representation Learning and Generative Modeling*, 2024. URL https://openreview.net/forum?id=WRWKy1M77y.

Edmond Jonckheere, Poonsuk Lohsoonthorn, and Francis Bonahon. Scaled gromov hyperbolic graphs. *Journal of Graph Theory*, 57(2):157–180, 2008.

Valentin Khrulkov, Leyla Mirvakhabova, Evgeniya Ustinova, Ivan Oseledets, and Victor Lempitsky. Hyperbolic image embeddings. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 6418–6428, 2020.

Sejong Kim and Jimmie Lawson. Unit balls, lorentz boosts, and hyperbolic geometry. *Results in Mathematics*, 63(3):1225–1242, 2013.

Dmitri Krioukov, Fragkiskos Papadopoulos, Maksim Kitsak, Amin Vahdat, and Marián Boguná. Hyperbolic geometry of complex networks. *Physical Review E—Statistical, Nonlinear, and Soft Matter Physics*, 82(3):036106, 2010.

Hao Li, Hao Jiang, Dongsheng Ye, Qiang Wang, Liang Du, Yuanyuan Zeng, Yingxue Wang, Cheng Chen, et al. DHGAT: Hyperbolic representation learning on dynamic graphs via attention networks. *Neurocomputing*, 568:127038, 2024a.

Yuekang Li, Yidan Mao, Yifei Yang, and Dongmian Zou. Improving robustness of hyperbolic neural networks by lipschitz analysis. In *Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pages 1713–1724, 2024b.

Qi Liu, Maximilian Nickel, and Douwe Kiela. Hyperbolic graph neural networks. *Advances in Neural Information Processing Systems*, 32:8230–8241, 2019.

Emile Mathieu, Charline Le Lan, Chris J Maddison, Ryota Tomioka, and Yee Whye Teh. Continuous hierarchical representations with poincaré variational auto-encoders. *Advances in neural information processing systems*, 32, 2019.

David McDonald and Shan He. HEAT: Hyperbolic embedding of attributed networks. In *International Conference on Intelligent Data Engineering and Automated Learning*, pages 28–40. Springer, 2020.

Aniss Aiman Medbouhi, Giovanni Luca Marchetti, Vladislav Polianskii, Alexander Kravberg, Petra Poklukar, Anastasia Varava, and Danica Kragic. Hyperbolic delaunay geometric alignment. In *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*, pages 111–126. Springer, 2024.

Gal Mishne, Zhengchao Wan, Yusu Wang, and Sheng Yang. The numerical stability of hyperbolic representation learning. In *International Conference on Machine Learning*, pages 24925–24949. PMLR, 2023.

Maximillian Nickel and Douwe Kiela. Learning continuous hierarchies in the lorentz model of hyperbolic geometry. In *International Conference on Machine Learning*, pages 3779–3788. PMLR, 2018.

Frank Nielsen and Richard Nock. Hyperbolic Voronoi diagrams made easy. In *2010 International Conference on Computational Science and Its Applications*, pages 74–80. IEEE, 2010.

W. Peng, T. Varanka, A. Mostafa, H. Shi, and G. Zhao. Hyperbolic deep neural networks: A survey. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, December 2021. doi: 10.1109/TPAMI.2021.3136921.

Eric Qu and Dongmian Zou. Lorentzian fully hyperbolic generative adversarial network. *arXiv preprint arXiv:2201.12825*, 2022.

Mahfouz Rostamzadeh and Sayed-Ghahreman Taherian. On trigonometry in beltrami–klein model of hyperbolic geometry. *Results in Mathematics*, 65:361–369, 2014.

Prithviraj Sen, Galileo Namata, Mustafa Bilgic, Lise Getoor, Brian Galligher, and Tina Eliassi-Rad. Collective classification in network data. *AI magazine*, 29(3):93–93, 2008.

Ryohei Shimizu, YUSUKE Mukuta, and Tatsuya Harada. Hyperbolic neural networks++. In *International Conference on Learning Representations*, 2021.

Martin Skrodzki, Hunter van Geffen, Nicolas F Chaves-de Plaza, Thomas Hollt, Elmar Eisemann, and Klaus Hildebrandt. Accelerating hyperbolic t-SNE. *IEEE Transactions on Visualization and Computer Graphics*, 2024.

Mingyang Song, Yi Feng, and Liping Jing. Hyperbolic relevance matching for neural keyphrase extraction. In Marine Carpuat, Marie-Catherine de Marneffe, and Ivan Vladimir Meza Ruiz, editors, *Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 5710–5720, Seattle, United States, July 2022. Association for Computational Linguistics. doi: 10.18653/v1/2022.naacl-main.419. URL https://aclanthology.org/2022.naacl-main.419.

Rishi Sonthalia and Anna Gilbert. Tree! I am no tree! I am a low dimensional hyperbolic embedding. *Advances in Neural Information Processing Systems*, 33:845–856, 2020.

Sayed-Ghahreman Taherian. On algebraic structures related to beltrami–klein model of hyperbolic geometry. *Results in Mathematics*, 57(3):205–219, 2010.

Chang-You Tai, Chien-Kun Huang, Liang-Ying Huang, and Lun-Wei Ku. Knowledge based hyperbolic propagation. In *Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 1945–1949, 2021.

Jie Tang, Jimeng Sun, Chi Wang, and Zi Yang. Social influence analysis in large-scale networks. In *Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 807–816, 2009.

Abraham Albert Ungar. *A Gyrovector Space Approach to Hyperbolic Geometry*. Synthesis Lectures on Mathematics & Statistics. Springer International Publishing, Cham, 2009. URL https://link.springer.com/10.1007/978-3-031-02396-5.

Abraham Albert Ungar. *Beyond the Einstein addition law and its gyroscopic Thomas precession: The theory of gyrogroups and gyrovector spaces*, volume 117. Springer Science & Business Media, 2012.

Max van Spengler, Philipp Wirth, and Pascal Mettes. HypLL: The hyperbolic learning library. In *Proceedings of the 31st ACM International Conference on Multimedia*, pages 9676–9679, 2023.

Richard C Wilson, Edwin R Hancock, Elżbieta Pekalska, and Robert PW Duin. Spherical and hyperbolic embeddings of data. *IEEE transactions on pattern analysis and machine intelligence*, 36(11):2255–2269, 2014.

Meimei Yang, Qiao Liu, Xinkai Sun, Na Shi, and Hui Xue. Towards kernelizing the classifier for hyperbolic data. *Frontiers of Computer Science*, 18(1):181301, 2024a.

Menglin Yang, Min Zhou, Zhihao Li, Jiahong Liu, Lujia Pan, Hui Xiong, and Irwin King. Hyperbolic graph neural networks: A review of methods and applications, 2022.

Menglin Yang, Harshit Verma, Delvin Ce Zhang, Jiahong Liu, Irwin King, and Rex Ying. Hypformer: Exploring efficient transformer fully in hyperbolic space. In *Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pages 3770–3781, 2024b.

Yifei Yang, Wonjun Lee, Dongmian Zou, and Gilad Lerman. Improving hyperbolic representations via gromov-wasserstein regularization. *arXiv preprint arXiv:2407.10495*, 2024c.

Chengkun Zhang and Junbin Gao. Hype-HAN: Hyperbolic hierarchical attention network for semantic embedding. In *Proceedings of the Twenty-Ninth International Conference on International Joint Conferences on Artificial Intelligence*, pages 3990–3996, 2021.

Yiding Zhang, Xiao Wang, Chuan Shi, Nian Liu, and Guojie Song. Lorentzian graph convolutional networks. In *Proceedings of the Web Conference 2021*, pages 1249–1261, 2021.

Yudong Zhu, Di Zhou, Jinghui Xiao, Xin Jiang, Xiao Chen, and Qun Liu. HyperText: Endowing FastText with hyperbolic geometry. In Trevor Cohn, Yulan He, and Yang Liu, editors, *Findings of the Association for Computational Linguistics: EMNLP 2020*, pages 1166–1171, Online, November 2020. Association for Computational Linguistics. doi: 10.18653/v1/2020.findings-emnlp.104. URL https://aclanthology.org/2020.findings-emnlp.104.

Appendix A. Hyperbolic Geometry

We review basic facts about the Poincaré ball model and the hyperboloid model of hyperbolic space. For readers seeking a more introductory overview of hyperbolic geometry, we recommend Anderson (2006). We also recommend the work of Peng et al. (2021) for a survey of HNNs. For a comprehensive study of operations within these models, including those used in hyperbolic neural networks, we refer to the detailed work of Ganea et al. (2018a) and Chami et al. (2019).

A.1. The Poincaré Ball Model

The n-dimensional Poincaré ball model with constant negative curvature of -1 is the Riemannian manifold $\mathbb{B}^n = \{\mathbf{x} \in \mathbb{R}^n \mid \|\mathbf{x}\| < 1\}$ with the metric tensor $g_{\mathbf{x}}^{\mathbb{B}} = (\rho_{\mathbf{x}})^2 \mathbf{I}_n$, where $\rho_{\mathbf{x}} = 2\left(1 - \|\mathbf{x}\|^2\right)^{-1}$ is the conformal factor and $\mathbf{I}_n$ is the Euclidean metric. The geodesic distance between two point $\mathbf{x}, \mathbf{y} \in \mathbb{B}$ is given by

$$d_{\mathcal{B}}(\mathbf{x}, \mathbf{y}) = \cosh^{-1}\left(1 + \frac{2\|\mathbf{x} - \mathbf{y}\|^2}{\left(1 - \|\mathbf{x}\|^2\right)\left(1 - \|\mathbf{y}\|^2\right)}\right). \tag{23}$$

Given $\mathbf{x} \in \mathbb{B}^n$, $\mathcal{T}_{\mathbf{x}}\mathbb{B}^n$ denotes the tangent space of $\mathbb{B}^n$ at $\mathbf{x}$. For $\mathbf{x}, \mathbf{y} \in \mathbb{B}^n$ with $\mathbf{x} \neq \mathbf{y}$ and $\mathbf{v} \in \mathcal{T}_{\mathbf{x}}\mathbb{B}^n \backslash \{\mathbf{0}\}$, the exponential map $\exp_{\mathbf{x}}^{\mathcal{B}} : \mathcal{T}_{\mathbf{x}}\mathbb{B}^n \to \mathbb{B}^n$ and the logarithmic map $\log_{\mathbf{x}}^{\mathcal{B}} : \mathbb{B}^n \to \mathcal{T}_{\mathbf{x}}\mathbb{B}^n$ satisfy

$$\exp_{\mathbf{x}}^{\mathcal{B}}(\mathbf{v}) = \mathbf{x} \oplus_{\mathrm{M}} \left(\tanh\left(\frac{\rho_{\mathbf{x}}\|\mathbf{v}\|}{2}\right)\frac{\mathbf{v}}{\|\mathbf{v}\|}\right), \tag{24}$$

$$\log_{\mathbf{x}}^{\mathcal{B}}(\mathbf{y}) = \frac{2}{\rho_{\mathbf{x}}}\tanh^{-1}\left(\|-\mathbf{x} \oplus_{\mathrm{M}} \mathbf{y}\|\right)\frac{-\mathbf{x} \oplus_{\mathrm{M}} \mathbf{y}}{\|-\mathbf{x} \oplus_{\mathrm{M}} \mathbf{y}\|}, \tag{25}$$

where $\oplus_{\mathrm{M}}$ is the Möbius addition, defined by

$$\mathbf{x} \oplus_{\mathrm{M}} \mathbf{y} = \frac{\left(1 + 2\mathbf{x} \cdot \mathbf{y} + \|\mathbf{y}\|^2\right)\mathbf{x} + \left(1 - \|\mathbf{x}\|^2\right)\mathbf{y}}{1 + 2\mathbf{x} \cdot \mathbf{y} + \|\mathbf{x}\|^2\|\mathbf{y}\|^2}. \tag{26}$$

In particular, the exponential and logarithmic maps at the hyperbolic origin $\mathbf{o}$ are given by

$$\exp_{\mathbf{o}}^{\mathcal{B}}(\mathbf{v}) = \tanh(\|\mathbf{v}\|)\frac{\mathbf{v}}{\|\mathbf{v}\|}, \quad \log_{\mathbf{o}}^{\mathcal{B}}(\mathbf{y}) = \tanh^{-1}(\|\mathbf{y}\|)\frac{\mathbf{y}}{\|\mathbf{y}\|}. \tag{27}$$

Ganea et al. (2018a) proved that, the parallel transport w.r.t. the Levi-Civita connection of $\mathbf{v} \in \mathcal{T}_{\mathbf{o}}\mathbb{B}^n$ to the tangent space $\mathcal{T}_{\mathbf{x}}\mathbb{B}^n$ at $\mathbf{x}$ is given by

$$P_{\mathbf{o} \to \mathbf{x}}(\mathbf{v}) = \log_{\mathbf{x}}^{\mathcal{B}}\left(\mathbf{x} \oplus_{\mathrm{M}} \exp_{\mathbf{o}}^{\mathcal{B}}(\mathbf{v})\right). \tag{28}$$

A.2. The Hyperboloid Model

The n-dimensional hyperboloid model with constant negative curvature -1 is represented as $\mathbb{L}^n = \{\mathbf{x} \in \mathbb{R}^{n+1} | \langle \mathbf{x}, \mathbf{x} \rangle_{\mathcal{L}} = -1, x_t > 0\}$, where $\langle \cdot, \cdot \rangle_{\mathcal{L}}$ denotes the Minkowski inner product, $\langle \mathbf{x}, \mathbf{y} \rangle_{\mathcal{L}} := -x_0 y_0 + x_1 y_1 + \cdots + x_n y_n$. It has the metric tensor $g^{\mathbb{L}} = \operatorname{diag}\left([-1, \mathbf{1}_n^\top]\right)$. From a special relativity perspective, we may also write $\mathbf{x} \in \mathbb{L}^n$ as $[x_t, \mathbf{x}_s^\top]^\top$, for x_t being the time axis and $\mathbf{x}_s$ being the spatial axes.

The tangent space of $\mathbb{L}^n$ at $\mathbf{x}$ is the orthogonal space of $\mathbb{L}^n$ with respect to the Minkowski inner product and is represented as $\mathcal{T}_{\mathbf{x}}\mathbb{L}^n := \{\mathbf{v} \in \mathbb{R}^{n+1} : \langle \mathbf{v}, \mathbf{x} \rangle_{\mathcal{L}} = 0\}$. Specifically, let $\{\mathbf{e}_0, \cdots, \mathbf{e}_n\}$ be the canonical basis of $\mathbb{R}^{n+1}$, the tangent space is represented with the understanding that

$$\left.\frac{\partial}{\partial x_i}\right|_{\mathbf{x}} = \frac{\mathbf{x} \cdot \mathbf{e}_i}{\sqrt{1 + \|\mathbf{x}\|^2}} \mathbf{e}_0 + \mathbf{e}_i. \tag{29}$$

The unit-speed geodesic $\phi_{\mathbf{x},\mathbf{v}}(t)$ with $\phi_{\mathbf{x},\mathbf{v}}(0) = \mathbf{x} \in \mathbb{L}^n$ and $\dot{\phi}_{\mathbf{x},\mathbf{v}}(0) = \mathbf{v} \in \mathcal{T}_{\mathbf{x}}\mathbb{L}^n$ is given by

$$\phi_{\mathbf{x},\mathbf{v}}(t) = \mathbf{x} \cosh(t) + \mathbf{v} \sinh(t). \tag{30}$$

The geodesic distance of $\mathbf{x}, \mathbf{y} \in \mathbb{L}^n$ is given by $d_{\mathcal{L}}(\mathbf{x}, \mathbf{y}) = \cosh^{-1}(-\langle \mathbf{x}, \mathbf{y} \rangle_{\mathcal{L}})$. For $\mathbf{x}, \mathbf{y} \in \mathbb{L}^n$ with $\mathbf{x} \neq \mathbf{y}$ and $\mathbf{v} \in \mathcal{T}_{\mathbf{x}}\mathbb{L}^n \backslash \{\mathbf{0}\}$, the exponential map $\exp_{\mathbf{x}}^{\mathcal{L}} : \mathcal{T}_{\mathbf{x}}\mathbb{L}^n \to \mathbb{L}^n$ and the logarithmic map $\log_{\mathbf{x}}^{\mathcal{L}} : \mathbb{L}^n \to \mathcal{T}_{\mathbf{x}}\mathbb{L}^n$ are given by

$$\exp_{\mathbf{x}}^{\mathcal{L}}(\mathbf{v}) = \cosh\left(\|\mathbf{v}\|_{\mathcal{L}}\right) \mathbf{x} + \sinh\left(\|\mathbf{v}\|_{\mathcal{L}}\right) \frac{\mathbf{v}}{\|\mathbf{v}\|_{\mathcal{L}}}, \tag{31}$$

$$\log_{\mathbf{x}}^{\mathcal{L}}(\mathbf{y}) = d_{\mathcal{L}}(\mathbf{x}, \mathbf{y}) \frac{\mathbf{y} + \langle \mathbf{x}, \mathbf{y} \rangle_{\mathcal{L}} \mathbf{x}}{\|\mathbf{y} + \langle \mathbf{x}, \mathbf{y} \rangle_{\mathcal{L}} \mathbf{x}\|_{\mathcal{L}}}, \tag{32}$$

where $\|\mathbf{v}\|_{\mathcal{L}} = \sqrt{\langle \mathbf{v}, \mathbf{v} \rangle_{\mathcal{L}}}$ is the induced norm from the Minkowski inner product.

The parallel transport of a tangent vector $\mathbf{v} \in \mathcal{T}_{\mathbf{o}}\mathbb{L}^n$ at the hyperbolic origin $\mathbf{o} = [1, \mathbf{0}^\top]^\top$ to the tangent space $\mathcal{T}_{\mathbf{x}^{\mathcal{L}}}\mathbb{L}^n$ is given by

$$P_{\mathbf{o} \to \mathbf{x}}(\mathbf{v}) = \mathbf{v} - \frac{\langle \log_{\mathbf{o}}(\mathbf{x}), \mathbf{v} \rangle_{\mathcal{L}}}{d_{\mathcal{L}}(\mathbf{o}, \mathbf{x})^2} (\log_{\mathbf{o}}(\mathbf{x}) + \log_{\mathbf{x}}(\mathbf{o})). \tag{33}$$

Appendix B. Additional Proofs

B.1. Proof of Lemma 1

1. The corresponding tangent vectors $\mathbf{v}^{\mathcal{B}} \in \mathcal{T}_{\mathbf{x}^{\mathcal{B}}}\mathbb{B}^n$ and $\mathbf{v}^{\mathcal{K}} \in \mathcal{T}_{\mathbf{x}^{\mathcal{K}}}\mathbb{K}^n$ are related using Jacobian matrices. Namely,

$$\mathbf{v}^{\mathcal{K}} = \frac{\partial \pi_{\mathcal{B} \to \mathcal{K}}(\mathbf{x}^{\mathcal{B}})}{\partial \mathbf{x}^{\mathcal{B}}} \mathbf{v}^{\mathcal{B}}, \quad \mathbf{v}^{\mathcal{B}} = \frac{\partial \pi_{\mathcal{K} \to \mathcal{B}}(\mathbf{x}^{\mathcal{K}})}{\partial \mathbf{x}^{\mathcal{K}}} \mathbf{v}^{\mathcal{K}}. \tag{34}$$

Since

$$\frac{\partial \pi_{\mathcal{B} \to \mathcal{K}}(\mathbf{x}^{\mathcal{B}})}{\partial \mathbf{x}^{\mathcal{B}}} = \frac{2}{1 + \|\mathbf{x}^{\mathcal{B}}\|^2} \mathbf{I} + \mathbf{x}^{\mathcal{B}} \cdot \frac{-2 \cdot 2(\mathbf{x}^{\mathcal{B}})^\top}{(1 + \|\mathbf{x}^{\mathcal{B}}\|^2)^2}, \tag{35}$$

$$\frac{\partial \pi_{\mathcal{K} \to \mathcal{B}}(\mathbf{x}^{\mathcal{K}})}{\partial \mathbf{x}^{\mathcal{K}}} = \frac{1}{1 + \sqrt{1 - \|\mathbf{x}^{\mathcal{K}}\|^2}} \mathbf{I} + \mathbf{x}^{\mathcal{K}} \cdot \frac{-1 \cdot 2(\mathbf{x}^{\mathcal{K}})^\top}{2\sqrt{1 - \|\mathbf{x}^{\mathcal{K}}\|^2}(1 + \sqrt{1 - \|\mathbf{x}^{\mathcal{K}}\|^2})^2}, \tag{36}$$

we have

$$\mathbf{v}^{\mathcal{K}} = \frac{\partial \pi_{\mathcal{B} \to \mathcal{K}}(\mathbf{x}^{\mathcal{B}})}{\partial \mathbf{x}^{\mathcal{B}}} \mathbf{v}^{\mathcal{B}} = \frac{2}{1 + \|\mathbf{x}^{\mathcal{B}}\|^2} \mathbf{v}^{\mathcal{B}} - \frac{4\mathbf{x}^{\mathcal{B}} \cdot \mathbf{v}^{\mathcal{B}}}{(1 + \|\mathbf{x}^{\mathcal{B}}\|^2)^2} \mathbf{x}^{\mathcal{B}}, \tag{37}$$

$$\mathbf{v}^{\mathcal{K}} = \frac{\partial \pi_{\mathcal{K} \to \mathcal{B}}(\mathbf{x}^{\mathcal{K}})}{\partial \mathbf{x}^{\mathcal{K}}} \mathbf{v}^{\mathcal{K}} = \frac{1}{1 + \sqrt{1 - \|\mathbf{x}^{\mathcal{K}}\|^2}} \mathbf{v}^{\mathcal{K}} + \frac{\mathbf{x}^{\mathcal{K}} \cdot \mathbf{v}^{\mathcal{K}}}{\sqrt{1 - \|\mathbf{x}^{\mathcal{K}}\|^2}(1 + \sqrt{1 - \|\mathbf{x}^{\mathcal{K}}\|^2})^2} \mathbf{x}^{\mathcal{K}}. \tag{38}$$

2. Similarly,

$$\mathbf{v}^{\mathcal{K}} = \frac{\partial \pi_{\mathcal{L}\to\mathcal{K}}(\mathbf{x}^{\mathcal{L}})}{\partial \mathbf{x}^{\mathcal{L}}}\mathbf{v}^{\mathcal{L}}, \quad \mathbf{v}^{\mathcal{L}} = \frac{\partial \pi_{\mathcal{K}\to\mathcal{L}}(\mathbf{x}^{\mathcal{K}})}{\partial \mathbf{x}^{\mathcal{K}}}\mathbf{v}^{\mathcal{K}}. \tag{39}$$

For $\frac{\partial \pi_{\mathcal{L}\to\mathcal{K}}(\mathbf{x}^{\mathcal{L}})}{\partial \mathbf{x}^{\mathcal{L}}}$, let $\mathbf{y} = \pi_{\mathcal{L}\to\mathcal{K}}(\mathbf{x}^{\mathcal{L}})$. Then

$$\frac{\partial \pi_{\mathcal{L}\to\mathcal{K}}(\mathbf{x}^{\mathcal{L}})}{\partial \mathbf{x}^{\mathcal{L}}} = \frac{\partial \mathbf{y}}{\partial \mathbf{x}^{\mathcal{L}}} = \begin{bmatrix} \frac{\partial \mathbf{y}}{\partial x_t} & \frac{\partial \mathbf{y}}{\partial \mathbf{x}_s} \end{bmatrix}, \tag{40}$$

with $\frac{\partial \mathbf{y}}{\partial x_t}$ being a $n \times 1$ column vector and $\frac{\partial \mathbf{y}}{\partial \mathbf{x}_s}$ being a $n \times n$ matrix. The isometric mapping in Equation (2) indicates that $\mathbf{y} = \frac{1}{x_t}\mathbf{x}_s$, and thus

$$\frac{\partial \mathbf{y}}{\partial x_t} = -\frac{1}{x_t^2}\mathbf{x}_s, \quad \frac{\partial \mathbf{y}}{\partial \mathbf{x}_s} = \frac{1}{x_t}\mathbf{I}. \tag{41}$$

Writing $\mathbf{v}^{\mathcal{L}}$ as $\begin{bmatrix} v_t, \mathbf{v}_s^\top \end{bmatrix}^\top$, we have

$$\mathbf{v}^{\mathcal{K}} = \frac{\partial \pi_{\mathcal{L}\to\mathcal{K}}(\mathbf{x}^{\mathcal{L}})}{\partial \mathbf{x}^{\mathcal{L}}}\mathbf{v}^{\mathcal{L}} = \begin{bmatrix} \frac{\partial \mathbf{y}}{\partial x_t} & \frac{\partial \mathbf{y}}{\partial \mathbf{x}_s} \end{bmatrix} \begin{bmatrix} v_t \\ \mathbf{v}_s \end{bmatrix} = v_t \frac{\partial \mathbf{y}}{\partial x_t} + \frac{\partial \mathbf{y}}{\partial \mathbf{x}_s}\mathbf{v}_s = -\frac{v_t}{x_t^2}\mathbf{x}_s + \frac{1}{x_t}\mathbf{v}_s. \tag{42}$$

Hence,

$$\mathbf{v}^{\mathcal{K}} = \frac{\partial \pi_{\mathcal{L}\to\mathcal{K}}(\mathbf{x}^{\mathcal{L}})}{\partial \mathbf{x}^{\mathcal{L}}}\mathbf{v}^{\mathcal{L}} = -\frac{v_t}{x_t^2}\mathbf{x}_s + \frac{1}{x_t}\mathbf{v}_s \in \mathcal{T}_{\mathbf{x}^{\mathcal{K}}}\mathbb{K}^n. \tag{43}$$

For $\frac{\partial \pi_{\mathcal{K}\to\mathcal{L}}(\mathbf{x}^{\mathcal{K}})}{\partial \mathbf{x}^{\mathcal{K}}}$, let $\mathbf{z} = \pi_{\mathcal{K}\to\mathcal{L}}(\mathbf{x}^{\mathcal{K}}) = \begin{bmatrix} z_t, \mathbf{z}_s^\top \end{bmatrix}^\top$. Then

$$\frac{\partial \pi_{\mathcal{K}\to\mathcal{L}}(\mathbf{x}^{\mathcal{K}})}{\partial \mathbf{x}^{\mathcal{K}}} = \frac{\partial \mathbf{z}}{\partial \mathbf{x}^{\mathcal{K}}} = \begin{bmatrix} \frac{\partial z_t}{\partial \mathbf{x}^{\mathcal{K}}} \\ \frac{\partial \mathbf{z}_s}{\partial \mathbf{x}^{\mathcal{K}}} \end{bmatrix}. \tag{44}$$

with $\frac{\partial z_t}{\partial \mathbf{x}^{\mathcal{K}}}$ being a $1 \times n$ row vector and $\frac{\partial \mathbf{z}_s}{\partial \mathbf{x}^{\mathcal{K}}}$ being a $n \times n$ matrix. From the mapping between the hyperboloid model and the Klein model, we know

$$z_t = \frac{1}{\sqrt{1 - \|\mathbf{x}^{\mathcal{K}}\|^2}}, \quad \mathbf{z}_s = \frac{1}{\sqrt{1 - \|\mathbf{x}^{\mathcal{K}}\|^2}}\mathbf{x}^{\mathcal{K}}. \tag{45}$$

We compute $\frac{\partial z_t}{\partial \mathbf{x}^{\mathcal{K}}}$ and $\frac{\partial \mathbf{z}_s}{\partial \mathbf{x}^{\mathcal{K}}}$ respectively:

$$\frac{\partial z_t}{\partial \mathbf{x}^{\mathcal{K}}} = -\frac{1}{2}\left(1 - \|\mathbf{x}^{\mathcal{K}}\|^2\right)^{-\frac{3}{2}} \cdot \left(-2(\mathbf{x}^{\mathcal{K}})^\top\right) = \frac{(\mathbf{x}^{\mathcal{K}})^\top}{(1 - \|\mathbf{x}^{\mathcal{K}}\|^2)^{\frac{3}{2}}}; \tag{46}$$

$$\frac{\partial \mathbf{z}_s}{\partial \mathbf{x}^{\mathcal{K}}} = \frac{1}{\sqrt{1 - \|\mathbf{x}^{\mathcal{K}}\|^2}}\mathbf{I} + \mathbf{x}^{\mathcal{K}} \cdot \frac{(\mathbf{x}^{\mathcal{K}})^\top}{(1 - \|\mathbf{x}^{\mathcal{K}}\|^2)^{\frac{3}{2}}}. \tag{47}$$

Putting everything together,

$$\mathbf{v}^{\mathcal{L}} = \frac{\partial \pi_{\mathcal{K}\to\mathcal{L}}(\mathbf{x}^{\mathcal{K}})}{\partial \mathbf{x}^{\mathcal{K}}}\mathbf{v}^{\mathcal{K}} = \begin{bmatrix} \frac{\mathbf{x}^{\mathcal{K}} \cdot \mathbf{v}^{\mathcal{K}}}{(1 - \|\mathbf{x}^{\mathcal{K}}\|^2)^{\frac{3}{2}}} \\ \frac{1}{\sqrt{1 - \|\mathbf{x}^{\mathcal{K}}\|^2}}\mathbf{v}^{\mathcal{K}} + \frac{\mathbf{x}^{\mathcal{K}} \cdot \mathbf{v}^{\mathcal{K}}}{(1 - \|\mathbf{x}^{\mathcal{K}}\|^2)^{\frac{3}{2}}}\mathbf{x}^{\mathcal{K}} \end{bmatrix} \in \mathcal{T}_{\mathbf{x}^{\mathcal{L}}}\mathbb{L}^n. \tag{48}$$

■

B.2. Proof of Lemma 3

The proof is analogous to that of $\mathbb{B}^n$ (Ganea et al., 2018b, Theorem 1). Let $\mathbf{x}^{\mathcal{L}} \in \mathbb{L}^n$ and $\mathbf{v}^{\mathcal{L}} \in \mathcal{T}_{\mathbf{x}^{\mathcal{L}}}\mathbb{L}^n$ with $\langle \mathbf{v}^{\mathcal{L}}, \mathbf{v}^{\mathcal{L}} \rangle_{\mathcal{L}} = 1$. According to Equation (30), the unit-speed geodesic $\phi(t) = \phi_{\mathbf{x}^{\mathcal{L}},\mathbf{v}^{\mathcal{L}}}(t)$ with $\phi_{\mathbf{x}^{\mathcal{L}},\mathbf{v}^{\mathcal{L}}}(0) = \mathbf{x}^{\mathcal{L}}$ and $\dot{\phi}_{\mathbf{x}^{\mathcal{L}},\mathbf{v}^{\mathcal{L}}}(0) = \mathbf{v}^{\mathcal{L}}$ is

$$\phi_{\mathbf{x}^{\mathcal{L}},\mathbf{v}^{\mathcal{L}}}(t) = \mathbf{x}^{\mathcal{L}} \cosh(t) + \mathbf{v}^{\mathcal{L}} \sinh(t). \tag{49}$$

Let $\mathbf{x}^{\mathcal{K}} \in \mathbb{K}^n$ and $\mathbf{v}^{\mathcal{K}} \in \mathcal{T}^{\mathcal{K}}_{\mathbf{x}}\mathbb{K}^n$ with $g_{\mathbf{x}^{\mathcal{K}}}(\mathbf{v}^{\mathcal{K}}, \mathbf{v}^{\mathcal{K}}) = 1$. Let $\gamma(t) = \gamma_{\mathbf{x}^{\mathcal{K}},\mathbf{v}^{\mathcal{K}}}(t) \in \mathbb{K}^n$ denote the unit-speed geodesic in $\mathbb{K}^n$ with $\gamma_{\mathbf{x}^{\mathcal{K}},\mathbf{v}^{\mathcal{K}}}(0) = \mathbf{x}^{\mathcal{K}}$ and $\dot{\gamma}_{\mathbf{x}^{\mathcal{K}},\mathbf{v}^{\mathcal{K}}}(0) = \mathbf{v}^{\mathcal{K}}$. Suppose $\mathbf{x}^{\mathcal{L}}$ and $\mathbf{x}^{\mathcal{K}}$ are corresponding points in $\mathbb{L}^n$ and $\mathbf{v}^{\mathcal{L}}$ and $\mathbf{v}^{\mathcal{K}}$ are corresponding vectors in their tangent spaces. Then

$$\mathbf{x}^{\mathcal{L}} = \pi_{\mathcal{K}\to\mathcal{L}}(\mathbf{x}^{\mathcal{K}}) = \frac{1}{\sqrt{1 - \|\mathbf{x}^{\mathcal{K}}\|^2}} \begin{bmatrix} 1 \\ \mathbf{x}^{\mathcal{K}} \end{bmatrix} = \lambda_{\mathbf{x}^{\mathcal{K}}} \begin{bmatrix} 1 \\ \mathbf{x}^{\mathcal{K}} \end{bmatrix}, \tag{50}$$

$$\mathbf{v}^{\mathcal{L}} = \dot{\phi}(0) = \left.\frac{\partial \pi_{\mathcal{K}\to\mathcal{L}}(\mathbf{y}^{\mathcal{K}})}{\partial \mathbf{y}^{\mathcal{K}}}\right|_{\gamma(0)} \dot{\gamma}(0) = \frac{\partial \pi_{\mathcal{K}\to\mathcal{L}}(\mathbf{x}^{\mathcal{K}})}{\partial \mathbf{x}^{\mathcal{K}}} \mathbf{v}^{\mathcal{K}}. \tag{51}$$

With

$$\frac{\partial \lambda_{\mathbf{x}^{\mathcal{K}}}}{\partial \mathbf{x}^{\mathcal{K}}} = \frac{\mathbf{x}^{\mathcal{K}}}{(1 - \|\mathbf{x}^{\mathcal{K}}\|^2)^{3/2}} = \lambda^3_{\mathbf{x}^{\mathcal{K}}} \mathbf{x}^{\mathcal{K}}, \tag{52}$$

it follows that

$$\begin{aligned} \mathbf{v}^{\mathcal{L}} = \frac{\partial \pi_{\mathcal{K}\to\mathcal{L}}(\mathbf{x}^{\mathcal{K}})}{\partial \mathbf{x}^{\mathcal{K}}} \mathbf{v}^{\mathcal{K}} &= \left(\frac{\partial}{\partial \mathbf{x}^{\mathcal{K}}} \left(\begin{bmatrix} \lambda_{\mathbf{x}^{\mathcal{K}}} \\ \lambda_{\mathbf{x}^{\mathcal{K}}} \mathbf{x}^{\mathcal{K}} \end{bmatrix} \right) \right) \mathbf{v}^{\mathcal{K}} = \begin{bmatrix} \frac{\partial \lambda_{\mathbf{x}^{\mathcal{K}}}}{\partial \mathbf{x}^{\mathcal{K}}} \\ \frac{\partial \lambda_{\mathbf{x}^{\mathcal{K}}}}{\partial \mathbf{x}^{\mathcal{K}}} (\mathbf{x}^{\mathcal{K}})^\top + \lambda_{\mathbf{x}^{\mathcal{K}}} \mathbf{I} \end{bmatrix} \mathbf{v}^{\mathcal{K}} \\ &= \begin{bmatrix} \lambda^3_{\mathbf{x}^{\mathcal{K}}} \left(\mathbf{x}^{\mathcal{K}} \cdot \mathbf{v}^{\mathcal{K}} \right) \\ \lambda^3_{\mathbf{x}^{\mathcal{K}}} \left(\mathbf{x}^{\mathcal{K}} \cdot \mathbf{v}^{\mathcal{K}} \right) \mathbf{x}^{\mathcal{K}} + \lambda_{\mathbf{x}^{\mathcal{K}}} \mathbf{v}^{\mathcal{K}} \end{bmatrix}. \end{aligned} \tag{53}$$

Making substitutions in Equation (49) yields

$$\phi(t) = \lambda_{\mathbf{x}^{\mathcal{K}}} \begin{bmatrix} 1 \\ \mathbf{x}^{\mathcal{K}} \end{bmatrix} \cosh(t) + \begin{bmatrix} \lambda^3_{\mathbf{x}^{\mathcal{K}}} \left(\mathbf{x}^{\mathcal{K}} \cdot \mathbf{v}^{\mathcal{K}} \right) \\ \lambda^3_{\mathbf{x}^{\mathcal{K}}} \left(\mathbf{x}^{\mathcal{K}} \cdot \mathbf{v}^{\mathcal{K}} \right) \mathbf{x}^{\mathcal{K}} + \lambda_{\mathbf{x}^{\mathcal{K}}} \mathbf{v}^{\mathcal{K}} \end{bmatrix} \sinh(t). \tag{54}$$

Since the geodesics are preserved by the isometric mapping, γ is given by

$$\begin{aligned} \gamma_{\mathbf{x}^{\mathcal{K}},\mathbf{v}^{\mathcal{K}}}(t) &= \pi_{\mathcal{L}\to\mathcal{K}} \circ \phi(t) \\ &= \frac{\left(\lambda_{\mathbf{x}^{\mathcal{K}}} \cosh(t) + \lambda^3_{\mathbf{x}^{\mathcal{K}}} \left(\mathbf{x}^{\mathcal{K}} \cdot \mathbf{v}^{\mathcal{K}} \right) \sinh(t) \right) \mathbf{x}^{\mathcal{K}} + \left(\lambda_{\mathbf{x}^{\mathcal{K}}} \sinh(t) \right) \mathbf{v}^{\mathcal{K}}}{\lambda_{\mathbf{x}^{\mathcal{K}}} \cosh(t) + \lambda^3_{\mathbf{x}^{\mathcal{K}}} \left(\mathbf{x}^{\mathcal{K}} \cdot \mathbf{v}^{\mathcal{K}} \right) \sinh(t)} \\ &= \frac{\left(\cosh(t) + \lambda^2_{\mathbf{x}^{\mathcal{K}}} \left(\mathbf{x}^{\mathcal{K}} \cdot \mathbf{v}^{\mathcal{K}} \right) \sinh(t) \right) \mathbf{x}^{\mathcal{K}} + \sinh(t) \mathbf{v}^{\mathcal{K}}}{\cosh(t) + \lambda^2_{\mathbf{x}^{\mathcal{K}}} \left(\mathbf{x}^{\mathcal{K}} \cdot \mathbf{v}^{\mathcal{K}} \right) \sinh(t)} \\ &= \mathbf{x}^{\mathcal{K}} + \frac{\sinh(t) \mathbf{v}^{\mathcal{K}}}{\cosh(t) + \lambda^2_{\mathbf{x}^{\mathcal{K}}} \left(\mathbf{x}^{\mathcal{K}} \cdot \mathbf{v}^{\mathcal{K}} \right) \sinh(t)}. \end{aligned} \tag{55}$$

■

B.3. Proof of Proposition 5

Write $\mathbf{x}^{\mathcal{L}}$ as $\left[x_t, \mathbf{x}_s^\top\right]^\top$. Since $\mathbf{o}^{\mathcal{L}} = [1, 0, \cdots, 0]^\top$, we have $d_{\mathcal{L}}(\mathbf{o}^{\mathcal{L}}, \mathbf{x}^{\mathcal{L}}) = \cosh^{-1}(x_t)$. Then

$$\log_{\mathbf{o}^{\mathcal{L}}}(\mathbf{x}^{\mathcal{L}}) = d_{\mathcal{L}}(\mathbf{o}^{\mathcal{L}}, \mathbf{x}^{\mathcal{L}}) \frac{\mathbf{x}^{\mathcal{L}} + \langle \mathbf{o}^{\mathcal{L}}, \mathbf{x}^{\mathcal{L}} \rangle_{\mathcal{L}} \mathbf{o}^{\mathcal{L}}}{\|\mathbf{x}^{\mathcal{L}} + \langle \mathbf{o}^{\mathcal{L}}, \mathbf{x}^{\mathcal{L}} \rangle_{\mathcal{L}} \mathbf{o}^{\mathcal{L}}\|_{\mathcal{L}}} = \cosh^{-1}(x_t) \frac{\mathbf{x}^{\mathcal{L}} - x_t \mathbf{o}^{\mathcal{L}}}{\|\mathbf{x}^{\mathcal{L}} - x_t \mathbf{o}^{\mathcal{L}}\|_{\mathcal{L}}}. \tag{56}$$

Here, $\mathbf{x}^{\mathcal{L}} - x_t \mathbf{o}^{\mathcal{L}} = \left[x_t, \mathbf{x}_s^\top\right]^\top - \left[x_t, \mathbf{0}^\top\right]^\top = \left[0, \mathbf{x}_s^\top\right]^\top$ and thus $\|\mathbf{x}^{\mathcal{L}} - x_t \mathbf{o}^{\mathcal{L}}\|_{\mathcal{L}} = \|\mathbf{x}_s\|$. Therefore, $\log_{\mathbf{o}^{\mathcal{L}}}(\mathbf{x}^{\mathcal{L}}) = \frac{\cosh^{-1}(x_t)}{\|\mathbf{x}_s\|} \left[0, \mathbf{x}_s^\top\right]^\top$. Similarly,

$$\log_{\mathbf{x}^{\mathcal{L}}}(\mathbf{o}^{\mathcal{L}}) = d_{\mathcal{L}}(\mathbf{x}^{\mathcal{L}}, \mathbf{o}^{\mathcal{L}}) \frac{\mathbf{o}^{\mathcal{L}} + \langle \mathbf{x}^{\mathcal{L}}, \mathbf{o}^{\mathcal{L}} \rangle_{\mathcal{L}} \mathbf{x}^{\mathcal{L}}}{\|\mathbf{o}^{\mathcal{L}} + \langle \mathbf{x}^{\mathcal{L}}, \mathbf{o}^{\mathcal{L}} \rangle_{\mathcal{L}} \mathbf{x}^{\mathcal{L}}\|_{\mathcal{L}}} = \cosh^{-1}(x_t) \frac{\mathbf{o}^{\mathcal{L}} - x_t \mathbf{x}^{\mathcal{L}}}{\|\mathbf{o}^{\mathcal{L}} - x_t \mathbf{x}^{\mathcal{L}}\|_{\mathcal{L}}}. \tag{57}$$

Here, $\mathbf{o}^{\mathcal{L}} - x_t \mathbf{x}^{\mathcal{L}} = \left[1, \mathbf{0}^\top\right]^\top - x_t \left[x_t, \mathbf{x}_s^\top\right]^\top = \left[1 - x_t^2, x_t \mathbf{x}_s^\top\right]^\top$, and thus

$$\begin{aligned} \|\mathbf{o}^{\mathcal{L}} - x_t \mathbf{x}^{\mathcal{L}}\|_{\mathcal{L}} &= \sqrt{-\left(1 - x_t^2\right)^2 + x_t^2 \|\mathbf{x}_s\|^2} = \sqrt{-\left(-\|\mathbf{x}_s\|^2\right)^2 + x_t^2 \|\mathbf{x}_s\|^2} \\ &= \sqrt{\|\mathbf{x}_s\|^2 (x_t^2 - \|\mathbf{x}_s\|^2)} = \sqrt{\|\mathbf{x}_s\|^2} = \|\mathbf{x}_s\|. \end{aligned} \tag{58}$$

Therefore, $\log_{\mathbf{x}^{\mathcal{L}}}(\mathbf{o}^{\mathcal{L}}) = \frac{\cosh^{-1}(x_t)}{\|\mathbf{x}_s\|} \left[1 - x_t^2, x_t \mathbf{x}_s^\top\right]^\top$. Since $\mathbf{v}^{\mathcal{L}} \in \mathcal{T}_{\mathbf{o}^{\mathcal{L}}} \mathbb{L}^n$, it holds that $\langle \mathbf{o}, \mathbf{v}^{\mathcal{L}} \rangle_{\mathcal{L}} = 0$. Writing $\mathbf{v}^{\mathcal{L}}$ as $\left[v_t, \mathbf{v}_s^\top\right]^\top$, this indicates that $v_t = 0$. Hence,

$$\begin{aligned} P^{\mathcal{L}}_{\mathbf{o}^{\mathcal{L}} \to \mathbf{x}^{\mathcal{L}}}(\mathbf{v}^{\mathcal{L}}) &= \mathbf{v}^{\mathcal{L}} - \frac{\langle \log^{\mathcal{L}}_{\mathbf{o}}(\mathbf{x}^{\mathcal{L}}), \mathbf{v}^{\mathcal{L}} \rangle_{\mathcal{L}}}{d_{\mathcal{L}}(\mathbf{o}^{\mathcal{L}}, \mathbf{x}^{\mathcal{L}})^2} (\log^{\mathcal{L}}_{\mathbf{o}^{\mathcal{L}}}(\mathbf{x}^{\mathcal{L}}) + \log^{\mathcal{L}}_{\mathbf{x}^{\mathcal{L}}}(\mathbf{o}^{\mathcal{L}})) \\ &= \begin{bmatrix} 0 \\ \mathbf{v}_s \end{bmatrix} - \frac{\frac{\cosh^{-1}(x_t)}{\|\mathbf{x}_s\|} \mathbf{x}_s \cdot \mathbf{v}_s}{(\cosh^{-1}(x_t))^2} \frac{\cosh^{-1}(x_t)}{\|\mathbf{x}_s\|} \left(\begin{bmatrix} 0 \\ \mathbf{x}_s \end{bmatrix} + \begin{bmatrix} 1 - x_t^2 \\ x_t \mathbf{x}_s \end{bmatrix} \right) \\ &= \begin{bmatrix} 0 \\ \mathbf{v}_s \end{bmatrix} - \frac{\mathbf{x}_s \cdot \mathbf{v}_s}{\|\mathbf{x}_s\|^2} \begin{bmatrix} 1 - x_t^2 \\ (1 + x_t) \mathbf{x}_s \end{bmatrix} = \begin{bmatrix} \mathbf{x}_s \cdot \mathbf{v}_s \\ \mathbf{v}_s - \frac{\mathbf{x}_s \cdot \mathbf{v}_s}{(x_t - 1)} \mathbf{x}_s \end{bmatrix}. \end{aligned} \tag{59}$$

Next, we use the isometric mapping in Equation (2) as well as Equations (6) and (7) to derive $P_{\mathbf{o}^{\mathcal{K}} \to \mathbf{x}^{\mathcal{K}}}(\mathbf{v}^{\mathcal{K}})$. For clarity, denote $P_{\mathbf{o}^{\mathcal{L}} \to \mathbf{x}^{\mathcal{L}}}(\mathbf{v}^{\mathcal{L}}) = \mathbf{w}^{\mathcal{L}} = \left[w_t, \mathbf{w}_s^\top\right]^\top \in \mathcal{T}_{\mathbf{x}^{\mathcal{L}}} \mathbb{L}^n$. Then, by Lemma 1, the corresponding tangent vector in $\mathcal{T}_{\mathbf{x}^{\mathcal{K}}} \mathbb{K}^n$ is given by

$$\mathbf{w}^{\mathcal{K}} = -\frac{w_t}{x_t^2} \mathbf{x}_s + \frac{1}{x_t} \mathbf{w}_s = -\frac{\mathbf{x}_s \cdot \mathbf{v}_s}{x_t^2} \mathbf{x}_s + \frac{1}{x_t} \left(\mathbf{v}_s - \frac{\mathbf{x}_s \cdot \mathbf{v}_s}{(x_t - 1)} \mathbf{x}_s \right). \tag{60}$$

Expressing $x_t, \mathbf{x}_s$ in terms of $\mathbf{x}^{\mathcal{K}}$, we have

$$x_t = \frac{1}{\sqrt{1 - \|\mathbf{x}^{\mathcal{K}}\|^2}}, \quad \mathbf{x}_s = \frac{1}{\sqrt{1 - \|\mathbf{x}^{\mathcal{K}}\|^2}} \mathbf{x}^{\mathcal{K}}. \tag{61}$$

We further express $\mathbf{v}^{\mathcal{L}} \in \mathcal{T}_{\mathbf{o}^{\mathcal{L}}} \mathbb{L}^n$ in terms of the corresponding tangent vector $\mathbf{v}^{\mathcal{K}}$ in $\mathcal{T}_{\mathbf{o}^{\mathcal{K}}} \mathbb{K}^n$. We already know that $v_t = 0$. Also,

$$\mathbf{v}_s = \frac{1}{\sqrt{1 - \|\mathbf{o}^{\mathcal{K}}\|^2}} \mathbf{v}^{\mathcal{K}} + \frac{\mathbf{o}^{\mathcal{K}} \cdot \mathbf{v}^{\mathcal{K}}}{(1 - \|\mathbf{o}^{\mathcal{K}}\|^2)^{\frac{3}{2}}} \mathbf{o}^{\mathcal{K}} = \mathbf{v}^{\mathcal{K}}. \tag{62}$$

Hence,

$$
\begin{aligned}
P_{\mathbf{o}^{\mathcal{L}}\to\mathbf{x}^{\mathcal{L}}}(\mathbf{v}^{\mathcal{L}}) = \mathbf{w}^{\mathcal{K}} &= -\frac{\mathbf{x}_s\cdot\mathbf{v}_s}{x_t^2}\mathbf{x}_s + \frac{1}{x_t}\left(\mathbf{v}_s - \frac{\mathbf{x}_s\cdot\mathbf{v}_s}{(x_t-1)}\mathbf{x}_s\right)\\
&= -\left(1-\|\mathbf{x}^{\mathcal{K}}\|^2\right)\frac{\mathbf{x}^{\mathcal{K}}\cdot\mathbf{v}^{\mathcal{K}}}{\sqrt{1-\|\mathbf{x}^{\mathcal{K}}\|^2}}\frac{1}{\sqrt{1-\|\mathbf{x}^{\mathcal{K}}\|^2}}\mathbf{x}^{\mathcal{K}}\\
&\quad + \sqrt{1-\|\mathbf{x}^{\mathcal{K}}\|^2}\left(\mathbf{v}^{\mathcal{K}} - \frac{1}{\frac{1}{\sqrt{1-\|\mathbf{x}^{\mathcal{K}}\|^2}}-1}\frac{\mathbf{x}^{\mathcal{K}}\cdot\mathbf{v}^{\mathcal{K}}}{\sqrt{1-\|\mathbf{x}^{\mathcal{K}}\|^2}}\frac{1}{\sqrt{1-\|\mathbf{x}^{\mathcal{K}}\|^2}}\mathbf{x}^{\mathcal{K}}\right)\\
&= -(\mathbf{x}^{\mathcal{K}}\cdot\mathbf{v}^{\mathcal{K}})\mathbf{x}^{\mathcal{K}} + \sqrt{1-\|\mathbf{x}^{\mathcal{K}}\|^2}\mathbf{v}^{\mathcal{K}} - \frac{\mathbf{x}^{\mathcal{K}}\cdot\mathbf{v}^{\mathcal{K}}}{1-\sqrt{1-\|\mathbf{x}^{\mathcal{K}}\|^2}}\mathbf{x}^{\mathcal{K}}\\
&= \frac{(\mathbf{x}^{\mathcal{K}}\cdot\mathbf{v}^{\mathcal{K}})(\sqrt{1-\|\mathbf{x}^{\mathcal{K}}\|^2}-2)}{1-\sqrt{1-\|\mathbf{x}^{\mathcal{K}}\|^2}}\mathbf{x}^{\mathcal{K}} + \sqrt{1-\|\mathbf{x}^{\mathcal{K}}\|^2}\mathbf{v}^{\mathcal{K}}. \qquad (63)
\end{aligned}
$$

■

B.4. Proof of Theorem 6

Recall that Corollary 4 indicates

$$\exp^{\mathcal{K}}_{\mathbf{o}^{\mathcal{K}}}(\mathbf{v}^{\mathcal{K}}) = \tanh(\|\mathbf{v}^{\mathcal{K}}\|)\frac{\mathbf{v}^{\mathcal{K}}}{\|\mathbf{v}^{\mathcal{K}}\|}, \qquad (64)$$

$$\log^{\mathcal{K}}_{\mathbf{o}^{\mathcal{K}}}(\mathbf{x}^{\mathcal{K}}) = \cosh^{-1}\left(\frac{1}{\sqrt{1-\|\mathbf{x}^{\mathcal{K}}\|^2}}\right)\frac{\mathbf{x}^{\mathcal{K}}}{\|\mathbf{x}^{\mathcal{K}}\|}. \qquad (65)$$

On the one hand,

$$
\begin{aligned}
\exp^{\mathcal{K}}_{\mathbf{o}^{\mathcal{K}}}(r\log^{\mathcal{K}}_{\mathbf{o}^{\mathcal{K}}}(\mathbf{x}^{\mathcal{K}})) &= \exp^{\mathcal{K}}_{\mathbf{o}^{\mathcal{K}}}\left(r\cosh^{-1}\left(\frac{1}{\sqrt{1-\|\mathbf{x}^{\mathcal{K}}\|^2}}\right)\frac{\mathbf{x}^{\mathcal{K}}}{\|\mathbf{x}^{\mathcal{K}}\|}\right)\\
&= \tanh\left(|r|\cosh^{-1}\left(\frac{1}{\sqrt{1-\|\mathbf{x}^{\mathcal{K}}\|^2}}\right)\right)\frac{r\mathbf{x}^{\mathcal{K}}}{|r|\|\mathbf{x}^{\mathcal{K}}\|} \qquad (66)\\
&= \tanh\left(r\cosh^{-1}\left(\frac{1}{\sqrt{1-\|\mathbf{x}^{\mathcal{K}}\|^2}}\right)\right)\frac{\mathbf{x}^{\mathcal{K}}}{\|\mathbf{x}^{\mathcal{K}}\|}.
\end{aligned}
$$

On the other hand,

$$r\otimes_{\mathrm{E}}\mathbf{x}^{\mathcal{K}} = \tanh\left(r\tanh^{-1}\left(\|\mathbf{x}^{\mathcal{K}}\|\right)\right)\frac{\mathbf{x}^{\mathcal{K}}}{\|\mathbf{x}^{\mathcal{K}}\|}. \qquad (67)$$

Therefore, to prove $r\otimes_{\mathrm{E}}\mathbf{x}^{\mathcal{K}} = \exp^{\mathcal{K}}_{\mathbf{o}^{\mathcal{K}}}\left(r\log^{\mathcal{K}}_{\mathbf{o}^{\mathcal{K}}}\left(\mathbf{x}^{\mathcal{K}}\right)\right)$, it suffices to show:

$$\cosh^{-1}\left(\frac{1}{\sqrt{1-\|\mathbf{x}^{\mathcal{K}}\|^2}}\right) = \tanh^{-1}\left(\|\mathbf{x}^{\mathcal{K}}\|\right). \qquad (68)$$

Let $\cosh^{-1}\left(\frac{1}{\sqrt{1-\|\mathbf{x}^{\mathcal{K}}\|^2}}\right) = t$, then $\cosh(t) = \frac{1}{\sqrt{1-\|\mathbf{x}^{\mathcal{K}}\|^2}}$, i.e.,

$$\frac{e^t + e^{-t}}{2} = \frac{1}{\sqrt{1-\|\mathbf{x}^{\mathcal{K}}\|^2}}. \tag{69}$$

This implies that

$$\|\mathbf{x}^{\mathcal{K}}\|^2 = 1 - \frac{4}{e^{2t}+e^{-2t}+2} = \frac{e^{2t}+e^{-2t}-2}{e^{2t}+e^{-2t}+2} = \left(\frac{e^t-e^{-t}}{e^t+e^{-t}}\right)^2. \tag{70}$$

Therefore, $\|\mathbf{x}^{\mathcal{K}}\| = \frac{e^t-e^{-t}}{e^t+e^{-t}}$ since $t > 0$, i.e. $\tanh^{-1}\left(\|\mathbf{x}^{\mathcal{K}}\|\right) = t$, completing the proof. ■

B.5. Proof of Theorem 7

Our proof is based on Equation (28) and the isometric mappings in Equations (3)–(5).

Note that proving Equation (18) is equivalent to showing that $P_{\mathbf{o}^{\mathcal{K}}\to\mathbf{x}^{\mathcal{K}}}(\mathbf{v}^{\mathcal{K}})$ given by $\log^{\mathcal{K}}_{\mathbf{x}^{\mathcal{K}}}\left(\mathbf{x}^{\mathcal{K}} \oplus_{\mathrm{E}} \exp^{\mathcal{K}}_{\mathbf{o}^{\mathcal{K}}}(\mathbf{v}^{\mathcal{K}})\right)$ and $P_{\mathbf{o}^{\mathcal{B}}\to\mathbf{x}^{\mathcal{B}}}(\mathbf{v}^{\mathcal{B}})$ given by $\log^{\mathcal{B}}_{\mathbf{x}^{\mathcal{B}}}\left(\mathbf{x}^{\mathcal{B}} \oplus_{\mathrm{M}} \exp^{\mathcal{B}}_{\mathbf{o}^{\mathcal{B}}}(\mathbf{v}^{\mathcal{B}})\right)$ are corresponding tangent vectors. This is further equivalent to showing that $\mathbf{x}^{\mathcal{K}} \oplus_{\mathrm{E}} \exp^{\mathcal{K}}_{\mathbf{o}^{\mathcal{K}}}(\mathbf{v}^{\mathcal{K}})$ and $\mathbf{x}^{\mathcal{B}} \oplus_{\mathrm{M}} \exp^{\mathcal{B}}_{\mathbf{o}^{\mathcal{B}}}(\mathbf{v}^{\mathcal{B}})$ are corresponding points in hyperbolic space.

On the one hand, applying Equation (26), we have

$$\begin{aligned}
&\mathbf{x}^{\mathcal{B}} \oplus_{\mathrm{M}} \exp^{\mathcal{B}}_{\mathbf{o}^{\mathcal{B}}}(\mathbf{v}^{\mathcal{B}}) = \mathbf{x}^{\mathcal{B}} \oplus_{\mathrm{M}} \left(\tanh\left(\|\mathbf{v}^{\mathcal{B}}\|\right)\frac{\mathbf{v}^{\mathcal{B}}}{\|\mathbf{v}^{\mathcal{B}}\|}\right) \\
&= \frac{\left(1 + \frac{2\tanh\left(\|\mathbf{v}^{\mathcal{B}}\|\right)}{\|\mathbf{v}^{\mathcal{B}}\|}\mathbf{x}^{\mathcal{B}}\cdot\mathbf{v}^{\mathcal{B}} + \tanh^2\left(\|\mathbf{v}^{\mathcal{B}}\|\right)\right)\mathbf{x}^{\mathcal{B}} + \left(1-\|\mathbf{x}^{\mathcal{B}}\|^2\right)\frac{\tanh\left(\|\mathbf{v}^{\mathcal{B}}\|\right)}{\|\mathbf{v}^{\mathcal{B}}\|}\mathbf{v}^{\mathcal{B}}}{1 + \frac{2\tanh\left(\|\mathbf{v}^{\mathcal{B}}\|\right)}{\|\mathbf{v}^{\mathcal{B}}\|}\mathbf{x}^{\mathcal{B}}\cdot\mathbf{v}^{\mathcal{B}} + \tanh^2\left(\|\mathbf{v}^{\mathcal{B}}\|\right)\|\mathbf{x}^{\mathcal{B}}\|^2} \\
&= \frac{\left(\|\mathbf{v}^{\mathcal{B}}\| + 2\tanh\left(\|\mathbf{v}^{\mathcal{B}}\|\right)\mathbf{x}^{\mathcal{B}}\cdot\mathbf{v}^{\mathcal{B}} + \tanh^2\left(\|\mathbf{v}^{\mathcal{B}}\|\right)\|\mathbf{v}^{\mathcal{B}}\|\right)\mathbf{x}^{\mathcal{B}} + \left(1-\|\mathbf{x}^{\mathcal{B}}\|^2\right)\tanh\left(\|\mathbf{v}^{\mathcal{B}}\|\right)\mathbf{v}^{\mathcal{B}}}{\|\mathbf{v}^{\mathcal{B}}\| + 2\tanh\left(\|\mathbf{v}^{\mathcal{B}}\|\right)\mathbf{x}^{\mathcal{B}}\cdot\mathbf{v}^{\mathcal{B}} + \tanh^2\left(\|\mathbf{v}^{\mathcal{B}}\|\right)\|\mathbf{x}^{\mathcal{B}}\|^2\|\mathbf{v}^{\mathcal{B}}\|}.
\end{aligned} \tag{71}$$

On the other hand, applying Equation (14), we have

$$\begin{aligned}
&\mathbf{x}^{\mathcal{K}} \oplus_{\mathrm{E}} \exp^{\mathcal{K}}_{\mathbf{o}^{\mathcal{K}}}(\mathbf{v}^{\mathcal{K}}) = \mathbf{x}^{\mathcal{K}} \oplus_{\mathrm{E}} \left(\tanh(\|\mathbf{v}^{\mathcal{K}}\|)\frac{\mathbf{v}^{\mathcal{K}}}{\|\mathbf{v}^{\mathcal{K}}\|}\right) \\
&= \frac{1}{1 + \frac{\tanh(\|\mathbf{v}^{\mathcal{K}}\|)}{\|\mathbf{v}^{\mathcal{K}}\|}\mathbf{x}^{\mathcal{K}}\cdot\mathbf{v}^{\mathcal{K}}}\left(\mathbf{x}^{\mathcal{K}} + \sqrt{1-\|\mathbf{x}^{\mathcal{K}}\|^2}\frac{\tanh(\|\mathbf{v}^{\mathcal{K}}\|)}{\|\mathbf{v}^{\mathcal{K}}\|}\mathbf{v}^{\mathcal{K}}\right. \\
&\quad \left. + \frac{\frac{1}{\sqrt{1-\|\mathbf{x}^{\mathcal{K}}\|^2}}}{1 + \frac{1}{\sqrt{1-\|\mathbf{x}^{\mathcal{K}}\|^2}}}\frac{\tanh(\|\mathbf{v}^{\mathcal{K}}\|)}{\|\mathbf{v}^{\mathcal{K}}\|}(\mathbf{x}^{\mathcal{K}}\cdot\mathbf{v}^{\mathcal{K}})\mathbf{x}^{\mathcal{K}}\right) \\
&= \frac{\|\mathbf{v}^{\mathcal{K}}\|}{\|\mathbf{v}^{\mathcal{K}}\| + \tanh(\|\mathbf{v}^{\mathcal{K}}\|)\mathbf{x}^{\mathcal{K}}\cdot\mathbf{v}^{\mathcal{K}}}\left(\mathbf{x}^{\mathcal{K}} + \sqrt{1-\|\mathbf{x}^{\mathcal{K}}\|^2}\frac{\tanh(\|\mathbf{v}^{\mathcal{K}}\|)}{\|\mathbf{v}^{\mathcal{K}}\|}\mathbf{v}^{\mathcal{K}}\right. \\
&\quad \left. + \frac{1}{1 + \sqrt{1-\|\mathbf{x}^{\mathcal{K}}\|^2}}\frac{\tanh(\|\mathbf{v}^{\mathcal{K}}\|)}{\|\mathbf{v}^{\mathcal{K}}\|}(\mathbf{x}^{\mathcal{K}}\cdot\mathbf{v}^{\mathcal{K}})\mathbf{x}^{\mathcal{K}}\right)
\end{aligned}$$

$$= \frac{\|\mathbf{v}^{\mathcal{K}}\|}{\|\mathbf{v}^{\mathcal{K}}\| + \tanh(\|\mathbf{v}^{\mathcal{K}}\|)\mathbf{x}^{\mathcal{K}} \cdot \mathbf{v}^{\mathcal{K}}} \mathbf{x}^{\mathcal{K}} + \frac{\tanh(\|\mathbf{v}^{\mathcal{K}}\|)\mathbf{x}^{\mathcal{K}} \cdot \mathbf{v}^{\mathcal{K}}}{\|\mathbf{v}^{\mathcal{K}}\| + \tanh(\|\mathbf{v}^{\mathcal{K}}\|)\mathbf{x}^{\mathcal{K}} \cdot \mathbf{v}^{\mathcal{K}}} \frac{\mathbf{x}^{\mathcal{K}}}{1 + \sqrt{1 - \|\mathbf{x}^{\mathcal{K}}\|^2}}$$
$$+ \frac{\tanh(\|\mathbf{v}^{\mathcal{K}}\|)\sqrt{1 - \|\mathbf{x}^{\mathcal{K}}\|^2}}{\|\mathbf{v}^{\mathcal{K}}\| + \tanh(\|\mathbf{v}^{\mathcal{K}}\|)\mathbf{x}^{\mathcal{K}} \cdot \mathbf{v}^{\mathcal{K}}} \mathbf{v}^{\mathcal{K}}.$$

Given $\mathbf{x}^{\mathcal{B}}, \mathbf{v}^{\mathcal{B}}$, the corresponding $\mathbf{x}^{\mathcal{K}} \in \mathbb{K}^n, \mathbf{v}^{\mathcal{K}} \in \mathcal{T}_{\mathbf{o}^{\mathcal{K}}}\mathbb{K}^n$ are

$$\mathbf{x}^{\mathcal{K}} = \frac{2}{1 + \|\mathbf{x}^{\mathcal{B}}\|^2} \mathbf{x}^{\mathcal{B}}, \quad \mathbf{v}^{\mathcal{K}} = \frac{2}{1 + \|\mathbf{o}^{\mathcal{B}}\|^2} \mathbf{v}^{\mathcal{B}} - \frac{4\mathbf{o}^{\mathcal{B}} \cdot \mathbf{v}^{\mathcal{B}}}{(1 + \|\mathbf{o}^{\mathcal{B}}\|^2)^2} \mathbf{o}^{\mathcal{B}} = 2\mathbf{v}^{\mathcal{B}}. \tag{72}$$

Writing $\mathbf{x}^{\mathcal{K}} \oplus_{\mathrm{E}} \exp^{\mathcal{K}}_{\mathbf{o}^{\mathcal{K}}}(\mathbf{v}^{\mathcal{K}})$ in terms of $\mathbf{x}^{\mathcal{B}}, \mathbf{v}^{\mathcal{B}}$ yields

$$\begin{aligned}
&\mathbf{x}^{\mathcal{K}} \oplus_{\mathrm{E}} \exp^{\mathcal{K}}_{\mathbf{o}^{\mathcal{K}}}(\mathbf{v}^{\mathcal{K}}) \\
&= \frac{2\|\mathbf{v}^{\mathcal{B}}\|}{2\|\mathbf{v}^{\mathcal{B}}\| + \tanh(2\|\mathbf{v}^{\mathcal{B}}\|)\frac{2\mathbf{x}^{\mathcal{B}} \cdot 2\mathbf{v}^{\mathcal{B}}}{1 + \|\mathbf{x}^{\mathcal{B}}\|^2}} \frac{2\mathbf{x}^{\mathcal{B}}}{1 + \|\mathbf{x}^{\mathcal{B}}\|^2} + \frac{\tanh(2\|\mathbf{v}^{\mathcal{B}}\|)\frac{2\mathbf{x}^{\mathcal{B}} \cdot 2\mathbf{v}^{\mathcal{B}}}{1 + \|\mathbf{x}^{\mathcal{B}}\|^2}}{2\|\mathbf{v}^{\mathcal{B}}\| + \tanh(2\|\mathbf{v}^{\mathcal{B}}\|)\frac{2\mathbf{x}^{\mathcal{B}} \cdot 2\mathbf{v}^{\mathcal{B}}}{1 + \|\mathbf{x}^{\mathcal{B}}\|^2}} \mathbf{x}^{\mathcal{B}} \\
&\quad + \frac{\tanh(2\|\mathbf{v}^{\mathcal{B}}\|)\sqrt{1 - \frac{4\|\mathbf{x}^{\mathcal{B}}\|^2}{(1 + \|\mathbf{x}^{\mathcal{B}}\|^2)^2}}}{2\|\mathbf{v}^{\mathcal{B}}\| + \tanh(2\|\mathbf{v}^{\mathcal{B}}\|)\frac{2\mathbf{x}^{\mathcal{B}} \cdot 2\mathbf{v}^{\mathcal{B}}}{1 + \|\mathbf{x}^{\mathcal{B}}\|^2}} 2\mathbf{v}^{\mathcal{B}} \\
&= \frac{\|\mathbf{v}^{\mathcal{B}}\| \left(1 + \|\mathbf{x}^{\mathcal{B}}\|^2\right)}{\|\mathbf{v}^{\mathcal{B}}\| \left(1 + \|\mathbf{x}^{\mathcal{B}}\|^2\right) + 2\tanh(2\|\mathbf{v}^{\mathcal{B}}\|)\mathbf{x}^{\mathcal{B}} \cdot \mathbf{v}^{\mathcal{B}}} \frac{2\mathbf{x}^{\mathcal{B}}}{1 + \|\mathbf{x}^{\mathcal{B}}\|^2} \\
&\quad + \frac{2\tanh(2\|\mathbf{v}^{\mathcal{B}}\|)\mathbf{x}^{\mathcal{B}} \cdot \mathbf{v}^{\mathcal{B}}}{\|\mathbf{v}^{\mathcal{B}}\| \left(1 + \|\mathbf{x}^{\mathcal{B}}\|^2\right) + 2\tanh(2\|\mathbf{v}^{\mathcal{B}}\|)\mathbf{x}^{\mathcal{B}} \cdot \mathbf{v}^{\mathcal{B}}} \mathbf{x}^{\mathcal{B}} \\
&\quad + \frac{\tanh(2\|\mathbf{v}^{\mathcal{B}}\|)\left(1 - \|\mathbf{x}^{\mathcal{B}}\|^2\right)}{\|\mathbf{v}^{\mathcal{B}}\| \left(1 + \|\mathbf{x}^{\mathcal{B}}\|^2\right) + 2\tanh(2\|\mathbf{v}^{\mathcal{B}}\|)\mathbf{x}^{\mathcal{B}} \cdot \mathbf{v}^{\mathcal{B}}} \mathbf{v}^{\mathcal{B}} \\
&= \frac{2\|\mathbf{v}^{\mathcal{B}}\| + 2\tanh(2\|\mathbf{v}^{\mathcal{B}}\|)\mathbf{x}^{\mathcal{B}} \cdot \mathbf{v}^{\mathcal{B}}}{\|\mathbf{v}^{\mathcal{B}}\| \left(1 + \|\mathbf{x}^{\mathcal{B}}\|^2\right) + 2\tanh(2\|\mathbf{v}^{\mathcal{B}}\|)\mathbf{x}^{\mathcal{B}} \cdot \mathbf{v}^{\mathcal{B}}} \mathbf{x}^{\mathcal{B}} \\
&\quad + \frac{\tanh(2\|\mathbf{v}^{\mathcal{B}}\|)\left(1 - \|\mathbf{x}^{\mathcal{B}}\|^2\right)}{\|\mathbf{v}^{\mathcal{B}}\| \left(1 + \|\mathbf{x}^{\mathcal{B}}\|^2\right) + 2\tanh(2\|\mathbf{v}^{\mathcal{B}}\|)\mathbf{x}^{\mathcal{B}} \cdot \mathbf{v}^{\mathcal{B}}} \mathbf{v}^{\mathcal{B}} \\
&= \frac{2\|\mathbf{v}^{\mathcal{B}}\| + \frac{4\tanh(\|\mathbf{v}^{\mathcal{B}}\|)\mathbf{x}^{\mathcal{B}} \cdot \mathbf{v}^{\mathcal{B}}}{1 + \tanh^2(\|\mathbf{v}^{\mathcal{B}}\|)}}{\|\mathbf{v}^{\mathcal{B}}\| \left(1 + \|\mathbf{x}^{\mathcal{B}}\|^2\right) + \frac{4\tanh(\|\mathbf{v}^{\mathcal{B}}\|)\mathbf{x}^{\mathcal{B}} \cdot \mathbf{v}^{\mathcal{B}}}{1 + \tanh^2(\|\mathbf{v}^{\mathcal{B}}\|)}} \mathbf{x}^{\mathcal{B}} + \frac{\frac{2\tanh(\|\mathbf{v}^{\mathcal{B}}\|)\left(1 - \|\mathbf{x}^{\mathcal{B}}\|^2\right)}{1 + \tanh^2(\|\mathbf{v}^{\mathcal{B}}\|)}}{\|\mathbf{v}^{\mathcal{B}}\| \left(1 + \|\mathbf{x}^{\mathcal{B}}\|^2\right) + \frac{4\tanh(\|\mathbf{v}^{\mathcal{B}}\|)\mathbf{x}^{\mathcal{B}} \cdot \mathbf{v}^{\mathcal{B}}}{1 + \tanh^2(\|\mathbf{v}^{\mathcal{B}}\|)}} \mathbf{v}^{\mathcal{B}} \\
&= \frac{2\|\mathbf{v}^{\mathcal{B}}\| \left(1 + \tanh^2(\|\mathbf{v}^{\mathcal{B}}\|)\right) + 4\tanh(\|\mathbf{v}^{\mathcal{B}}\|)\mathbf{x}^{\mathcal{B}} \cdot \mathbf{v}^{\mathcal{B}}}{\|\mathbf{v}^{\mathcal{B}}\| \left(1 + \|\mathbf{x}^{\mathcal{B}}\|^2\right)\left(1 + \tanh^2(\|\mathbf{v}^{\mathcal{B}}\|)\right) + 4\tanh(\|\mathbf{v}^{\mathcal{B}}\|)\mathbf{x}^{\mathcal{B}} \cdot \mathbf{v}^{\mathcal{B}}} \mathbf{x}^{\mathcal{B}} \\
&\quad + \frac{2\tanh(\|\mathbf{v}^{\mathcal{B}}\|)\left(1 - \|\mathbf{x}^{\mathcal{B}}\|^2\right)}{\|\mathbf{v}^{\mathcal{B}}\| \left(1 + \|\mathbf{x}^{\mathcal{B}}\|^2\right)\left(1 + \tanh^2(\|\mathbf{v}^{\mathcal{B}}\|)\right) + 4\tanh(\|\mathbf{v}^{\mathcal{B}}\|)\mathbf{x}^{\mathcal{B}} \cdot \mathbf{v}^{\mathcal{B}}} \mathbf{v}^{\mathcal{B}} \\
&= \frac{\left(2\|\mathbf{v}^{\mathcal{B}}\| + 2\tanh^2(\|\mathbf{v}^{\mathcal{B}}\|)\|\mathbf{v}^{\mathcal{B}}\| + 4\tanh(\|\mathbf{v}^{\mathcal{B}}\|)\mathbf{x}^{\mathcal{B}} \cdot \mathbf{v}^{\mathcal{B}}\right)\mathbf{x}^{\mathcal{B}} + \left(2\tanh(\|\mathbf{v}^{\mathcal{B}}\| - 2\tanh(\|\mathbf{v}^{\mathcal{B}}\|)\|\mathbf{x}^{\mathcal{B}}\|^2\right)\mathbf{v}^{\mathcal{B}}}{\|\mathbf{v}^{\mathcal{B}}\| + \|\mathbf{x}^{\mathcal{B}}\|^2\|\mathbf{v}^{\mathcal{B}}\| + \tanh^2(\|\mathbf{v}^{\mathcal{B}}\|)\|\mathbf{v}^{\mathcal{B}}\| + \tanh^2(\|\mathbf{v}^{\mathcal{B}}\|)\|\mathbf{x}^{\mathcal{B}}\|^2\|\mathbf{v}^{\mathcal{B}}\| + 4\tanh(\|\mathbf{v}^{\mathcal{B}}\|)\mathbf{x}^{\mathcal{B}} \cdot \mathbf{v}^{\mathcal{B}}}
\end{aligned}$$

Meanwhile, we need to map $\mathbf{x}^{\mathcal{B}} \oplus_{\mathrm{M}} \exp^{\mathcal{B}}_{\mathbf{o}^{\mathcal{B}}}(\mathbf{v}^{\mathcal{B}}) \in \mathbb{B}^n$ to $\mathbb{K}^n$ and check whether it matches the simplified result of $\mathbf{x}^{\mathcal{K}} \oplus_{\mathrm{E}} \exp^{\mathcal{K}}_{\mathbf{o}^{\mathcal{K}}}(\mathbf{v}^{\mathcal{K}})$. Denote

$$\|\mathbf{v}^{\mathcal{B}}\| + 2\tanh\left(\|\mathbf{v}^{\mathcal{B}}\|\right)\mathbf{x}^{\mathcal{B}} \cdot \mathbf{v}^{\mathcal{B}} + \tanh^2\left(\|\mathbf{v}^{\mathcal{B}}\|\right)\|\mathbf{x}^{\mathcal{B}}\|^2\|\mathbf{v}^{\mathcal{B}}\| = A, \tag{73}$$

$$\|\mathbf{v}^{\mathcal{B}}\| + 2\tanh\left(\|\mathbf{v}^{\mathcal{B}}\|\right)\mathbf{x}^{\mathcal{B}} \cdot \mathbf{v}^{\mathcal{B}} + \tanh^2\left(\|\mathbf{v}^{\mathcal{B}}\|\right)\|\mathbf{v}^{\mathcal{B}}\| = B, \tag{74}$$

$$\left(1 - \|\mathbf{x}^{\mathcal{B}}\|^2\right)\tanh\left(\|\mathbf{v}^{\mathcal{B}}\|\right) = C. \tag{75}$$

Then $\mathbf{x}^{\mathcal{B}} \oplus_{\mathrm{M}} \exp^{\mathcal{B}}_{\mathbf{o}^{\mathcal{B}}}(\mathbf{v}^{\mathcal{B}}) \in \mathbb{B}^n$ can be written as $\frac{B\mathbf{x}^{\mathcal{B}}+C\mathbf{v}^{\mathcal{B}}}{A}$. Thus, the corresponding point in $\mathbb{K}^n$ is given by

$$\begin{aligned}\frac{2\frac{B\mathbf{x}^{\mathcal{B}}+C\mathbf{v}^{\mathcal{B}}}{A}}{1+\|\frac{B\mathbf{x}^{\mathcal{B}}+C\mathbf{v}^{\mathcal{B}}}{A}\|^2} &= \frac{2AB\mathbf{x}^{\mathcal{B}} + 2AC\mathbf{v}^{\mathcal{B}}}{A^2 + \|B\mathbf{x}^{\mathcal{B}} + C\mathbf{v}^{\mathcal{B}}\|^2} \\ &= \frac{2AB\mathbf{x}^{\mathcal{B}} + 2AC\mathbf{v}^{\mathcal{B}}}{A^2 + B^2\|\mathbf{x}^{\mathcal{B}}\|^2 + C^2\|\mathbf{v}^{\mathcal{B}}\|^2 + 2BC(\mathbf{x}^{\mathcal{B}} \cdot \mathbf{v}^{\mathcal{B}})}.\end{aligned} \tag{76}$$

For clarity, let $\|\mathbf{x}^{\mathcal{B}}\| = x, \|\mathbf{v}^{\mathcal{B}}\| = v$, and $\mathbf{x}^{\mathcal{B}} \cdot \mathbf{v}^{\mathcal{B}} = xv\cos\theta$. We calculate:

$$\begin{aligned}A^2 &= (v + 2xv\cos(\theta)\tanh(v) + x^2 v\tanh^2(v))^2 \\ &= v^2 + 4xv^2\cos(\theta)\tanh(v) + 2x^2v^2\tanh^2(v) \\ &\quad + 4x^2v^2\cos^2(\theta)\tanh^2(v) + 4x^3v^2\cos(\theta)\tanh^3(v) + x^4v^2\tanh^4(v),\end{aligned} \tag{77}$$

$$\begin{aligned}B^2\|\mathbf{x}^{\mathcal{B}}\|^2 &= \left(v + 2xv\cos(\theta)\tanh(v) + v\tanh^2(v)\right)^2 x^2 \\ &= x^2v^2 + 4x^3v^2\cos(\theta)\tanh(v) + 2x^2v^2\tanh^2(v) \\ &\quad + 4x^4v^2\cos^2(\theta)\tanh^2(v) + 4x^3v^2\cos(\theta)\tanh^3(v) + x^2v^2\tanh^4(v),\end{aligned} \tag{78}$$

$$C^2\|\mathbf{v}^{\mathcal{B}}\|^2 = \left(1-x^2\right)^2 v^2\tanh^2(v) = v^2\tanh^2(v) - 2x^2v^2\tanh^2(v) + x^4v^2\tanh^2(v), \tag{79}$$

$$\begin{aligned}2BC(\mathbf{x}^{\mathcal{B}} \cdot \mathbf{v}^{\mathcal{B}}) &= 2xv\cos(\theta)\left(v + 2xv\cos(\theta)\tanh(v) + v\tanh^2(v)\right)\left(1-x^2\right)\tanh(v) \\ &= 2xv^2\cos(\theta)\tanh(v) - 2x^3v^2\cos(\theta)\tanh(v) + 4x^2v^2\cos^2(\theta)\tanh^2(v) \\ &\quad - 4x^4v^2\cos^2(\theta)\tanh^2(v) + 2xv^2\cos(\theta)\tanh^3(v) - 2x^3v^2\cos(\theta)\tanh^3(v).\end{aligned} \tag{80}$$

Thus, the denominator in Equation (76) is

$$\begin{aligned}&A^2 + B^2\|\mathbf{x}^{\mathcal{B}}\|^2 + C^2\|\mathbf{v}^{\mathcal{B}}\|^2 + 2BC(\mathbf{x}^{\mathcal{B}} \cdot \mathbf{v}^{\mathcal{B}}) \\ &= v^2 + x^2v^2 + 6xv^2\cos(\theta)\tanh(v) + 2x^3v^2\cos(\theta)\tanh(v) \\ &\quad + v^2\tanh^2(v) + 2x^2v^2\tanh^2(v) + x^4v^2\tanh^2(v) + 8x^2v^2\cos^2(\theta)\tanh^2(v) \\ &\quad + 2xv^2\cos(\theta)\tanh^3(v) + 6x^3v^2\cos(\theta)\tanh^3(v) + x^2v^2\tanh^4(v) + x^4v^2\tanh^4(v).\end{aligned} \tag{81}$$

We can factor this expression and get

$$\begin{aligned}&A^2 + B^2\|\mathbf{x}^{\mathcal{B}}\|^2 + C^2\|\mathbf{v}^{\mathcal{B}}\|^2 + 2BC(\mathbf{x}^{\mathcal{B}} \cdot \mathbf{v}^{\mathcal{B}}) \\ &= v^2\left(1 + 2x\cos(\theta)\tanh(v) + x^2\tanh^2(v)\right) \\ &\quad \cdot \left(1 + x^2 + 4x\cos(\theta)\tanh(v) + \tanh^2(v) + x^2\tanh^2(v)\right).\end{aligned} \tag{82}$$

The numerator contains the following terms:

$$2AB = 2\left(v + 2xv\cos(\theta)\tanh(v) + x^2v\tanh^2(v)\right)\left(v + 2xv\cos(\theta)\tanh(v) + v\tanh^2(v)\right)$$

$$= 2v^2\left(1+2x\cos(\theta)\tanh(v)+x^2\tanh^2(v)\right)\left(1+2x\cos(\theta)\tanh(v)+\tanh^2(v)\right), \quad (83)$$

$$2AC = 2\left(v+2xv\cos(\theta)\tanh(v)+x^2v\tanh^2(v)\right)\left(1-x^2\right)\tanh(v)$$
$$= 2v\left(1+2x\cos(\theta)\tanh(v)+x^2\tanh^2(v)\right)\left(\tanh(v)-x^2\tanh(v)\right). \quad (84)$$

Altogether, Equation (76) can be simplified as

$$\frac{2AB\mathbf{x}^{\mathcal{B}}+2AC\mathbf{v}^{\mathcal{B}}}{A^2+B^2\|\mathbf{x}^{\mathcal{B}}\|^2+C^2\|\mathbf{v}^{\mathcal{B}}\|^2+2BC(\mathbf{x}^{\mathcal{B}}\cdot\mathbf{v}^{\mathcal{B}})}$$
$$= \frac{2v^2\left(1+2x\cos(\theta)\tanh(v)+x^2\tanh^2(v)\right)\left(1+2x\cos(\theta)\tanh(v)+\tanh^2(v)\right)}{v^2\left(1+2x\cos(\theta)\tanh(v)+x^2\tanh^2(v)\right)\left(1+x^2+4x\cos(\theta)\tanh(v)+\tanh^2(v)+x^2\tanh^2(v)\right)}\mathbf{x}^{\mathcal{B}}$$
$$+\frac{2v\left(1+2x\cos(\theta)\tanh(v)+x^2\tanh^2(v)\right)\left(\tanh(v)-x^2\tanh(v)\right)}{v^2\left(1+2x\cos(\theta)\tanh(v)+x^2\tanh^2(v)\right)\left(1+x^2+4x\cos(\theta)\tanh(v)+\tanh^2(v)+x^2\tanh^2(v)\right)}\mathbf{v}^{\mathcal{B}}$$
$$= \frac{2\left(1+2x\cos(\theta)\tanh(v)+\tanh^2(v)\right)}{1+x^2+4x\cos(\theta)\tanh(v)+\tanh^2(v)+x^2\tanh^2(v)}\mathbf{x}^{\mathcal{B}}$$
$$+\frac{2\left(\tanh(v)-x^2\tanh(v)\right)}{v\left(1+x^2+4x\cos(\theta)\tanh(v)+\tanh^2(v)+x^2\tanh^2(v)\right)}\mathbf{v}^{\mathcal{B}}. \quad (85)$$

If we express $\mathbf{x}^{\mathcal{K}}\oplus_{\mathrm{E}}\exp^{\mathcal{K}}_{\mathbf{o}^{\mathcal{K}}}(\mathbf{v}^{\mathcal{K}})$ in terms of x, v and $\cos(\theta)$, we get

$$\mathbf{x}^{\mathcal{K}}\oplus_{\mathrm{E}}\exp^{\mathcal{K}}_{\mathbf{o}^{\mathcal{K}}}(\mathbf{v}^{\mathcal{K}})$$
$$= \frac{\left(2\|\mathbf{v}^{\mathcal{B}}\|+2\tanh^2(\|\mathbf{v}^{\mathcal{B}}\|)\|\mathbf{v}^{\mathcal{B}}\|+4\tanh(\|\mathbf{v}^{\mathcal{B}}\|)\mathbf{x}^{\mathcal{B}}\cdot\mathbf{v}^{\mathcal{B}}\right)\mathbf{x}^{\mathcal{B}}+\left(2\tanh(\|\mathbf{v}^{\mathcal{B}}\|-2\tanh(\|\mathbf{v}^{\mathcal{B}}\|)\|\mathbf{x}^{\mathcal{B}}\|^2\right)\mathbf{v}^{\mathcal{B}}}{\|\mathbf{v}^{\mathcal{B}}\|+\|\mathbf{x}^{\mathcal{B}}\|^2\|\mathbf{v}^{\mathcal{B}}\|+\tanh^2(\|\mathbf{v}^{\mathcal{B}}\|)\|\mathbf{v}^{\mathcal{B}}\|+\tanh^2(\|\mathbf{v}^{\mathcal{B}}\|)\|\mathbf{x}^{\mathcal{B}}\|^2\|\mathbf{v}^{\mathcal{B}}\|+4\tanh(\|\mathbf{v}^{\mathcal{B}}\|)\mathbf{x}^{\mathcal{B}}\cdot\mathbf{v}^{\mathcal{B}}}$$
$$= \frac{2\left(v+2xv\cos(\theta)\tanh(v)+v\tanh^2(v)\right)\mathbf{x}^{\mathcal{B}}+2\left(\tanh(v)-x^2\tanh(v)\right)\mathbf{v}^{\mathcal{B}}}{v+x^2v+4xv\cos(\theta)\tanh(v)+\tanh^2(v)v+x^2v\tanh^2(v)}$$
$$= \frac{2\left(1+2x\cos(\theta)\tanh(v)+\tanh^2(v)\right)}{1+x^2+4x\cos(\theta)\tanh(v)+\tanh^2(v)+x^2\tanh^2(v)}\mathbf{x}^{\mathcal{B}}$$
$$+\frac{2\left(\tanh(v)-x^2\tanh(v)\right)}{v\left(1+x^2+4x\cos(\theta)\tanh(v)+\tanh^2(v)+x^2\tanh^2(v)\right)}\mathbf{v}^{\mathcal{B}}. \quad (86)$$

The expressions in Equation (85) and Equation (86) agree. Therefore, $\mathbf{x}^{\mathcal{K}}\oplus_{\mathrm{E}}\exp^{\mathcal{K}}_{\mathbf{o}^{\mathcal{K}}}(\mathbf{b}^{\mathcal{K}})$ and $\mathbf{x}^{\mathcal{B}}\oplus_{\mathrm{M}}\exp^{\mathcal{B}}_{\mathbf{o}^{\mathcal{B}}}(\mathbf{b}^{\mathcal{B}})$ are the corresponding points in the hyperbolic space. We conclude that

$$P_{\mathbf{o}^{\mathcal{K}}\to\mathbf{x}^{\mathcal{K}}}(\mathbf{v}^{\mathcal{K}}) = \log^{\mathcal{K}}_{\mathbf{x}^{\mathcal{K}}}\left(\mathbf{x}^{\mathcal{K}}\oplus_{\mathrm{E}}\exp^{\mathcal{K}}_{\mathbf{o}^{\mathcal{K}}}(\mathbf{v}^{\mathcal{K}})\right). \quad (87)$$

■

B.6. Proof of Theorem 9

A simple application of Equation (12) yields

$$\mathbf{M}^{\otimes_{\mathrm{E}}}(\mathbf{x}^{\mathcal{K}}) = \exp^{\mathcal{K}}_{\mathbf{o}^{\mathcal{K}}}(\mathbf{M}\log^{\mathcal{K}}_{\mathbf{o}^{\mathcal{K}}}(\mathbf{x}^{\mathcal{K}}))$$
$$= \tanh\left(\frac{2\|\mathbf{M}\mathbf{x}^{\mathcal{K}}\|}{\|\mathbf{x}^{\mathcal{K}}\|}\tanh^{-1}\left(\frac{\|\mathbf{x}^{\mathcal{K}}\|}{1+\sqrt{1-\|\mathbf{x}^{\mathcal{K}}\|^2}}\right)\right)\frac{\mathbf{M}\mathbf{x}^{\mathcal{K}}}{\|\mathbf{M}\mathbf{x}^{\mathcal{K}}\|}. \quad (88)$$

■

Appendix C. Statistics of Datasets

The following table summarizes the statistics of the datasets.

Table 3: Statistics of datasets

Dataset	**Texas**	**Wisconsin**	**Chameleon**	**Actor**	**Cora**	**Pubmed**
Nodes	183	251	2,277	7,600	2,708	19,717
Edges	280	466	31,421	26,752	5,278	44,327
Features	1,703	1,703	2,325	931	1,433	500
Class	5	5	5	5	7	3
Hyperbolicity	1.0	1.0	1.5	1.5	3.0	2.5