SRBGCN: Tangent space-Free Lorentz Transformations for Graph Feature Learning (Supplementary Material)

Abdelrahman Mostafa[1]
abdelrahman.mostafa@oulu.fi
Wei Peng[2]
wepeng@stanford.edu
Guoying Zhao[1,*]
guoying.zhao@oulu.fi

[1] Center for Machine Vision and Signal Analysis, University of Oulu, Oulu, Finland
[2] Department of Psychiatry and Behavioral Sciences, Stanford University, California, USA

A Hyperbolic Geometry Review

We give a quick review about hyperbolic geometry to make the paper self-contained.

A.1 Topological Space and Topological Homeomorphism

A topological space is a geometrical space which has the notion of closeness. The closeness can, but not necessarily, be measured by the notion of distance to determine if points are close to each other. A homeomorphism is a continuous one-to-one mapping function or a bicontinuous function between topological spaces that has a continuous inverse function.

A.2 Manifold and Tangent Space

A d-dimensional Manifold $\mathcal{M}^d$ (which can be embedded in R^{d+1}) is a topological space which can be locally approximated by a d-dimensional Euclidean space R^d. For any point $x \in \mathcal{M}^d$, there is a homeomorphism between the neighbourhood of x and the Euclidean space R^d. Lines and circles are examples of one-dimensional manifolds. Planes and spheres are examples of two-dimensional manifolds which are called surfaces. The notion of manifold is a generalization of surfaces in any dimension d. The tangent space $\mathcal{T}_x\mathcal{M}^d$ at point $x \in \mathcal{M}^d$ is a d-dimensional hyperplane which is embedded in R^{d+1} that locally approximates the manifold $\mathcal{M}^d$ around the point x.

*Corresponding author.

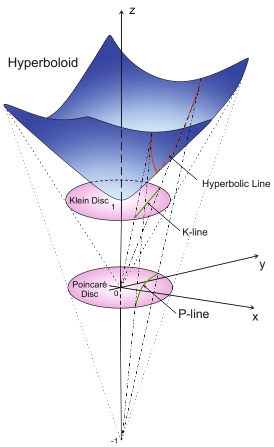

Figure 1: Projection of a hyperbolic geodesic from $H^{2,K}$ onto the Klein disk and the Poincaré disk.

A.3 Riemannian Metric and Riemannian Manifold

A Riemannian metric g is used to define geometric notions on the manifold such as distances, angles, areas or volumes. It is a collection of smoothly varying inner products on tangent spaces, $g_x : \mathcal{T}_x\mathcal{M}^d \times \mathcal{T}_x\mathcal{M}^d \to R$. A Riemannian manifold can then be defined as $(\mathcal{M}^d, g)$.

A.4 Curvature and Geodesics

A curvature measures how much a curve deviates from being a straight line. Euclidean spaces have zero curvature whereas non-Euclidean spaces have non-zero curvature. For example, spheres have constant positive curvatures whereas hyperbolic spaces have constant negative curvatures. Geodesics are the generalizations of shortest paths in graphs or lines in Euclidean geometry to non-Euclidean geometry. These are the curves that give the shortest paths between pairs of points.

A.5 Hyperbolic Space

A hyperbolic space is a Riemannian manifold with a constant negative curvature. Many models have been proposed to model a hyperbolic space such as the Lorentz model (also called the hyperboloid model), the Poincaré model and the Klein model. The Lorentz model is the upper sheet of a two-sheeted hyperboloid. The Poincaré model and the Klein model are the projections of the Lorentz model onto the hyperplanes $x_0 = 0$ and $x_0 = 1$, respectively. There are bijection functions to map between different hyperbolic models as they are all isomorphic. Figure 1 shows the three different models which model the hyperbolic space.

Let $\langle .,. \rangle_{\mathcal{L}} : R^{d+1} \times R^{d+1} \to R$ represents the Lorentz-Minkowski inner product where $\langle x, y \rangle_{\mathcal{L}} := \sum_{i=1}^{d} x_i y_i - x_0 y_0 = x^T g_{\mathcal{L}} y$ where $g_{\mathcal{L}} = diag(-1, 1, \ldots, 1)$ is a diagonal matrix that represents the Riemannian metric for the hyperbolic manifold. Let $H^{d,K}$ be a d dimensional hyperboloid model with a constant negative curvature $-1/K$ where $K > 0$. Then we have:

$$H^{d,K} := \{x \in R^{d+1} : \langle x, x \rangle_{\mathcal{L}} = -K, x_0 > 0\} \tag{1}$$

Note that $x_0 > 0$ to indicate the upper half of the hyperboloid manifold. In special relativity, x_0 is referred to as the time axis whereas the rest of axes are called the spatial axes.

A.6 Exponential and Logarithmic maps

The exponential and logarithmic maps are used to map between the hyperbolic space and the tangent space and represent a bijection between the tangent space at a point and the hyperboloid. The exponential map maps a point $v \in \mathcal{T}_x H^{d,K}$ where $x \in H^{d,K}$ to the hyperboloid $H^{d,K}$ such that $v \neq 0$ and is defined as:

$$exp_x^K(v) = \cosh(\frac{\|v\|_{\mathcal{L}}}{\sqrt{K}})x + \sqrt{K}\sinh(\frac{\|v\|_{\mathcal{L}}}{\sqrt{K}})\frac{v}{\|v\|_{\mathcal{L}}} \tag{2}$$

where $\|v\|_{\mathcal{L}} = \sqrt{\langle v, v \rangle_{\mathcal{L}}}$ is the norm of v. The logarithmic map maps a point $y \in H^{d,K}$ to the tangent space $\mathcal{T}_x H^{d,K}$ centered at point $x \in H^{d,K}$ such that $x \neq y$ and is defined as:

$$log_x^K(y) = d_{\mathcal{L}}^K(x,y)\frac{y + {}^1\!/\!_K\langle x, y \rangle_{\mathcal{L}} x}{\|y + {}^1\!/\!_K\langle x, y \rangle_{\mathcal{L}} x\|_{\mathcal{L}}} \tag{3}$$

where $d_{\mathcal{L}}^K(x,y)$ is the Minkowskian distance between two points x and y in $H^{d,K}$ and is given by:

$$d_{\mathcal{L}}^K(x,y) = \sqrt{K}\,\mathrm{arcosh}(-\langle x,y \rangle_{\mathcal{L}}/K) \tag{4}$$

B Lorentz Transformations

A Lorentz transformation matrix Λ should satisfy the following constraint:

$$\Lambda^T g_{\mathcal{L}} \Lambda = g_{\mathcal{L}} \tag{5}$$

where $\Lambda \in R^{(d+1)\times(d+1)}$ and T represents the transpose operation of the matrix.

B.1 Spatial Rotation Operation

Let $\mathbf{P} = \begin{bmatrix} 1 & 0 \\ 0 & \mathbf{Q} \end{bmatrix}$ to rotate the spatial coordinates and keep the time/first coordinate fixed. From Equation 5, we have:

$$\mathbf{P}^T g_{\mathcal{L}} \mathbf{P} = \begin{bmatrix} 1 & 0 \\ 0 & \mathbf{Q}^T \end{bmatrix} \begin{bmatrix} -1 & 0 \\ 0 & \mathbf{I} \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & \mathbf{Q} \end{bmatrix} = \begin{bmatrix} -1 & 0 \\ 0 & \mathbf{Q}^T \mathbf{I} \mathbf{Q} \end{bmatrix} = \begin{bmatrix} -1 & 0 \\ 0 & \mathbf{I} \end{bmatrix} = g_{\mathcal{L}}$$

So, we have $\mathbf{Q^T Q} = \mathbf{I}$ i.e. $\mathbf{Q}$ belongs to the special orthogonal group $SO(d)$.

Table 1: Datasets statistics

Dataset	nodes	node features	nodes classes	edges	δ-hyperbolicity
Disease	1044	1000	2	1043	0
Airport	3188	11	4	18631	1
PubMed	19717	500	3	88651	3.5
Cora	2708	1433	7	5429	11

B.2 Boost Operation

Let $\mathbf{L} = \begin{bmatrix} a & b_d^T \\ b_d & \mathbf{C}_{d\times d} \end{bmatrix}$ where $\mathbf{C} = \mathbf{C}^T$ as $\mathbf{L}$ is a symmetric matrix.
From Equation 5, we have:

$$\mathbf{L}^T g_{\mathcal{L}} \mathbf{L} = \begin{bmatrix} a & b_d^T \\ b_d & \mathbf{C}^T \end{bmatrix} \begin{bmatrix} -1 & 0 \\ 0 & \mathbf{I} \end{bmatrix} \begin{bmatrix} a & b_d^T \\ b_d & \mathbf{C} \end{bmatrix} = \begin{bmatrix} a & b_d^T \\ b_d & \mathbf{C} \end{bmatrix} \begin{bmatrix} -a & -b_d^T \\ b_d & \mathbf{C} \end{bmatrix}$$

$$= \begin{bmatrix} -a^2 + b_d \cdot b_d & -ab_d + \mathbf{C}b_d \\ -ab_d + \mathbf{C}b_d & -b_d \otimes b_d + \mathbf{C}^2 \end{bmatrix} = g_{\mathcal{L}} = \begin{bmatrix} -1 & 0 \\ 0 & \mathbf{I} \end{bmatrix}$$

So, we get $-a^2 + b_d \cdot b_d = -1$ and using the hyperbolic identity: $-\cosh^2\omega + \sinh^2\omega = -1$, we have $a = \cosh\omega$ and $b_d = (\sinh\omega)n_d$ where n_d is a unit vector. To solve for $\mathbf{C}$, we have $-b_d \otimes b_d + \mathbf{C}^2 = \mathbf{I}$. So, we get:

$$\mathbf{C}^2 = \mathbf{I} + (\sinh^2\omega)n_d \otimes n_d = \mathbf{I} + (-1 + \cosh^2\omega)n_d \otimes n_d$$

$$= \mathbf{I} + (1 - 2 + \cosh^2\omega + 2\cosh\omega - 2\cosh\omega)n_d \otimes n_d$$

$$= \mathbf{I} + (1 - \cosh\omega)^2 n_d \otimes n_d - 2(1 - \cosh\omega)n_d \otimes n_d = (\mathbf{I} - (1 - \cosh\omega)n_d \otimes n_d)^2$$

and using $-ab_d + \mathbf{C}b_d = 0$, we omit the negative solution and accept the positive one. So, we have $\mathbf{C} = \mathbf{I} - (1 - \cosh\omega)n_d \otimes n_d$ as the solution.

C Datasets Statistics and Hyperparameters Details

Table 1 shows the statistics for all datasets. The δ-hyperbolicity represents the Gromovs hyperbolicity and is reported on these datasets by [1]. The lower the δ-hyperbolicity value, the closer the graph to a tree i.e. more hyperbolic where a tree structure has a δ-hyperbolicity value of zero as the case for the Disease dataset. The higher the δ-hyperbolicity value, the closer the graph to a complete graph. The hyperparameters used for network training are shown in Table 2.

References

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

Table 2: Hyperparameters used for network training.

Dataset	Disease		Airport		PubMed		Cora	
Parameter	LP	NC	LP	NC	LP	NC	LP	NC
Learning rate	0.001	0.005	0.5	0.2	0.05	0.04	0.001	0.001
Number of layers	2	6	1	2	1	5	1	3
Weight decay	0.0	0.0	1e-05	0.0	0.0	0.01	1e-04	0.01
Dropout	0.0	0.0	0.0	0.6	0.5	0.8	0.7	0.9
Margin	2	2	0.1	2	0.1	1	0.1	2
Normalize features	0	1	1	0	1	1	1	1