HoroPCA: Hyperbolic Dimensionality Reduction via Horospherical Projections

Ines Chami [*1] **Albert Gu** [*1] **Dat Nguyen** [*1] **Christopher Ré** [1]

Abstract

This paper studies Principal Component Analysis (PCA) for data lying in hyperbolic spaces. Given directions, PCA relies on: (1) a parameterization of subspaces spanned by these directions, (2) a method of *projection* onto subspaces that preserves information in these directions, and (3) an *objective* to optimize, namely the variance explained by projections. We generalize each of these concepts to the hyperbolic space and propose HoroPCA, a method for hyperbolic dimensionality reduction. By focusing on the core problem of extracting *principal directions*, HoroPCA theoretically better preserves information in the original data such as distances, compared to previous generalizations of PCA. Empirically, we validate that HoroPCA outperforms existing dimensionality reduction methods, significantly reducing error in distance preservation. As a data whitening method, it improves downstream classification by up to 3.9% compared to methods that don't use whitening. Finally, we show that HoroPCA can be used to visualize hyperbolic data in two dimensions.

1. Introduction

Learning representations of data in hyperbolic spaces has recently attracted important interest in Machine Learning (ML) (Nickel & Kiela, 2017; Sala et al., 2018) due to their ability to represent hierarchical data with high fidelity in low dimensions (Sarkar, 2011). Many real-world datasets exhibit hierarchical structures, and hyperbolic embeddings have led to state-of-the-art results in applications such as question answering (Tay et al., 2018), node classification (Chami et al., 2019), link prediction (Balazevic et al., 2019; Chami et al., 2020b) and word embeddings (Tifrea et al., 2018). These developments motivate the need for algorithms that operate in hyperbolic spaces such as nearest neighbor search (Krauthgamer & Lee, 2006; Wu & Charikar, 2020), hierarchical clustering (Monath et al., 2019; Chami et al., 2020a), or dimensionality reduction which is the focus of this work.

Euclidean Principal Component Analysis (PCA) is a fundamental dimensionality reduction technique which seeks directions that best explain the original data. PCA is an important primitive in data analysis and has many important uses such as (i) dimensionality reduction (e.g. for memory efficiency), (ii) data whitening and pre-processing for downstream tasks, and (iii) data visualization.

Here, we seek a generalization of PCA to hyperbolic geometry. Given a core notion of **directions**, PCA involves the following ingredients:

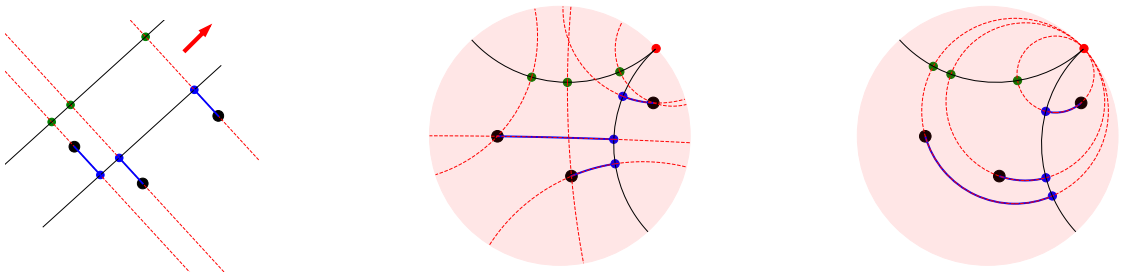
1. A nested sequence of affine **subspaces** (flags) spanned by a set of directions.
2. A **projection** method which maps points to these subspaces while preserving information (e.g. dot-product) along each direction.
3. A **variance** objective to help choose the best directions.

The PCA algorithm is then defined as combining these primitives: given a dataset, it chooses directions that maximize the *variance* of *projections* onto a *subspace* spanned by those directions, so that the resulting sequence of directions optimally explains the data. Crucially, the algorithm only depends on the directions of the *affine* subspaces and not their locations in space (Fig. 1a). Thus, in practice we can assume that they all go through the origin (and hence become *linear* subspaces), which greatly simplifies computations.

Generalizing PCA to manifolds is a challenging problem that has been studied for decades, starting with Principal Geodesic Analysis (PGA) (Fletcher et al., 2004) which parameterizes subspaces using tangent vectors at the mean of the data, and maximizes distances from projections to the mean to find optimal directions. More recently, the Barycentric Subspace Analysis (BSA) method (Pennec, 2018) was introduced. It finds a more general parameterization of nested sequences of submanifolds by minimizing the unexplained variance. However, both PGA and BSA map points onto submanifolds using closest-point or *geodesic projections*, which do not attempt to preserve information along principal directions; for example, they cannot isometrically

[*]Equal contribution [1]Stanford University, CA, USA. Correspondence to: Ines Chami <chami@cs.stanford.edu>.

Proceedings of the 38^{th} International Conference on Machine Learning, PMLR 139, 2021.



(a) Euclidean projections. (b) Hyperbolic geodesic projections. (c) Hyperbolic horospherical projections.

Figure 1: Given datapoints (black dots), Euclidean and horospherical projections preserve distance information across different subspaces (black lines) pointing towards the same direction or point at infinity, while geodesic projections do not. (a): Distances between points, and therefore the explained variance, are invariant to translations along orthogonal directions (red lines). (b): Geodesic projections do not preserve this property: distances between green projections are not the same as distances between blue projections. (c): Horospherical projections project data by sliding it along the complementary direction defined by horospheres (red circles) and there exist an isometric mapping between the blue and green projections.

preserve hyperbolic distances between *any* points and shrink all path lengths by an exponential factor (Proposition 3.5).

Fundamentally, all previous methods only look for *subspaces* rather than directions that explain the data, and can perhaps be better understood as principal *subspace* analysis rather than principal *component* analysis. Like with PCA, they assume that all optimal subspaces go through a chosen *base point*, but unlike in the Euclidean setting, this assumption is now unjustified: "translating" the submanifolds does not preserve distances between projections (Fig. 1b). Furthermore, the dependence on the base point is sensitive: as noted above, the shrink factor of the projection depends exponentially on the distances between the subspaces and the data. Thus, having to choose a base point increases the number of necessary parameters and reduces stability.

Here, we propose HoroPCA, a dimensionality reduction method for data defined in hyperbolic spaces which better preserves the properties of Euclidean PCA. We show how to interpret **directions** using *points at infinity* (or *ideal points*), which then allows us to generalize core properties of PCA:

1. To generalize the notion of affine **subspace**, we propose parameterizing geodesic subspaces as the sets spanned by these ideal points. This yields multiple viable nested subspaces (flags) (Section 3.1).
2. To maximally preserve information in the original data, we propose a new **projection** method that uses horospheres, a generalization of complementary directions for hyperbolic space. In contrast with geodesic projections, these projections exactly preserve information – specifically, distances to ideal points – along each direction. Consequently, they preserve distances between points much better than geodesic projections (Section 3.2).
3. Finally, we introduce a simple generalization of explained **variance** that is a function of distances only and can be computed in hyperbolic space (Section 3.3).

Combining these notions, we propose an algorithm that seeks a sequence of *principal components* that best explain variations in hyperbolic data. We show that this formulation retains the *location-independence* property of PCA: translating target submanifolds along orthogonal directions (horospheres) preserves projected distances (Fig. 1c). In particular, the algorithm's objective depends only on the directions and not locations of the submanifolds (Section 4).

We empirically validate HoroPCA on real datasets and for three standard PCA applications. First, (i) we show that it yields much lower distortion and higher explained variance than existing methods, reducing average distortion by up to 77%. Second, (ii) we validate that it can be used for data pre-processing, improving downstream classification by up to 3.8% in Average Precision score compared to methods that don't use whitening. Finally, (iii) we show that the low-dimensional representations learned by HoroPCA can be visualized to qualitatively interpret hyperbolic data.

2. Background

We first review some basic notions from hyperbolic geometry; a more in-depth treatment is available in standard texts (Lee, 2013). We discuss the generalization of coordinates and directions in hyperbolic space and then review geodesic projections. We finally describe generalizations of the notion of mean and variance to non-Euclidean spaces.

2.1. The Poincaré Model of Hyperbolic Space

Hyperbolic geometry is a Riemannian geometry with constant negative curvature -1, where curvature measures deviation from flat Euclidean geometry. For easier visualizations, we work with the d-dimensional Poincaré model of

	Euclidean	Hyperbolic
Component	Unit vector w	Ideal point p
Coordinate	Dot product $x \cdot w$	Busemann func. $B_p(x)$

Table 1: Analogies of components and their corresponding coordinates, in both Euclidean and hyperbolic space.

hyperbolic space: $\mathbb{H}^d = \{x \in \mathbb{R}^d : \|x\| < 1\}$, where $\|\cdot\|$ is the Euclidean norm. In this model, the Riemannian distance can be computed in cartesian coordinates by:

$$d_{\mathbb{H}}(x, y) = \cosh^{-1}\left(1 + 2\frac{\|x-y\|^2}{(1-\|x\|^2)(1-\|y\|^2)}\right). \quad (1)$$

Geodesics Shortest paths in hyperbolic space are called *geodesics*. In the Poincaré model, they are represented by straight segments going through the origin and circular arcs perpendicular to the boundary of the unit ball (Fig. 2).

Geodesic submanifolds A submanifold $M \subset \mathbb{H}^d$ is called *(totally) geodesic* if for every $x, y \in M$, the geodesic line connecting x and y belongs to M. This generalizes the notion of affine subspaces in Euclidean spaces. In the Poincaré model, geodesic submanifolds are represented by linear subspaces going through the origin and spherical caps perpendicular to the boundary of the unit ball.

2.2. Directions in Hyperbolic space

The notions of directions, and coordinates in a given direction can be generalized to hyperbolic spaces as follows.

Ideal points As with parallel rays in Euclidean spaces, geodesic rays in $\mathbb{H}^d$ that stay close to each other can be viewed as sharing a common *endpoint at infinity*, also called an *ideal point*. Intuitively, ideal points represent *directions* along which points in $\mathbb{H}^d$ can move toward infinity. The set of ideal points $\mathbb{S}^{d-1}_\infty$, called the *boundary at infinity* of $\mathbb{H}^d$, is represented by the unit sphere $\mathbb{S}^{d-1}_\infty = \{\|x\| = 1\}$ in the Poincaré model. We abuse notations and say that a geodesic submanifold $M \subset \mathbb{H}^d$ contains an ideal point p if the boundary of M in $\mathbb{S}^{d-1}_\infty$ contains p.

Busemann coordinates In Euclidean spaces, each *direction* can be represented by a unit vector w. The *coordinate* of a point x in the direction of w is simply the dot product $w \cdot x$. In hyperbolic geometry, directions can be represented by ideal points but dot products are not well-defined. Instead, we take a ray-based perspective: note that in Euclidean spaces, if we shoot a ray in the direction of w from the origin, the coordinate $w \cdot x$ can be viewed as the *normalized distance to infinity in the direction of that ray*. In other words, as a point $y = tw$, $(t > 0)$ moves toward infinity in the direction of w:

$$w \cdot x = \lim_{t\to\infty} \left(d(0, tw) - d(x, tw)\right).$$

This approach generalizes to other geometries: given a unit-speed geodesic ray $\gamma(t)$, the *Busemann function* $B_\gamma(x)$ of γ is defined as:[1]

$$B_\gamma(x) = \lim_{t\to\infty} \left(d(x, \gamma(t)) - t\right).$$

Up to an additive constant, this function only depends on the endpoint at infinity of the geodesic ray, and not the starting point $\gamma(0)$. Thus, given an ideal point p, we define the Busemann function $B_p(x)$ of p to be the Busemann function of the geodesic ray that starts from the origin of the unit ball model and has endpoint p. Intuitively, it represents the *coordinates* of x in the direction of p. In the Poincaré model, there is a closed formula:

$$B_p(x) = \ln \frac{\|p-x\|^2}{1-\|x\|^2}.$$

Horospheres The level sets of Busemann functions $B_p(x)$ are called *horospheres centered at* p. In this sense, they resemble spheres, which are level sets of distance functions. However, intrinsically as Riemannian manifolds, horospheres have curvature zero and thus also exhibit many properties of planes in Euclidean spaces.

Every geodesic with endpoint p is orthogonal to every horosphere centered at p. Given two horospheres with the same center, every orthogonal geodesic segment connecting them has the same length. In this sense, *concentric horospheres resemble parallel planes in Euclidean spaces*. In the Poincaré model, horospheres are Euclidean spheres that touch the boundary sphere $\mathbb{S}^{d-1}_\infty$ at their ideal centers (Fig. 2). Given an ideal point p and a point x in $\mathbb{H}^d$, there is a unique horosphere $S(p, x)$ passing through x and centered at p.

2.3. Geodesic Projections

PCA uses orthogonal projections to project data onto subspaces. Orthogonal projections are usually generalized to other geometries as *closest-point projections*. Given a target submanifold M, each point x in the ambient space is mapped to the closest-point to it in M:

$$\pi^{\mathrm{G}}_M(x) = \operatorname*{argmin}_{y\in M} d_M(x, y).$$

One could view $\pi^{\mathrm{G}}_M(\cdot)$ as the map that pushes each point x along an orthogonal geodesic until it hits M. For this reason, it is also called *geodesic projection*. In the Poincaré model, these can be computed in closed-form (see Appendix C).

2.4. Manifold Statistics

PCA relies on data statistics which do not generalize easily to hyperbolic geometry. One approach to generalize the

[1] Note that compared to the above formula, the sign convention is flipped due to historical reasons.

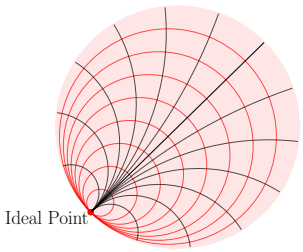


Figure 2: Hyperbolic geodesics (black lines) going through an ideal point (in red), and horospheres (red circles) centered at that same ideal point. The hyperbolic lengths of geodesic segments between two horospheres are equal.

arithmetic mean is to notice that it is the minimizer of the sum of squared distances to the inputs. Motivated by this, the Fréchet mean (Fréchet, 1948) of a set of points S in a Riemannian manifold (M, d_M) is defined as:

$$\mu_M(S) \coloneqq \operatorname*{argmin}_{y \in M} \sum_{x \in S} d_M(x, y)^2.$$

This definition only depends on the intrinsic distance of the manifold. For hyperbolic spaces, since squared distance functions are convex, $\mu(S)$ always exists and is unique.[2] Analogously, the Fréchet variance is defined as:

$$\sigma_M^2(S) \coloneqq \frac{1}{|S|} \sum_{x \in S}^{N} d_M(x, \mu(S))^2. \tag{2}$$

We refer to (Huckemann & Eltzner, 2020) for a discussion on different intrinsic statistics in non-Euclidean spaces, and a study of their asymptotic properties.

3. Generalizing PCA to the Hyperbolic Space

We now describe our approach to generalize PCA to hyperbolic spaces. The starting point of HOROPCA is to pick $1 \leq K \leq d$ ideal points $p_1, \ldots, p_K \in \mathbb{S}_\infty^{d-1}$ to represent K directions in hyperbolic spaces (Section 2.2). Then, we generalize the core concepts of Euclidean PCA. In Section 3.1, we show how to generalize *flags*. In Section 3.2, we show how to *project* points onto the lower-dimensional submanifold spanned by a given set of directions, while preserving information along each direction. In Section 3.3, we introducing a *variance objective* to optimize and show that it is a function of the directions only.

3.1. Hyperbolic Flags

In Euclidean spaces, one can take the linear spans of more and more components to define a nested sequence of linear subspaces, called a flag. To generalize this to hyperbolic spaces, we first need to adapt the notion of linear/affine spans. Recall that geodesic submanifolds are generalizations of affine subspaces in Euclidean spaces.

Definition 3.1. Given a set of points S (that could be inside $\mathbb{H}^d$ or on the boundary sphere $\mathbb{S}_\infty^{d-1}$), the smallest geodesic submanifold of $\mathbb{H}^d$ that contains S is called the *geodesic hull* of S and denoted by $\mathrm{GH}(S)$.

Thus, given K ideal points $p_1, p_2, \ldots, p_K$ and a base point $b \in \mathbb{H}^d$, we can define a nested sequence of geodesic submanifolds $\mathrm{GH}(b, p_1) \subset \mathrm{GH}(b, p_1, p_2) \subset \cdots \subset \mathrm{GH}(b, p_1, \ldots, p_K)$. This will be our notion of flags.

Remark 3.2. The base point b is only needed here for technical reasons, just like an origin $\mathbf{o}$ is needed to define linear spans in Euclidean spaces. We will see next that it does not affect the projection results or objectives (Theorem 4.1).

Remark 3.3. We assume that none of $b, p_1, \ldots, p_K$ are in the geodesic hull of the other K points. This is analogous to being linearly independent in Euclidean spaces.

3.2. Projections via Horospheres

In Euclidean PCA, points are projected to the subspaces spanned by the given directions in a way that preserves coordinates in those directions. We seek a projection method in hyperbolic spaces with a similar property.

Recall that coordinates are generalized by Busemann functions (Table 1), and that horospheres are level sets of Busemann functions. Thus, we propose a projection that preserves coordinates by moving points along horospheres. It turns out that this projection method also preserves distances better than the traditional geodesic projection.

As a toy example, we first show how the projection is defined in the $K = 1$ case (i.e. projecting onto a geodesic) and why it tends to preserve distances well. We will then show how to use $K \geq 1$ ideal points simultaneously.

3.2.1. Projecting onto $K = 1$ Directions

For $K = 1$, we have one ideal point p and base point b, and the geodesic hull $\mathrm{GH}(b, p)$ is just a geodesic γ. Our goal is to map every $x \in \mathbb{H}^d$ to a point $\pi_{b,p}^{\mathrm{H}}(x)$ on γ that has the same Busemann coordinate in the direction of p:

$$B_p(x) = B_p(\pi_{b,p}^{\mathrm{H}}(x)).$$

Since level sets of $B_p(x)$ are horospheres centered at p, the above equation simply says that $\pi_{b,p}^{\mathrm{H}}(x)$ belongs to the horosphere $S(p, x)$ centered at p and passing through x. Thus, we define:

$$\pi_{b,p}^{\mathrm{H}}(x) \coloneqq \gamma \cap S(p, x). \tag{3}$$

Another important property that $\pi_{b,p}^{\mathrm{H}}(\cdot)$ shares with orthogonal projections in Euclidean spaces is that it preserves

[2] For more general geometries, existence and uniqueness hold if the data is well-localized (Kendall, 1990).

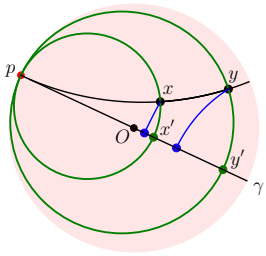


Figure 3: x', y' are horospherical (green) projections of x, y. Proposition 3.4 shows $d_{\mathbb{H}}(x', y') = d_{\mathbb{H}}(x, y)$. The distance between the two geodesic (blue) projections is smaller.

distances along a direction – lengths of geodesic segments that point to p are preserved after projection (Fig. 3):

Proposition 3.4. *For any $x \in \mathbb{H}^d$, if $y \in \mathrm{GH}(x, p)$ then:*

$$d_{\mathbb{H}}(\pi^{\mathrm{H}}_{b,p}(x), \pi^{\mathrm{H}}_{b,p}(y)) = d_{\mathbb{H}}(x, y).$$

Proof. This follows from the remark in Section 2.2 about horospheres: every geodesic going through p is orthogonal to every horosphere centered at p, and every orthogonal geodesic segment connecting concentric horospheres has the same length (Fig. 2). In this case, the segments from x to y and from $\pi^{\mathrm{H}}_{b,p}(x)$ to $\pi^{\mathrm{H}}_{b,p}(y)$ are two such segments, connecting $S(p, x)$ and $S(p, y)$. □

3.2.2. Projecting onto $K > 1$ Directions

We now generalize the above construction to projections onto higher-dimensional submanifolds. We describe the main ideas here; Appendix A contains more details, including an illustration in the case $K = 2$ (Fig. 5).

Fix a base point $b \in \mathbb{H}^d$ and $K > 1$ ideal points $\{p_1, \ldots, p_K\}$. We want to define a map from $\mathbb{H}^d$ to $M := \mathrm{GH}(b, p_1, \ldots, p_K)$ that preserves the Busemann coordinates in the directions of $p_1, \ldots, p_K$, i.e.:

$$B_{p_j}(x) = B_{p_j}\left(\pi^{\mathrm{H}}_{b,p_1,\ldots,p_K}(x)\right) \text{ for every } j = 1, \ldots, K.$$

As before, the idea is to take the intersection with the horospheres centered at p_j's and passing through x:

$$\pi^{\mathrm{H}}_{b,p_1,\ldots,p_K} : \mathbb{H}^d \to M$$
$$x \mapsto M \cap S(p_1, x) \cap \cdots \cap S(p_K, x).$$

It turns out that this intersection generally consists of two points instead of one. When that happens, one of them will be strictly closer to the base point b, and we define $\pi^{\mathrm{H}}_{b,p_1,\ldots,p_K}(x)$ to be that point.

As with Proposition 3.4, $\pi^{\mathrm{H}}_{b,p_1,\ldots,p_K}(\cdot)$ preserves distances along K-dimensional manifolds (Corollary A.10). In contrast, geodesic projections in hyperbolic spaces *never* preserve distances (except between points already in the target):

Proposition 3.5. *Let $M \subset \mathbb{H}^d$ be a geodesic submanifold. Then every geodesic segment of distance at least r from M gets at least $\cosh(r)$ times shorter under the geodesic projection $\pi^{\mathrm{G}}_M(\cdot)$ to M:*

$$\mathrm{length}(\pi^{\mathrm{G}}_M(I)) \leq \frac{1}{\cosh(r)} \mathrm{length}(I).$$

In particular, the shrink factor grows exponentially as the segment I moves away from M. □

The proof is in Appendix B.

Computation Interestingly, horosphere projections can be computed without actually computing the horospheres. The key idea is that if we let $P = \mathrm{GH}(p_1, \ldots, p_K)$ be the geodesic hull of the horospheres' centers, then the intersection $S(p_1, x) \cap \cdots \cap S(p_K, x)$ is simply the orbit of x under the rotations around P. (This is true for the same reason that spheres whose centers lie on the same axis must intersect along a circle around that axis). Thus, $\pi^{\mathrm{H}}_{b,p_1,\ldots,p_K}(\cdot)$ can be viewed as the map that rotates x around until it hits M. It follows that it can be computed by:

1. Find the geodesic projection $c = \pi^{\mathrm{G}}_P(x)$ of x onto P.
2. Find the geodesic α on M that is orthogonal to P at c.
3. Among the two points on α whose distance to c equals $d_{\mathbb{H}}(x, c)$, returns the one closer to b.

The detailed computations and proof that this recovers horospherical projections are provided in Appendix A.

3.3. Intrinsic Variance Objective

In Euclidean PCA, directions are chosen to maximally preserve information from the original data. In particular, PCA chooses directions that maximize the Euclidean variance of projected data. To generalize this to hyperbolic geometry, we define an analog of variance that is intrinsic, i.e. dependent only on the distances between data points. As we will see in Section 4, having an intrinsic objective helps make our algorithm *location (or base point) independent*.

The usual notion of Euclidean variance is the squared sum of distances to the mean of the projected datapoints. Generalizing this is challenging because non-Euclidean spaces do not have a canonical choice of mean. Previous works have generalized variance either by using the *unexplained variance* or *Fréchet variance*. The former is the squared sum of residual distances to the projections, and thus avoids computing a mean. However, it is not intrinsic. The latter is intrinsic (Fletcher et al., 2004) but involves finding the Fréchet mean, which is not necessarily a canonical notion of mean and can only be computed by gradient descent.

Our approach uses the observation that in Euclidean space:

$$\sigma^2(S) = \frac{1}{n} \sum_{x \in S} \|x - \mu(S)\|^2 = \frac{1}{n^2} \sum_{x,y \in S} \|x - y\|^2.$$

Thus, we propose the following generalization of variance:

$$\sigma^2_{\mathbb{H}}(S) = \frac{1}{n^2} \sum_{x,y \in S} d_{\mathbb{H}}(x,y)^2. \tag{4}$$

This function agrees with the usual variance in Euclidean space, while being a function of distances only. Thus it is well defined in non-Euclidean space, is easily computed, and, as we will show next, has the desired invariance due to isometry properties of horospherical projections.

4. HoroPCA

Section 3 formulated several simple primitives – including directions, flags, projections, and variance – in a way that is generalizable to hyperbolic geometry. We now revisit standard PCA, showing how it has a simple definition that combines these primitives using optimization. This directly leads to the full HoroPCA algorithm by simply using the hyperbolic analogs of these primitives.

Euclidean PCA Given a dataset S and a target dimension K, Euclidean PCA greedily finds a sequence of *principal components* $p_1, \ldots, p_K$ that maximizes the *variance* of orthogonal *projections* $\pi^{\mathrm{E}}_{\mathbf{o},p_1,\ldots,p_k}(\cdot)$ onto the linear [3] *subspaces* spanned by these components:

$$p_1 = \operatorname*{argmax}_{||p||=1} \sigma^2(\pi^{\mathrm{E}}_{\mathbf{o},p}(S))$$
$$\text{and } p_{k+1} = \operatorname*{argmax}_{||p||=1} \sigma^2(\pi^{\mathrm{E}}_{\mathbf{o},p_1,\ldots,p_k,p}(S)).$$

Thus, for each $1 \le k \le K$, $\{p_1, \ldots, p_k\}$ is the optimal set of directions of dimension k.

HoroPCA Because we have generalized the notions of flag, projection, and variance to hyperbolic geometry, the HoroPCA algorithm can be defined in the same fashion. Given a dataset S in $\mathbb{H}^d$ and a base point $b \in \mathbb{H}^d$, we seek a sequence of K directions that maximizes the variance of horosphere-projected data:

$$p_1 = \operatorname*{argmax}_{p \in \mathbb{S}^{d-1}_{\infty}} \sigma^2_{\mathbb{H}}(\pi^{\mathrm{H}}_{b,p}(S))$$
$$\text{and } p_{k+1} = \operatorname*{argmax}_{p \in \mathbb{S}^{d-1}_{\infty}} \sigma^2_{\mathbb{H}}(\pi^{\mathrm{H}}_{b,p_1,\ldots,p_k,p}(S)). \tag{5}$$

Base point independence Finally, we show that algorithm (5) always returns the same results regardless of the choice of a base point $b \in \mathbb{H}^d$. Since our variance objective only depends on the distances between projected data points, it suffices to show that these distances do not depend on b.

[3] Here $\mathbf{o}$ denotes the origin, and $\pi^{\mathrm{E}}_{\mathbf{o},p_1,\ldots,p_k}(\cdot)$ denotes the projection onto the *affine* span of $\{\mathbf{o}, p_1, \ldots, p_k\}$, which is equivalent to the *linear* span of $\{p_1, \ldots, p_k\}$.

Theorem 4.1. *For any b, b' and any $x, y \in \mathbb{H}^d$, the two projected distances $d_{\mathbb{H}}(\pi^{\mathrm{H}}_{b,p_1,\ldots,p_K}(x), \pi^{\mathrm{H}}_{b,p_1,\ldots,p_K}(y))$ and $d_{\mathbb{H}}(\pi^{\mathrm{H}}_{b',p_1,\ldots,p_K}(x), \pi^{\mathrm{H}}_{b',p_1,\ldots,p_K}(y))$ are equal.*

The proof is included in Appendix A. Thus, HoroPCA retains the *location-independence* property of Euclidean PCA: only the *directions* of target subspaces matter; their exact locations do not (Fig. 1). Therefore, just like in the Euclidean setting, we can assume without loss of generality that b is the origin $\mathbf{o}$ of the Poincaré model. This:

1. alleviates the need to use d extra parameters to search for an appropriate base point, and
2. simplifies computations, since in the Poincaré model, geodesics submanifolds that go through the origin are simply linear subspaces, which are easier to deal with.

After computing the principal directions which span the target $M = \mathrm{GH}(\mathbf{o}, p_1, \ldots, p_K)$, the reduced dimensionality data can be found by applying an Euclidean rotation that sends M to $\mathbb{H}^K$, which also preserves hyperbolic distances.

5. Experiments

We now validate the empirical benefits of HoroPCA on three PCA uses. First, for dimensionality reduction, HoroPCA preserves information (distances and variance) better than previous methods which are sensitive to base point choices and distort distances more (Section 5.2). Next, we validate that our notion of hyperbolic coordinates captures variation in the data and can be used for whitening in classification tasks (Section 5.3). Finally, we visualize the representations learned by HoroPCA in two dimensions (Section 5.4).

5.1. Experimental Setup

Baselines We compare HoroPCA to several dimensionality reduction methods, including: (1) Euclidean PCA, which should perform poorly on hyperbolic data, (2) Exact PGA, (3) Tangent PCA (tPCA), which approximates PGA by moving the data in the tangent space of the Fréchet mean and then solves Euclidean PCA, (4) BSA, (5) Hyperbolic Multi-dimensional Scaling (hMDS) (Sala et al., 2018), which takes a distance matrix as input and recovers a configuration of points that best approximates these distances, (6) Hyperbolic autoencoder (hAE) trained with gradient descent (Ganea et al., 2018; Hinton & Salakhutdinov, 2006). To demonstrate their dependence on base points, we also include two baselines that perturb the base point in PGA and BSA. We open-source our implementation[4] and refer to Appendix C for implementation details on how we implemented all baselines and HoroPCA.

[4] https://github.com/HazyResearch/HoroPCA

	Balanced Tree		Phylo Tree		Diseases		CS Ph.D.	
	distortion (↓)	variance (↑)	distortion (↓)	variance (↑)	distortion (↓)	variance (↑)	distortion (↓)	variance (↑)
PCA	0.84	0.34	0.94	0.40	0.90	0.26	0.84	1.68
tPCA	0.70	1.16	0.63	14.34	0.63	3.92	0.56	11.09
PGA	0.63 ± 0.07	2.11 ± 0.47	0.64 ± 0.01	15.29 ± 0.51	0.66 ± 0.02	3.16 ± 0.39	0.73 ± 0.02	6.14 ± 0.60
PGA-Noise	0.87 ± 0.08	0.29 ± 0.30	0.64 ± 0.02	15.08 ± 0.77	0.88 ± 0.04	0.53 ± 0.19	0.79 ± 0.03	4.58 ± 0.64
BSA	0.50 ± 0.00	3.02 ± 0.01	0.61 ± 0.03	18.60 ± 1.16	0.52 ± 0.02	5.95 ± 0.25	0.70 ± 0.01	8.15 ± 0.96
BSA-Noise	0.74 ± 0.12	1.06 ± 0.67	0.68 ± 0.02	13.71 ± 0.72	0.80 ± 0.11	1.62 ± 1.30	0.79 ± 0.02	4.41 ± 0.59
hAE	0.26 ± 0.00	6.91 ± 0.00	$\underline{0.32} \pm 0.04$	$\underline{45.87} \pm 3.52$	$\underline{0.18} \pm 0.00$	14.23 ± 0.06	$\underline{0.37} \pm 0.02$	$\underline{22.12} \pm 2.47$
hMDS	$\underline{0.22}$	**7.54**	0.74	40.51	0.21	$\underline{15.05}$	0.83	19.93
HoroPCA	**0.19** ± 0.00	$\underline{7.15} \pm 0.00$	**0.13** ± 0.01	**69.16** ± 1.96	**0.15** ± 0.01	**15.46** ± 0.19	**0.16** ± 0.02	**36.79** ± 0.70

Table 2: Dimensionality reduction results on 10-dimensional hyperbolic embeddings reduced to two dimensions. Results are averaged over 5 runs for non-deterministic methods. Best in **bold** and second best underlined.

Datasets For dimensionality reduction experiments, we consider standard hierarchical datasets previously used to evaluate the benefits of hyperbolic embeddings. More specifically, we use the datasets in (Sala et al., 2018) including a fully balanced tree, a phylogenetic tree, a biological graph comprising of diseases' relationships and a graph of Computer Science (CS) Ph.D. advisor-advisee relationships. These datasets have respectively 40, 344, 516 and 1025 nodes, and we use the code from (Gu et al., 2018) to embed them in the Poincaré ball. For data whitening experiments, we reproduce the experimental setup from (Cho et al., 2019) and use the Polbooks, Football and Polblogs datasets which have 105, 115 and 1224 nodes each. These real-world networks are embedded in two-dimensions using Chamberlain et al. (2017)'s embedding method.

Evaluation metrics To measure distance-preservation after projection, we use average distortion. If $\pi(\cdot)$ denotes a mapping from high- to low-dimensional representations, the average distortion of a dataset S is computed as:

$$\frac{1}{\binom{|S|}{2}} \sum_{x \neq y \in S} \frac{|d_{\mathbb{H}}(\pi(x), \pi(y)) - d_{\mathbb{H}}(x, y)|}{d_{\mathbb{H}}(x, y)}.$$

We also measure the Fréchet variance in Eq. (2), which is the analogue of the objective that Euclidean PCA optimizes[5]. Note that the mean in Eq. (2) cannot be computed in closed-form and we therefore compute it with gradient-descent.

5.2. Dimensionality Reduction

We report metrics for the reduction of 10-dimensional embeddings to two dimensions in Table 2, and refer to Appendix C for additional results, such as more component and dimension configurations. All results suggest that HoroPCA better preserves information contained in the high-dimensional representations.

On distance preservation, HoroPCA outperforms all methods with significant improvements on larger datasets. This supports our theoretical result that horospherical projections better preserve distances than geodesic projections. Furthermore, HoroPCA also outperforms existing methods on the explained Fréchet variance metric on all but one dataset. This suggests that our distance-based formulation of the variance (Eq. (4)) effectively captures variations in the data. We also note that as expected, both PGA and BSA are sensitive to base point choices: adding Gaussian noise to the base point leads to significant drops in performance. In contrast, HoroPCA is by construction base-point independent.

5.3. Hyperbolic Data Whitening

An important use of PCA is for data whitening, as it allows practitioners to remove noise and decorrelate the data, which can improve downstream tasks such as regression or classification. Recall that standard PCA data whitening consists of (i) finding principal directions that explain the data, (ii) calculating the coordinates of each data point along these directions, and (iii) normalizing the coordinates for each direction (to have zero mean and unit variance).

Because of the close analogy between HoroPCA and Euclidean PCA, these steps can easily map to the hyperbolic case, where we (i) use HoroPCA to find principal directions (ideal points), (ii) calculate the Busemann coordinates along these directions, and (iii) normalize them as Euclidean coordinates. Note that this yields Euclidean representations, which allow leveraging powerful tools developed specifically for learning on Euclidean data.

We evaluate the benefit of this whitening step on a simple classification task. We compare to directly classifying the data with Euclidean Support Vector Machine (eSVM) or its hyperbolic counterpart (hSVM), and also to whitening with tPCA. Note that most baselines in Section 5.1 are incompatible with data whitening: hMDS does not learn a transformation that can be applied to unseen test data, while methods like PGA and BSA do not naturally return Euclidean coordinates for us to normalize. To obtain another

[5] All mentioned PCA methods, including HoroPCA, optimize for some forms of variance but *not* Fréchet variance or distortion.

	Polbooks	Football	Polblogs
eSVM	69.9± 1.2	20.7 ± 3.0	92.3 ± 1.5
hSVM	68.3 ± 0.6	20.9 ± 2.5	92.2 ± 1.6
tPCA+eSVM	68.5 ± 0.9	21.2 ± 2.2	92.4 ± 1.5
PGA+eSVM	64.4 ± 4.1	21.7 ± 2.2	82.3 ± 1.2
HoroPCA+eSVM	**72.2** ± 2.8	**25.0** ± 1.0	**92.8** ± 0.9

Table 3: Data whitening experiments. We report classification accuracy averaged over 5 embedding configurations. Best in **bold** and second best underlined.

baseline, we use a logarithmic map to extract Euclidean coordinates from PGA.

We reproduce the experimental setup from (Cho et al., 2019) who split the datasets in 50% train and 50% test sets, run classification on 2-dimensional embeddings and average results over 5 different embedding configurations as was done in the original paper (Table 3). [6] HoroPCA whitening improves downstream classification on all datasets compared to eSVM and hSVM or tPCA and PGA whitening. This suggests that HoroPCA can be leveraged for hyperbolic data whitening. Further, this confirms that Busemann coordinates do capture variations in the original data.

5.4. Visualizations

When learning embeddings for ML applications (e.g. classification), increasing the dimensionality can significantly improve the embeddings' quality. To effectively work with these higher-dimensional embeddings, it is useful to visualize their structure and organization, which often requires reducing their representations to two or three dimensions. Here, we consider embeddings of the mammals subtree of the Wordnet noun hierarchy learned with the algorithm from (Nickel & Kiela, 2017). We reduce embeddings to two dimensions using PGA and HoroPCA and show the results in Fig. 4. We also include more visualizations for PCA and BSA in Fig. 8 in the Appendix. As we can see, the reduced representations obtained with HoroPCA yield better visualizations. For instance, we can see some hierarchical patterns such as "feline hypernym of cat" or "cat hypernym of burmese cat". These patterns are harder to visualize for other methods, since these do not preserve distances as well as HoroPCA, e.g. PGA has 0.534 average distortion on this dataset compared to 0.078 for HoroPCA.

6. Related Work

PCA methods in Riemannian manifolds We first review some approaches to extending PCA to general Riemannian geometries, of which hyperbolic geometry is a special case. For a more detailed discussion, see Pennec (2018). The simplest such approach is tangent PCA (tPCA), which maps the data to the tangent space at the Fréchet mean μ using the logarithm map, then applies Euclidean PCA. A similar approach, Principal Geodesic Analysis (PGA) (Fletcher et al., 2004), seeks geodesic subspaces at μ that minimize the sum of squared Riemannian distances to the data. Compared to tPCA, PGA searches through the same subspaces but uses a more natural loss function.

Both PGA and tPCA project on submanifolds that go through the Fréchet mean. When the data is not well-centered, this may be sub-optimal, and Geodesic PCA (GPCA) was proposed to alleviate this issue (Huckemann & Ziezold, 2006; Huckemann et al., 2010). GPCA first finds a geodesic γ that best fits the data, then finds other orthogonal geodesics that go through *some* common point b on γ. In other words, GPCA removes the constraint of PGA that b is the Fréchet mean. Extensions of GPCA have been proposed such as probabilistic methods (Zhang & Fletcher, 2013) and Horizontal Component Analysis (Sommer, 2013).

Pennec (2018) proposes a more symmetric approach. Instead of using the exponential map at a base point, it parameterizes K-dimensional subspaces as the *barycenter* loci of $K+1$ points. Nested sequences of subspaces (flags) can be formed by simply adding more points. In hyperbolic geometry, this construction coincides with the one based on geodesic hulls that we use in Section 3, except that it applies to points inside $\mathbb{H}^d$ instead of ideal points and thus needs more parameters to parameterize a flag (see Remark A.8).

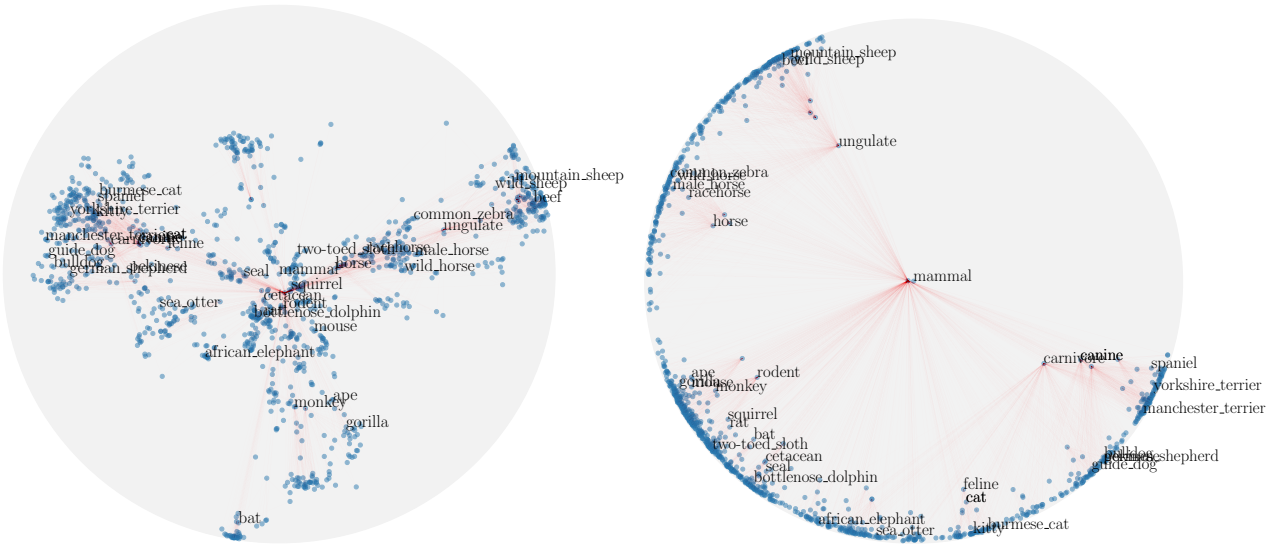
By considering a more general type of submanifolds, Hauberg (2016) gives another way to avoid the sensitive dependence on Fréchet mean. However, its bigger search space also makes the method computationally expensive, especially when the target dimension K is bigger than 1.

In contrast with all methods so far, HoroPCA relies on horospherical projections instead of geodesic projections. This yields a generalization of PCA that depends only on the directions and not specific locations of the subspaces.

Dimension reduction in hyperbolic geometry We now review some dimension reduction methods proposed specifically for hyperbolic geometry. Cvetkovski & Crovella (2011) and Sala et al. (2018) are examples of hyperbolic multidimensional scaling methods, which seek configurations of points in lower-dimensional hyperbolic spaces whose pairwise distances best approximate a given dissimilarity matrix. Unlike HoroPCA, they do not learn a projection that can be applied to unseen data.

Tran & Vu (2008) constructs a map to lower-dimensional hyperbolic spaces whose preimages of compact sets are compact. Unlike most methods, it is data-agnostic and does not optimize any objective.

[6] Note that the results slightly differ from (Cho et al., 2019) which could be because of different implementations or data splits.



(a) PGA (average distortion: 0.534)

(b) HoroPCA (average distortion: 0.078)

Figure 4: Visualization of embeddings of the WordNet mammal subtree computed by reducing 10-dimensional Poincaré embeddings (Nickel & Kiela, 2017).

Benjamini & Makarychev (2009) adapt the Euclidean Johnson–Lindenstrauss transform to hyperbolic geometry and obtain a distortion bound when the dataset size is not too big compared to the target dimension. They do not seek an analog of Euclidean directions or projections, but nevertheless implicitly use a projection method based on pushing points along *horocycles*, which shares many properties with our horo*spherical* projections. In fact, the latter converges to the former as the ideal points get closer to each other.

From hyperbolic to Euclidean Liu et al. (2019) use "distances to centroids" to compute Euclidean representations of hyperbolic data. The Busemann functions we use bear resemblances to these centroid-based functions but are better analogs of coordinates along given directions, which is a central concept in PCA, and have better regularity properties (Busemann, 1955). Recent works have also used Busemann functions for hyperbolic prototype learning (Keller-Ressel, 2020; Wang, 2021). These works do not define projections to lower-dimensional hyperbolic spaces. In contrast, HoroPCA naturally returns both hyperbolic representations (via horospherical projections) and Euclidean representations (via Busemann coordinates). This allows leveraging techniques in both settings.

7. Conclusion

We proposed HoroPCA, a method to generalize PCA to hyperbolic spaces. In contrast with previous PCA generalizations, HoroPCA preserves the core location-independence PCA property. Empirically, HoroPCA significantly outperforms previous methods on the reduction of hyperbolic data. Future extensions of this work include deriving a closed-form solution, analyzing the stability properties of HoroPCA, or using the concepts introduced in this work to derive efficient nearest neighbor search algorithms or neural network operations.

Acknowledgements

We gratefully acknowledge the support of NIH under No. U54EB020405 (Mobilize), NSF under Nos. CCF1763315 (Beyond Sparsity), CCF1563078 (Volume to Velocity), and 1937301 (RTML); ONR under No. N000141712266 (Unifying Weak Supervision); the Moore Foundation, NXP, Xilinx, LETI-CEA, Intel, IBM, Microsoft, NEC, Toshiba, TSMC, ARM, Hitachi, BASF, Accenture, Ericsson, Qualcomm, Analog Devices, the Okawa Foundation, American Family Insurance, Google Cloud, Swiss Re, Total, the HAI-AWS Cloud Credits for Research program, the Stanford Data Science Initiative (SDSI), and members of the Stanford DAWN project: Facebook, Google, and VMWare. The Mobilize Center is a Biomedical Technology Resource Center, funded by the NIH National Institute of Biomedical Imaging and Bioengineering through Grant P41EB027060. The U.S. Government is authorized to reproduce and distribute reprints for Governmental purposes notwithstanding any copyright notation thereon. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views, policies, or endorsements, either expressed or implied, of NIH, ONR, or the U.S. Government.

References

Balazevic, I., Allen, C., and Hospedales, T. Multi-relational poincaré graph embeddings. In *Advances in Neural Information Processing Systems*, pp. 4463–4473, 2019.

Benjamini, I. and Makarychev, Y. Dimension reduction for hyperbolic space. *Proceedings of the American Mathematical Society*, pp. 695–698, 2009.

Busemann, H. *The geometry of geodesics*. Academic Press Inc., New York, N. Y., 1955.

Chamberlain, B., Clough, J., and Deisenroth, M. Neural embeddings of graphs in hyperbolic space. In *CoRR*. MLG Workshop 2017, 2017.

Chami, I., Ying, Z., Ré, C., and Leskovec, J. Hyperbolic graph convolutional neural networks. In *Advances in neural information processing systems*, pp. 4868–4879, 2019.

Chami, I., Gu, A., Chatziafratis, V., and Ré, C. From trees to continuous embeddings and back: Hyperbolic hierarchical clustering. *Advances in Neural Information Processing Systems*, 33, 2020a.

Chami, I., Wolf, A., Juan, D.-C., Sala, F., Ravi, S., and Ré, C. Low-dimensional hyperbolic knowledge graph embeddings. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pp. 6901–6914, 2020b.

Cho, H., DeMeo, B., Peng, J., and Berger, B. Large-margin classification in hyperbolic space. In *The 22nd International Conference on Artificial Intelligence and Statistics*, pp. 1832–1840. PMLR, 2019.

Cvetkovski, A. and Crovella, M. Multidimensional scaling in the poincaré disk. *arXiv preprint arXiv:1105.5332*, 2011.

Fletcher, P. T., Conglin Lu, Pizer, S. M., and Sarang Joshi. Principal geodesic analysis for the study of nonlinear statistics of shape. *IEEE Transactions on Medical Imaging*, 23(8):995–1005, 2004.

Fréchet, M. Les éléments aléatoires de nature quelconque dans un espace distancié. In *Annales de l'institut Henri Poincaré*, volume 10, pp. 215–310, 1948.

Ganea, O.-E., Bécigneul, G., and Hofmann, T. Hyperbolic neural networks. In *Proceedings of the 32nd International Conference on Neural Information Processing Systems*, pp. 5350–5360, 2018.

Gu, A., Sala, F., Gunel, B., and Ré, C. Learning mixed-curvature representations in product spaces. In *International Conference on Learning Representations*, 2018.

Hauberg, S. Principal curves on riemannian manifolds. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 38(9):1915–1921, 2016. doi: 10.1109/TPAMI.2015.2496166.

Hinton, G. E. and Salakhutdinov, R. R. Reducing the dimensionality of data with neural networks. *science*, 313 (5786):504–507, 2006.

Huckemann, S. and Eltzner, B. Statistical methods generalizing principal component analysis to non-euclidean spaces. In *Handbook of Variational Methods for Nonlinear Geometric Data*, pp. 317–338. Springer, 2020.

Huckemann, S. and Ziezold, H. Principal component analysis for riemannian manifolds, with an application to triangular shape spaces. *Advances in Applied Probability*, 38(2):299–319, 2006.

Huckemann, S., Hotz, T., and Munk, A. Intrinsic shape analysis: Geodesic pca for riemannian manifolds modulo isometric lie group actions. *Statistica Sinica*, pp. 1–58, 2010.

Keller-Ressel, M. A theory of hyperbolic prototype learning. *arXiv preprint arXiv:2010.07744*, 2020.

Kendall, W. S. Probability, convexity, and harmonic maps with small image i: uniqueness and fine existence. *Proceedings of the London Mathematical Society*, 3(2):371–406, 1990.

Krauthgamer, R. and Lee, J. R. Algorithms on negatively curved spaces. In *2006 47th Annual IEEE Symposium on Foundations of Computer Science (FOCS'06)*, pp. 119–132. IEEE, 2006.

Lee, J. M. Smooth manifolds. In *Introduction to Smooth Manifolds*, pp. 1–31. Springer, 2013.

Liu, Q., Nickel, M., and Kiela, D. Hyperbolic graph neural networks. In *Advances in Neural Information Processing Systems*, pp. 8230–8241, 2019.

Monath, N., Zaheer, M., Silva, D., McCallum, A., and Ahmed, A. Gradient-based hierarchical clustering using continuous representations of trees in hyperbolic space. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pp. 714–722, 2019.

Nickel, M. and Kiela, D. Poincaré embeddings for learning hierarchical representations. In *Advances in neural information processing systems*, pp. 6338–6347, 2017.

Pennec, X. Barycentric subspace analysis on manifolds. *Annals of Statistics*, 46(6A):2711–2746, 2018.

Sala, F., De Sa, C., Gu, A., and Ré, C. Representation tradeoffs for hyperbolic embeddings. In *International Conference on Machine Learning*, pp. 4460–4469, 2018.

Sarkar, R. Low distortion delaunay embedding of trees in hyperbolic plane. In *International Symposium on Graph Drawing*, pp. 355–366. Springer, 2011.

Sommer, S. Horizontal dimensionality reduction and iterated frame bundle development. In *International Conference on Geometric Science of Information*, pp. 76–83. Springer, 2013.

Tay, Y., Tuan, L. A., and Hui, S. C. Hyperbolic representation learning for fast and efficient neural question answering. In *Proceedings of the Eleventh ACM International Conference on Web Search and Data Mining*, pp. 583–591, 2018.

Thurston, W. P. Geometry and topology of three-manifolds, 1978. Lecture notes. Available at http://library.msri.org/books/gt3m/.

Tifrea, A., Becigneul, G., and Ganea, O.-E. Poincare glove: Hyperbolic word embeddings. In *International Conference on Learning Representations*, 2018.

Tran, D. A. and Vu, K. Dimensionality reduction in hyperbolic data spaces: Bounding reconstructed-information loss. In *Seventh IEEE/ACIS International Conference on Computer and Information Science (icis 2008)*, pp. 133–139. IEEE, 2008.

Wang, M.-X. Laplacian eigenspaces, horocycles and neuron models on hyperbolic spaces, 2021. URL https://openreview.net/forum?id=ZglaBL5inu.

Wu, X. and Charikar, M. Nearest neighbor search for hyperbolic embeddings. *arXiv preprint arXiv:2009.00836*, 2020.

Zhang, M. and Fletcher, T. Probabilistic principal geodesic analysis. In *Advances in Neural Information Processing Systems*, pp. 1178–1186, 2013.