A. Horospherical Projection: Proofs and Discussions

We show that the horospherical projection $\pi^{\mathrm{H}}_{b,p_1,\dots,p_K}$ in Section 3.2.2 is well-defined, shares many nice properties with Euclidean orthogonal projections, and has a closed-form expression.

More specifically, this section is organized as follows. In Appendix A.1, we show that horospherical projections are well-defined (Theorem A.4). In Appendix A.2, we show that they are base-point independent (Corollary A.7), non-expanding (Corollary A.12), and distance-preserving along a family of K-dimensional submanifolds (Corollary A.10). Finally, in Appendix A.4, we explain how they can be computed efficiently in the hyperboloid model.

For an illustration in the case of $K = 2$ ideal points in $\mathbb{H}^3$, see Fig. 5. This figure might help explain the intuitions behind the theorems in Appendix A.1 and Appendix A.2.

A.1. Well-definedness

Recall that given a base point $b \in \mathbb{H}^d$ and $K > 1$ ideal points $\{p_1, \dots, p_K\}$, we would like to define $\pi^{\mathrm{H}}_{b,p_1,\dots,p_K}$ by

$$x \to M \cap S(p_1, x) \cap S(p_2, x) \cap \cdots \cap S(p_K, x),$$

where $M = \mathrm{GH}(b, p_1, \dots, p_K)$ is the target submanifold and $S(p_j, x)$ is the horosphere centered at p_j and passing through x. For this definition to make sense, the intersection in the right hand side must contain exactly one point for each $x \in \mathbb{H}^d$. Unfortunately, this is not the case: In fact, the intersection generally consists of two points. Nevertheless, we will show that there is a consistent way to choose one from these two points, making the function $\pi^{\mathrm{H}}_{b,p_1,\dots,p_K}$ well-defined. This is the result of Theorem A.4.

First, to understand the above intersection, we give a more concrete description of $\cap_j S(p_j, x)$.

Lemma A.1. *Let $P = \mathrm{GH}(p_1, p_2, \dots, p_K)$. Then for every $x \in \mathbb{H}^d$, the intersection of horospheres*

$$S(x) = S(p_1, x) \cap S(p_2, x) \cap \cdots \cap S(p_k, x)$$

is precisely the orbit of x under the group G of rotations around P.

Proof. First, note that every rotation around P preserves the horospheres $S(p_j, x)$ - just like how every rotation around an axis preserves every sphere whose center is on that axis. It follows that $S(x)$ is preserved by G. In particular, the orbit of x under G is contained in $S(x)$.

It remains to show that $S(x)$ contains no other points. To this end, consider any $y \neq x$ in $S(x)$. The perpendicular bisector B of x and y is a totally geodesic hyperplane of $\mathbb{H}^d$ that contains every p_j (because each p_j is intuitively the center of a sphere that goes through x and y). Thus, by the definition of geodesic hull, $B \supset P$. In particular, the reflection through B sends x to y and fixes every point in P.

Now take any geodesic hyperplane A that contains both P and y, so that the reflection through A fixes y and every point in P. Then the composition of the reflections through B and A is a rotation that sends x to y and fixes every point in P. In other words, it is a rotation around P that sends x to y. Therefore, y belongs to the orbit of x under G. □

Corollary A.2. *If $x \in P$ then $S(x) = \{x\}$. Otherwise, let $\pi^{\mathrm{G}}_P(x)$ be the geodesic projection of x onto P, and $Q(x)$ be the geodesic submanifold that orthogonally complements P at $\pi^{\mathrm{G}}_P(x)$. Then $S(x) \subset Q(x)$ and is precisely the (hyper)sphere in $Q(x)$ that is centered at $\pi^{\mathrm{G}}_P(x)$ and passing through x.*

Proof. This follows from Lemma A.1. If $x \in P$ then every rotation around P fixes x, so the orbit of x is just itself.

Now consider the case $x \notin P$. All rotations around P must preserve $\pi^{\mathrm{G}}_P(x)$ and the orthogonal complement $Q(x)$ of P at $\pi^{\mathrm{G}}_P(x)$. Furthermore, when restricted to the space $Q(x)$, these rotations are precisely the rotations in $Q(x)$ around the point $\pi^{\mathrm{G}}_P(x)$. Thus, for every $y \in Q(x)$, the orbit of y under G is a sphere in $Q(x)$ centered at $\pi^{\mathrm{G}}_P(x)$. In particular, $S(x)$, which is the orbit of x, is the sphere in $Q(x)$ that is centered at $\pi^{\mathrm{G}}_P(x)$ and passing through x. □

Corollary A.2 gives the following characterization of the intersection

$$M \cap S(p_1, x) \cap S(p_2, x) \cap \cdots \cap S(p_K, x) = M \cap S(x) :$$

Note that $P = \mathrm{GH}(p_1, \dots, p_K)$ is a geodesic submanifold of $M = \mathrm{GH}(b, p_1, \dots, p_K)$ and that $\dim P = \dim M - 1$. Thus, through every point $y \in P$, there is a unique geodesic α on M that goes through y and is perpendicular to P.

Corollary A.3. *If $x \in P$ then $M \cap S(x) = \{x\}$. Otherwise, let α be the geodesic on M that goes through $\pi^{\mathrm{G}}_P(x)$ and is perpendicular to P. Then $M \cap S(x)$ consists of the two points on α whose distance to $\pi^{\mathrm{G}}_P(x)$ equals $d_{\mathbb{H}}(x, \pi^{\mathrm{G}}_P(x))$.*

Proof. The case $x \in P$ is clear since $x \in M$ and $S(x) = \{x\}$ by Corollary A.2.

For the other case, let $Q(x)$ be the orthogonal complement of P at $\pi^{\mathrm{G}}_P(x)$. Then by Corollary A.2, $S(x)$ is precisely the sphere in $Q(x)$ that is centered at $c(x)$ and passing through x.

Now note that $M \cap Q(x) = \alpha$. Since $S(x) \subset Q(x)$, this gives $M \cap S(x) = M \cap Q(x) \cap S(x) = \alpha \cap S(x)$. We know

that every sphere intersects every line through the center at two points. □

Therefore, to define $\pi^{\mathrm{H}}_{b,p_1,\dots,p_K}(x)$, we just need to choose one of the two poins in $M \cap S(x)$ in a consistent way (so that the map is differentiable). To this end, note that P cuts M into two half-spaces, and exactly one of them contains the base point b. (Recall that $b \in M$ and $b \notin P$ by the "independence" condition). We denote this half by P_b. Then, while $S(x)$ contains two points in M, it only contains one point in P_b:

Theorem A.4. *Let α be the geodesic on M that goes through $\pi^{\mathrm{G}}_P(x)$ and is perpendicular to P. Let α^+ be the half of α contained in P_b. Then α^+ intersects the sphere $S(x)$ at a unique point x', which is also the unique intersection point between P_b and $S(x)$. Thus, we can define $\pi^{\mathrm{H}}_{b,p_1,\dots,p_K}$ by*

$$\pi^{\mathrm{H}}_{b,p_1,\dots,p_K}(x) = \alpha^+ \cap S(x) = P_b \cap S(x).$$

Equivalently, $\pi^{\mathrm{H}}_{b,p_1,\dots,p_K}(x)$ is the point in $M \cap S(x)$ that is strictly closer to b.

Proof. By Corollary A.2, $S(x)$ is a sphere centered at $\pi^{\mathrm{G}}_P(x)$. Then, since α^+ is a geodesic ray starting at $\pi^{\mathrm{G}}_P(x)$, it must intersect $S(x)$ at a unique point x'.

Next, we have

$$P_b \cap S(x) = P_b \cap M \cap S(x) = P_b \cap \alpha \cap S(x) = \alpha^+ \cap S(x),$$

where the first equality holds because $P_b \subset M$, the second because $M \cap S(x) = \alpha \cap S(x)$ by the proof of Corollary A.3, and the third because $P_b \cap \alpha = \alpha^+$. Thus, $P_b \cap S(x)$ is precisely x'.

Finally, let x'' be the other point of $M \cap S(x) = \alpha \cap S(x)$. Then P is the perpendicular bisector of x' and x'' in M. Thus, every point on the same side of P in M as x' (but not on the boundary P) is strictly closer to x' than to x''. By definition, b is one of such point. Thus, $\pi^{\mathrm{H}}_{b,p_1,\dots,p_K}(x)$ is the point in $M \cap S(x)$ that is closer to b. □

A.2. Geometric properties

From Lemma A.1 and Theorem A.4, we obtain another interpretation of $\pi^{\mathrm{H}}_{b,p_1,\dots,p_K}$: It maps $\mathbb{H}^d$ to $P_b \subset M$ by rotating every point $x \in \mathbb{H}^d$ around P until it hits P_b. In other words, we have

Theorem A.5 (The "open book" interpretation)**.** *For any $x \notin P$, let $M_x = \mathrm{GH}(P \cup \{x\})$. Then P cuts M_x into two half-spaces; we denote the half that contains x by P_x. Then, when restricted to P_x, the map*

$$\pi^{\mathrm{H}}_{b,p_1,\dots,p_K} : P_x \to P_b$$

is simply a rotation around P. □

Following this, we call P the *spine* of the horosphere projection. The identity $\mathbb{H}^d = \cup_{x \notin P} P_x$ can be thought of as an *open book decomposition* of $\mathbb{H}^d$ into *pages* P_x that are bounded by the spine P. The horosphere projection $\pi^{\mathrm{H}}_{b,p_1,\dots,p_K}$ then simply acts by collapsing every page onto a specified page P_b.

Here are some consequences of this interpretation:

Corollary A.6. *$\pi^{\mathrm{H}}_{b,p_1,\dots,p_K}$ only depends on the spine P and not specifically on $p_1, \dots p_K$. Thus, when we are not interested in the specific ideal points, we simply write $\pi^{\mathrm{H}}_{b,P}$.*

Proof. As noted above, $\pi^{\mathrm{H}}_{b,p_1,\dots,p_K}(x)$ can be obtained by rotating x around P until it hits P_b. This operation does not used the exact positions of p_j at all. □

Corollary A.7. *The choice of b does not affect the geometry of the projection $\pi^{\mathrm{H}}_{b,P}$. More precisely, for any two base points $b, b' \notin P$, the horosphere projections*

$$\pi^{\mathrm{H}}_{b,P} : \mathbb{H}^d \to P_b \quad \text{and} \quad \pi^{\mathrm{H}}_{b',P} : \mathbb{H}^d \to P_{b'}$$

only differ by a rotation $P_b \to P_{b'}$ around P.

Proof. By Theorem A.5, when restricted to any page P_x, the maps $\pi^{\mathrm{H}}_{b,P} : \mathbb{H}^d \to P_b$ and $\pi^{\mathrm{H}}_{b',P} : \mathbb{H}^d \to P_{b'}$ are just rotations around P. The difference between any two rotations around P is another rotation around P. □

In particular, Corollary A.7 implies Theorem 4.1:

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

As discussed in Section 4, this theorem helps reduces parameters and simplifies the computation of $\pi^{\mathrm{H}}_{b,P}$.

Remark A.8. Corollary A.6 and Corollary A.7 together imply that the HoroPCA algorithm (5) only depends on the geodesic hulls $\mathrm{GH}(p_1), \mathrm{GH}(p_1, p_2), \dots \mathrm{GH}(p_1, \dots, p_K)$ of the ideal points and not the specific ideal points themselves. It follows that, theoretically, the search space of (5) has dimension $dK - \frac{1}{2}K(K+1)$ – the same as the dimension of the space of flags in Euclidean spaces.

In our implementation, for simplicity we parametrize the K ideal points independently, which results in a suboptimal search space dimension $(d-1)K$. Nevertheless, this is still slightly more efficient than the parametrizations used in PGA and BSA, which require have $(d+1)K$-dimensions.

The following corollaries say that horospherical projections share a nice property with Euclidean orthogonal projections: When projecting to a K-dimensional submanifold, they preserve the distances along K dimensions and collapse the distances along the other $d-K$ orthogonal dimensions:

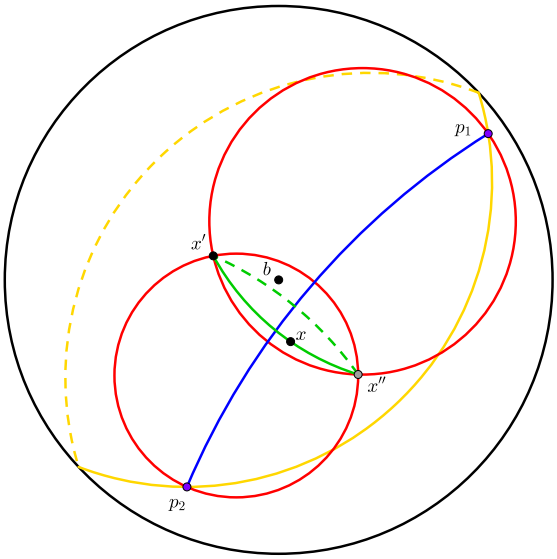


Figure 5: The horospherical projection $\pi^{\mathrm{H}}_{b,p_1,p_2}$ from $\mathbb{H}^3$ to 2 dimensions. Here p_1, p_2 are ideal points, and the base point $b \in \mathbb{H}^3$ is chosen to be the origin of the Poincaré ball. The geodesic hull $\mathrm{GH}(b, p_1, p_2)$ is the hyperbolic plane bounded by the yellow circle. The geodesic between p_1 and p_2 is shown in blue and is called the spine $P = \mathrm{GH}(p_1, p_2)$ of this projection. For any input $x \in \mathbb{H}^3$, the two horospheres $S(p_1, x)$ and $S(p_2, x)$ (shown in red) intersect along a circle $S(x)$ (shown in green). Note that this circle is precisely the set $\{y \in \mathbb{H}^3 : B_{p_j}(y) = B_{p_j}(x), j = 1, 2\}$ and is symmetric around the spine P. It intersects $\mathrm{GH}(b, p_1, p_2)$ at two points x' and x'', which lie on opposite sides of the spine. Since x' belongs to the side containing b, it is closer to b than x'' is. Thus, we define $\pi^{\mathrm{H}}_{b,p_1,p_2}(x)$ to be x'.

Corollary A.9. *$\pi^{\mathrm{H}}_{b,P}$ is invariant under rotations around P. In other words, if a rotation around P takes x to y then $\pi^{\mathrm{H}}_{b,P}(x) = \pi^{\mathrm{H}}_{b,P}(y)$.*

Consequently, every $x \notin P$ belongs to a $(d-K)$-dimensional submanifold that is collapsed to a point by $\pi^{\mathrm{H}}_{b,P}$.

Proof. The open book interpretation tells us that $\pi^{\mathrm{H}}_{b,P}(x)$ and $\pi^{\mathrm{H}}_{b,P}(y)$ are precisely the intersections of P_b with $S(x)$ and $S(y)$, respectively. If y belongs to the rotation orbit $S(x)$ of x then the rotation orbit $S(y)$ of y is the same as $S(x)$. Thus $\pi^{\mathrm{H}}_{b,P}(x) = \pi^{\mathrm{H}}_{b,P}(y)$.

Hence, for every $x \in \mathbb{H}^d$, $S(x)$ is collapsed to a point by $\pi^{\mathrm{H}}_{b,P}$. To see that $\dim S(x) = d-K$ when $x \notin P$, recall that by Corollary A.2, if $Q(x)$ is the orthogonal complement of P at $\pi^{\mathrm{G}}_P(x)$ then $S(x)$ is a hypersphere inside $Q(x)$. Since the ideal points p_j are assumed to be "affinely independent," we have $\dim P = K-1$, so $\dim Q(x) = d-(K-1)$, and $\dim S(x) = \dim Q(x) - 1 = d-K$. □

Corollary A.10. *For every $x \in \mathbb{H}^d$, there exists a K-dimensional totally geodesic submanifold (with boundary) that contains x and is mapped isometrically to P_b by $\pi^{\mathrm{H}}_{b,P}$. If $x \notin P$ then such a manifold is unique.*

Proof. If $x \notin P$ then the submanifold P_x in Theorem A.5 is a geodesic submanifold that contains x and is mapped isometrically to P_b by $\pi^{\mathrm{H}}_{b,P}$. As in the proof of Corollary A.9, we have $\dim P = K-1$ and $\dim P_x = \dim P + 1 = K$.

Since Corollary A.9 implies that the other $d-K$ dimensions are collapsed by $\pi^{\mathrm{H}}_{b,P}$, no other distances from x can be preserved. Thus, P_x is the unique submanifold with the desired properties.

If $x \in P$ then $x \in P_y$ for every $y \notin P$. Note that this means *every* distance from x is preserved by $\pi^{\mathrm{H}}_{b,P}$. □

The following corollaries say that, like Euclidean orthogonal projections, horospherical projections never increase distances. Thus, minimizing distortion is roughly equivalent to maximizing projected distances. This is another motivation for Eq. (4).

Corollary A.11. *For every $x \in \mathbb{H}^d$ and every tangent vector $\vec{v}$ at x,*

$$\|\pi^{\mathrm{H}}_{b,P}(\vec{v})\|_{\mathbb{H}} \leq \|\vec{v}\|_{\mathbb{H}}.$$

Proof. This follows from Corollary A.10 and Corollary A.9:

If $x \in P$ then the proof of Corollary A.10 implies that $\pi^{\mathrm{H}}_{b,P}$ preserves *every* distance from x. Thus, the desired inequality is actually an equality.

If $x \notin P$ then by $\vec{v}$ has an orthogonal decomposition $\vec{v} = \vec{u} + \vec{u}^{\perp}$, where $\vec{u}$ and $\vec{u}^{\perp}$ are tangent and perpendicular to P_x, respectively. By Corollary A.10 and Corollary A.9, $\pi^{\mathrm{H}}_{b,P}$ preserves the length of $\vec{u}$ while collapsing $\vec{u}^{\perp}$ to 0. It follows that $\|\pi^{\mathrm{H}}_{b,P}(\vec{v})\|_{\mathbb{H}} = \|\vec{u}\|_{\mathbb{H}} \le \|\vec{v}\|_{\mathbb{H}}$. □

Corollary A.12. *$\pi^{\mathrm{H}}_{b,P}$ is non-expanding. In other words, for every $x, y \in \mathbb{H}^d$,*

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

Proof. We first show that for any path $\gamma(t)$ in $\mathbb{H}^d$,

$$\mathrm{length}(\pi^{\mathrm{H}}_{b,P}(\gamma)) \le \mathrm{length}(\gamma).$$

Indeed, note that the velocity vector of $\pi^{\mathrm{H}}_{b,P}(\gamma)$ is precisely $\pi^{\mathrm{H}}_{b,P}(\dot{\gamma})$. Then, by Corollary A.11,

$$\begin{aligned}\mathrm{length}(\pi^{\mathrm{H}}_{b,P}(\gamma)) &= \int \|\pi^{\mathrm{H}}_{b,P}(\dot{\gamma}(t))\|_{\mathbb{H}} dt \\ &\le \int \|\dot{\gamma}(t)\|_{\mathbb{H}} dt = \mathrm{length}(\gamma).\end{aligned}$$

Now for any $x, y \in \mathbb{H}^d$, let $\gamma(t)$ be the geodesic segment from x to y. Then $\pi^{\mathrm{H}}_{b,P}(\gamma)$ is a path connecting $\pi^{\mathrm{H}}_{b,P}(x)$ and $\pi^{\mathrm{H}}_{b,P}(y)$. Thus, the projected distance is at most the length of $\pi^{\mathrm{H}}_{b,P}(\gamma)$, which by the above argument is at most $\mathrm{length}(\gamma) = d_{\mathbb{H}}(x, y)$. □

A.3. Detour: A Review of the Hyperboloid Model

Theorem A.4 suggests that $\pi^{\mathrm{H}}_{b,P}(x)$ can be computed in three steps:

1. Find the geodesic projection $\pi^{\mathrm{G}}_{P}(x)$ of x onto P.
2. Find a geodesic ray α^+ on P_b that starts at $\pi^{\mathrm{G}}_{P}(x)$ and is orthogonal to P.
3. Return the unique point on α^+ that is of distance $d_{\mathbb{H}}(x, \pi^{\mathrm{G}}_{P}(x))$ from $\pi^{\mathrm{G}}_{P}(x)$.

It turns out that these subroutines are easier to implement in the *hyperboloid model* instead of the Poincaré model of hyperbolic spaces. Thus, we first briefly review the basic definitions and properties of this model. The readers who are already familiar with the hyperboloid model can skip to Appendix A.4, where we describe the full algorithm.

For a more detailed treatment, see (Thurston, 1978).

Remark A.13. The above steps are slightly different from the ones mentioned in Section 3.2.2. However, by Theorem A.4, these two descriptions are equivalent. We decided to use the "closer-to-b" description in Section 3.2.2 because it is slightly more self-contained, but the actual implementation will be based on the above three steps.

Minkowski spaces We first describe *Minkowski spaces*, which is the embedding space where the hyperboloid model sits in as hypersurfaces.

The $(d+1)$-dimensional Minkowski space $\mathbb{R}^{1,d}$ is like the flat Euclidean space $\mathbb{R}^{1+d}$ except with a dot product that has a negative sign in the first coordinate. More precisely, $\mathbb{R}^{1,d}$ is the vector space $\mathbb{R}^{1+d}$ equipped with the indefinite, non-degenerate bilinear form

$$B\left((t, x_1, \ldots, x_d), (u, y_1, \ldots, y_d)\right) = -tu + \sum_{i=1}^{d} x_i y_i,$$

which serves as the "dot product."

Like with Euclidean spaces, the quantity $B(\vec{v}, \vec{v})$ is called the *(Minkowski) squared norm* of the vector $\vec{v}$. Two vectors $\vec{u}, \vec{v} \in \mathbb{R}^{1,d}$ are called *orthogonal* if $B(\vec{u}, \vec{v}) = 0$. The *orthogonal complement* of a linear subspace $V \subset \mathbb{R}^{1,d}$ is the set $V^{\perp} = \{\vec{w} \in \mathbb{R}^{1,d} : B(\vec{w}, \vec{v}) = 0 \text{ for every } \vec{v} \in V\}$, which still has dimension $\dim \mathbb{R}^{1,d} - \dim V$. However, unlike in Euclidean spaces, vectors in $\mathbb{R}^{1,d}$ can have negative or zero squared norms, and linear subspaces can intersect their orthogonal complements.

Types of vectors in $\mathbb{R}^{1,d}$ Vectors with negative, zero, and positive (Minkowski) squared norms are called *time-like*, *light-like*, and *space-like*, respectively. Time-like and light-like vectors together form a solid double cone in $\mathbb{R}^{1,d}$.

If $\vec{v}$ is a time-like or light-like vector, we call it *future-pointing* if its first coordinate is positive, otherwise we call it *past-pointing*. It follows from Cauchy-Schwarz inequality that.

Proposition A.14. *If $\vec{u}, \vec{v}$ are future-pointing time-like or light-like vectors then $B(\vec{u}, \vec{v}) < 0$.*

A linear subspace V of $\mathbb{R}^{1,d}$ is called *space-like* if every non-zero vector in V is space-like. In that case, the bilinear form $B(\cdot, \cdot)$ restricts to a positive-definite bilinear form on V, thus making it isometric to an Euclidean vector space. It follows from Proposition A.14 that

Proposition A.15. *The orthogonal complement of a time-like vector is a space-like linear subspace.*

The hyperboloid model of hyperbolic spaces We are now ready to introduce the hyperboloid model $\mathbb{H}^d$. It sits inside $\mathbb{R}^{1,d}$ in a similar way to how the unit sphere $\mathbb{S}^d$ sits inside the Euclidean space $\mathbb{R}^{1+d}$.

Definition A.16. The *hyperboloid model* of d-dimensional hyperbolic spaces is the set $\mathbb{H}^d$ of future-pointing vectors in $\mathbb{R}^{1,d}$ with Minkowsi squared norm -1.

Remark A.17. For the rest of Appendix A, we will use $\mathbb{H}^d$ to denote this hyperboloid model as a subset of $\mathbb{R}^{1,d}$, and not the abstract hyperbolic space or its other models (e.g.

Poincaré ball). This should not lead to any ambiguities because we will not work with any other model.

The following properties of $\mathbb{H}^d$ further illustrate the analogy with spheres in Euclidean spaces.

Proposition A.18. *For every $\vec{x} \in \mathbb{H}^d$, the tangent space $T_{\vec{x}}\mathbb{H}^d$ of $\mathbb{H}^d$ at $\vec{x}$ is (parallel to) the orthogonal complement of $\vec{x}$ in $\mathbb{R}^{1,d}$.*

Proposition A.19. *When restricted to each tangent space of $\mathbb{H}^d$, the bilinear form $B(\cdot,\cdot)$ is positive definite. This defines a Riemannian metric on $\mathbb{H}^d$ which has constant curvature -1. The* stereographic projection

$$(t, x_1, \dots, x_d) \to \left(\frac{x_1}{1+t}, \dots, \frac{x_d}{1+t}\right)$$

is an isometry between $\mathbb{H}^d$ and the Poincaré model. Its inverse map is given by

$$(y_1, \dots, y_d) \to \frac{(1 + \sum_i y_i^2, 2y_1, \dots, 2y_d)}{1 - \sum_i y_i^2}.$$

Proposition A.20. *Every k-dimensional geodesic submanifold of $\mathbb{H}^d$ is the intersection of $\mathbb{H}^d$ with a $(k+1)$-dimensional linear subspace of $\mathbb{R}^{1,d}$.*

In the hyperboloid model, the ideal points of hyperbolic spaces are represented by light-like *directions* (instead of individual vectors):

Proposition A.21. *Each ideal point of $\mathbb{H}^d$ is represented by a 1-dimensional linear subspace spanned by some $\vec{v} \in \mathbb{R}^{1,d}$ with $B(\vec{v}, \vec{v}) = 0$. The map*

$$(t, x_1, \dots, x_d) \to \left(\frac{x_1}{t}, \dots, \frac{x_d}{t}\right)$$

gives a correspondence between a light-like vector $\vec{v} \in \mathbb{R}^{1,d}$ and an ideal point $p \in \mathbb{S}^{d-1}_\infty$ in the Poincaré model that is represented by $\operatorname{span}(\vec{v})$. This correspondence is compatible with the stereographic projection in Proposition A.19. Its inverse map is given by

$$(y_1, \dots, y_d) \to (1, y_1, \dots, y_d).$$

We conclude this section by noting that geodesic hulls in the hyperboloid model are closely related to linear spans in Minkowski space:

Proposition A.22. *Let S be a set of vectors that are either in $\mathbb{H}^d$ or represent ideal directions of $\mathbb{H}^d$. Then the geodesic hull of S in $\mathbb{H}^d$ is the intersection of $\mathbb{H}^d$ with the linear span of S.*

In particular, since the spine P in our setting (Theorem A.4) is the geodesic hull of some ideal points, it is cut out by the linear span of the corresponding ideal directions.

Algorithm 1 Horospherical Projection

Input: point x, ideal points $\{p_1, \dots, p_K\}$, base point b.
$z \leftarrow \pi^{\text{Mink}}_{\text{span}(p_1,\dots,p_K)}(x)$ {orthogonal projection in ambient space}
$w \leftarrow z/\|z\|$ {rescale to a vector on the hyperboloid}
$u \leftarrow b - w$ {subtraction in ambient space}
$u \leftarrow u - \pi^{\text{Mink}}_{\text{span}(p_1,\dots,p_K)}(u)$ {orthogonal projection in ambient space}
$u \leftarrow u/\|u\|$ {make u the unit tangent vector orthogonal to the spine P}
$y \leftarrow \exp_w(d_{\mathbb{H}}(x, w) \cdot u)$ {exponential map in $\mathbb{H}^d$}

return y

A.4. Computation of Horospherical Projections

Now we describe how the three steps mentioned at the beginning of Appendix A.3 can be implemented in the hyperboloid model. The results of this section are summarized in Algorithm 1. To transfer back and forth between the hyperboloid and Poincaré models, we use the formulas in Proposition A.19 and Proposition A.21.

Step 1: Computing geodesic projections We first describe how to compute the geodesic (or closest-point) projection from $\mathbb{H}^d$ to a geodesic submanifold P in $\mathbb{H}^d$. This process is very similar to how geodesic projections work in Euclidean spheres: We first perform an orthogonal projection onto the linear subspace V that cuts out P, then rescale the result to get a vector on $\mathbb{H}^d$.

Generally, orthogonal projections in Minkowski spaces are very similar to those in Euclidean spaces. However, since vectors in $\mathbb{R}^{1,d}$ can have norm zero, the orthogonal projection π^{Mink}_V is not well-defined for every subspace V. Thus, a little extra argument is needed.

Proposition A.23 (Orthogonal projections onto time-containing linear subspaces)**.** *Let V be a linear subspace of $\mathbb{R}^{1,d}$ that contains some time-like vectors. Then*

1. *$V \cap V^\perp = \{0\}$. Consequently, we can define a linear orthogonal projection $\pi^{\text{Mink}}_V : \mathbb{R}^{1,d} \to V$ as follows: Since $\mathbb{R}^{1,d} = V \oplus V^\perp$, every vector $\vec{x} \in \mathbb{R}^{1,d}$ can be uniquely written as $\vec{x} = \vec{z} + \vec{n}$ for some $\vec{z} \in V$ and $\vec{n} \in V^\perp$. Then, we let $\pi^{\text{Mink}}_V(\vec{x}) := \vec{z}$.*
2. *Let A be a matrix whose column vectors form a linear basis of V. Let B be the $(1+d) \times (1+d)$ symmetric matrix associated to the bilinear form B, i.e. the diagonal matrix with entries $(-1, 1, 1, 1, \dots, 1)$. Then $A^\top BA$ is non-singular, and the linear projection π^{Mink}_V is given by $\vec{x} \to A(A^\top BA)^{-1}A^\top B\vec{x}$*
3. *If $\vec{x}$ is a future-pointing time-like vector then so is $\pi^{\text{Mink}}_V(\vec{x})$.*

Proof.

1. Let $\vec{v}$ be a time-like vector in V. Then $\vec{v}^\perp$ is space-like by Proposition A.15. Since $V \cap V^\perp \subset V^\perp \subset \vec{v}^\perp$, the intersection $V \cap V^\perp$ must be space-like. On the other hand, for every $\vec{w} \in V \cap V^\perp$, we have $B(\vec{w}, \vec{w}) = 0$, so $\vec{w}$ must be light-like. It follows that $V \cap V^\perp = \{0\}$. The $\mathbb{R}^{1,d} = V \oplus V^\perp$ part of the claim is a standard linear algebra fact.
2. This follows from the same argument that deduces the formula for Euclidean orthogonal projections.
3. To avoid cumbersome notations, let $\vec{z} = \pi_V^{\text{Mink}}(\vec{x})$. Then since $\vec{x} - \vec{z}$ is orthogonal to V and in particular to $\vec{z}$, we have the Pythagorean formula

$$B(\vec{z}, \vec{z}) + B(\vec{x} - \vec{z}, \vec{x} - \vec{z}) = B(\vec{x}, \vec{x}).$$

On the other hand, since $\vec{x} - \vec{z}$ is orthogonal to V and in particular orthogonal to some time-like vectors in V, by Proposition A.15, it must be space-like. Thus, if $B(\vec{x}, \vec{x}) < 0$ then the above equation implies $B(\vec{z}, \vec{z}) < 0$, which means $\vec{z}$ is time-lile.
Now either $\vec{z}$ or $-\vec{z}$ is future-pointing. In the latter case, since $\vec{x}$ is future-pointing, Proposition A.14 implies $B(-\vec{z}, \vec{x}) < 0$. On the other hand, we have $B(\vec{z}, \vec{x}) = B(\vec{z}, \vec{z}) < 0$, contradicting the above inequality. Thus, $\vec{z}$ is future-pointing. □

Proposition A.24 (Geodesic projections in $\mathbb{H}^d$)**.** *Let P be a geodesic submanifold of $\mathbb{H}^d$. Recall that by Proposition A.20, $P = \mathbb{H}^d \cap V$ for some linear subspace V of $\mathbb{R}^{1,d}$. Then for every $\vec{x} \in \mathbb{H}^d$, the geodesic projection $\pi_P^{\text{G}}(\vec{x})$ of $\vec{x}$ onto P in $\mathbb{H}^d$ is given by*

$$\pi_P^{\text{G}}(\vec{x}) = \frac{\vec{z}}{\sqrt{-B(\vec{z}, \vec{z})}},$$

where $\vec{z} = \pi_V^{\text{Mink}}(\vec{x})$ is the linear orthogonal projection of $\vec{x}$ onto V.

Proof. Again, to avoid cumbersome notations, let

$$\vec{w} = \frac{\vec{z}}{\sqrt{-B(\vec{z}, \vec{z})}}.$$

We will show that $\vec{w} = \pi_P^{\text{G}}(\vec{x})$. First, note that since $\vec{z}$ is a future-pointing time-like vector by Proposition A.23, $\vec{w} \in \mathbb{H}^d$. Since $\vec{w} \in V$, it follows that $\vec{w} \in P$.

Let W be the linear span of $\vec{w}$ and $\vec{x}$, so that $W \cap \mathbb{H}^d$ is the geodesic γ in $\mathbb{H}^d$ that connects $\vec{w}$ and $\vec{x}$. Note that $\vec{w}$ and $\vec{x} - \vec{z}$ form a basis of W. They are both orthogonal to the tangent space $T_{\vec{w}}P$ of P at $\vec{w}$ because:

- By Proposition A.18, $\vec{w}$ is orthogonal to every tangent vector of $\mathbb{H}^d$ at $\vec{w}$.
- $T_{\vec{w}}P$ is contained in V, which is orthogonal to $\vec{x} - \vec{z}$.

Thus, W is orthogonal to $T_{\vec{w}}P$. It follows that the geodesic γ is orthogonal to P, which means $\vec{w} = \pi_P^{\text{G}}(\vec{x})$. □

Step 2: Finding orthogonal geodesic ray Recall that if P is a geodesic submanifold of $\mathbb{H}^d$ and $\vec{b} \in \mathbb{H}^d$ is a point not in P, then the geodesic hull M of $P \cup \{\vec{b}\}$ in $\mathbb{H}^d$ is a geodesic submanifold of $\mathbb{H}^d$ with dimension $\dim P + 1$. Thus, P cuts M into two halves, and we denote the half that contains $\vec{b}$ by P_b.

Given a point $\vec{w} \in P$, there exists a unique geodesic ray α^+ that starts at $\vec{w}$, stays on P_b, and is orthogonal to P. In this section, we describe how to compute α^+.

Recall that by Proposition A.20, $P = V \cap \mathbb{H}^d$ for some linear subspace V.

Proposition A.25. *The vector*

$$\vec{u} = (\vec{b} - \vec{w}) - \pi_V^{\text{Mink}}(\vec{b} - \vec{w})$$

is tangent to M and orthogonal to P at $\vec{w}$. Furthermore, it points toward the side of P_b.

Proof. By construction, $\vec{u}$ is orthogonal to V. Note that V contains both $\vec{w}$ and the tangent space $T_{\vec{w}}P$ of P at $\vec{w}$. Thus, $\vec{u}$ is orthogonal to both of them. Together with Proposition A.18, this implies that $\vec{u}$ is a tangent vector of $\mathbb{H}^d$ at $\vec{w}$ that is orthogonal to P.

Next, note that M is the intersection of $\mathbb{H}^d$ with the linear span of $V \cup \{\vec{b}\}$, and since $\vec{b}$ and $\vec{w}$ both belong to this linear span, so does $\vec{u}$. Thus, since $\vec{u}$ is a tangent vector of $\mathbb{H}^d$ at $\vec{w}$, it must be tangent to M.

By construction, $\vec{u}$ points from $\vec{w}$ toward the side of $\vec{b}$ instead of away from it. More rigorously, note that by essentially the same argument as above,

$$\vec{a} \coloneqq (\vec{b} - \vec{w}) - \pi_{\vec{w}}^{\text{Mink}}(\vec{b} - \vec{w})$$

is a tangent vector at $\vec{w}$ of the geodesic on $\mathbb{H}^d$ that goes from $\vec{w}$ to $\vec{b}$. Also, note that

$$\vec{a} - \vec{u} = \pi_V^{\text{Mink}}(\vec{b} - \vec{w}) - \pi_{\vec{w}}^{\text{Mink}}(\vec{b} - \vec{w}).$$

Since $V = T_{\vec{w}}P \oplus \text{span}(\vec{w})$ is an orthogonal decomposition, this means

$$\vec{a} - \vec{u} = \pi_{T_{\vec{w}}P}^{\text{Mink}}(\vec{b} - \vec{w}),$$

which in particular implies $\vec{a} - \vec{u} \in T_{\vec{w}}P$. It follows that $\vec{u}$ is the projection of $\vec{a}$ onto the orthogonal complement of $T_{\vec{w}}P$ in $T_{\vec{w}}\mathbb{H}^d$.

In an Euclidean vector space, the dot product of any vector with its projection onto any direction is positive unless the projection is zero. Since $T_{\vec{w}}\mathbb{H}^d$ is a space-like subspace, this statement applies to $\vec{a}$ and $\vec{u}$. Note that $\vec{u}$ cannot be zero because otherwise $\vec{b}$ would belong to V and hence P. Thus, we conclude that $\vec{a} \cdot \vec{u} > 0$, which means that $\vec{u}$ points toward the side of P_b. □

Thus, $\vec{u}$ is the tangent vector at $\vec{w}$ of the desired ray α^+. To compute points on this ray, we can use the exponential map at $\vec{w}$.

Step 3: The exponential map Finally, given a distance d and a tangent vector $\vec{u}$ at $\vec{w}$ of a geodesic ray α^+ in $\mathbb{H}^d$, we need to compute the point on α^+ that is of hyperbolic distance d from $\vec{w}$. This is based on the following lemma:

Lemma A.26. *Suppose that $\vec{u}$ is a unit tangent vector of $\mathbb{H}^d$ at $\vec{w}$. Then*

$$\alpha(t) = (\cosh t)\vec{w} + (\sinh t)\vec{u}$$

is a unit-speed geodesic with $\alpha(0) = \vec{w}$ and $\dot{\alpha}(0) = \vec{u}$.

Proof. Note that $\vec{w}$ and $\vec{u}$ are orthogonal by Proposition A.18. The facts that $\alpha(0) = \vec{w}$, $\dot{\alpha}(0) = \vec{u}$, and that $\alpha(t)$ is a unit-speed curve on $\mathbb{H}^d$ follow from simple, direct computations. To see that $\alpha(t)$ is a geodesic, note that it is the intersection of $\mathbb{H}^d$ with the linear subspace $\mathrm{span}(\vec{w}, \vec{u})$ and apply Proposition A.20. □

It follows that

Proposition A.27. *If $\vec{u}$ is the tangent vector at the starting point $\vec{w}$ of a geodesic ray α^+ in $\mathbb{H}^d$ then*

$$(\cosh d)\vec{w} + (\sinh d)\frac{\vec{u}}{\sqrt{B(\vec{u},\vec{u})}}$$

is the point on α^+ that is of distance d from $\vec{w}$. □

From the results in this section, we conclude that Algorithm 1 computes the horospherical projection $\pi^{\mathrm{H}}_{b,P}(x)$.

B. Geodesic Projection: Distortion Analysis

We show that in hyperbolic geometry, the geodesic projection π^{G}_M (also called closest-point projection) to a geodesic submanifold M always strictly decreases distances unless the inputs are already in M; furthermore, all paths of distance at least r from M get shorter by at least $\cosh r$ times under the projection. Note that most ideas and results in this section already exist in the literature. For a comprehensive treatment, see (Thurston, 1978).

This section is organized as follows. First, we use basic hyperbolic trigonometry to prove a bound on projected distances onto geodesics in dimension two (Proposition B.3). Then we generalize this bound to higher dimensions (Theorem B.4) and obtain an infinitesimal version of it (Corollary B.5). The main theorem (Theorem B.6) then follows immediately.

Hyperbolic trigonometry in two dimensions We first recall the laws of sine and cosine for triangle in hyperbolic planes:

Proposition B.1. *Consider any triangle in $\mathbb{H}^2$ with edge lengths A, B, C. Let α, β, γ be the angles at the vertices opposite to edges A, B, C. Then*

$$\frac{\sinh A}{\sin\alpha} = \frac{\sinh B}{\sin\beta} = \frac{\sinh C}{\sin\gamma},$$

$$\cos\gamma = \frac{\cosh A\cosh B - \cosh C}{\sinh A\sinh B}.$$

Proof. See (Thurston, 1978), section 2.6. □

Proposition B.2. *Consider any quadrilateral in $\mathbb{H}^2$ with vertices x, y, z, w (in that order). Suppose that the angles at z and w are 90 degrees. Then*

$$\begin{aligned}\cosh(xy) =& \cosh(zw)\cosh(xw)\cosh(yz)\\ &- \sinh(xw)\sinh(yz).\end{aligned}$$

Here $\cosh(xy)$ is a shorthand for $\cosh(d_{\mathbb{H}}(x, y))$.

Proof. Applying the laws of sine to triangle wxz gives

$$\frac{\sinh(xw)}{\sin\angle(zx, zw)} = \frac{\sinh(xz)}{\sin\angle(wx, wz)} = \sinh(xz),$$

so

$$\sin\angle(zx, zw) = \frac{\sinh(xw)}{\sinh(xz)}.$$

On the other hand, applying the law of cosine to triangle xyz gives

$$\cos\angle(zx, zy) = \frac{\cosh(zx)\cosh(zy) - \cosh(xy)}{\sinh(zx)\sinh(zy)}.$$

Now note that $\sin\angle(zx, zw) = \cos\angle(zx, zy)$. Thus,

$$\frac{\sinh(xw)}{\sinh(xz)} = \frac{\cosh(zx)\cosh(zy) - \cosh(xy)}{\sinh(zx)\sinh(zy)},$$

which can be simplified to

$$\cosh(zx)\cosh(zy) - \cosh(xy) = \sinh(xw)\sinh(zy).$$

Finally, note that the law of cosine for triangle wxz gives

$$\frac{\cosh(wx)\cosh(wz) - \cosh(xz)}{\sinh(wx)\sinh(wz)} = \cos\angle(wx, wz) = 0,$$

so $\cosh(xz) = \cosh(wx)\cosh(wz)$. Plugging this into the above equation gives

$$\begin{aligned}&\cosh(wx)\cosh(wz)\cosh(zy) - \cosh(xy)\\ =& \sinh(xw)\sinh(zy),\end{aligned}$$

which can be arranged to get the desired identity. □

Now we are ready to prove a bound on the projected distances onto a geodesic in $\mathbb{H}^2$:

Proposition B.3. *Let L be a geodesic in $\mathbb{H}^2$. Consider any $x, y \in \mathbb{H}^2$ that are on the same side of L and are of distances at least r from L. Let $d = d_{\mathbb{H}}(x, y)$ and $d' = d_{\mathbb{H}}(\pi^{\mathrm{G}}_L(x), \pi^{\mathrm{G}}_L(y))$ denote the original and projected distances of x and y. Then,*

$$\sinh(d'/2) \le \frac{\sinh(d/2)}{\cosh r}.$$

Proof. Let $r_x = d_{\mathbb{H}}(x, \pi^{\mathrm{G}}_L(x))$ and $r_y = d_{\mathbb{H}}(y, \pi^{\mathrm{G}}_L(y))$. Then by Proposition B.2,

$$\cosh d = \cosh d' \cosh r_x \cosh r_y - \sinh r_x \sinh r_y.$$

Since $\cosh a \cosh b - \sinh a \sinh b = \cosh(a-b)$, this gives

$$\cosh d = (\cosh d' - 1)\cosh r_x \cosh r_y + \cosh(r_x - r_y).$$

Subtracting 1 from both sides and using $\cosh(a) - 1 = 2\sinh^2(a/2)$ gives

$$\begin{aligned}&2\sinh^2(d/2)\\ =&2\sinh^2(d'/2)\cosh r_x \cosh r_y + 2\sinh^2((r_x - r_y)/2).\end{aligned}$$

Thus,

$$\sinh^2(d/2) \ge \sinh^2(d'/2)\cosh r_x \cosh r_y.$$

Since $r_x, r_y \ge r$, this implies

$$\sinh^2(d/2) \ge \sinh^2(d'/2)\cosh^2 r,$$

which is equivalent to the desired inequality. □

General dimensions Now we show that Proposition B.3 generalizes to higher dimensions.

Theorem B.4. *Let $M \subset \mathbb{H}^d$ be a totally geodesic submanifold. Consider any $x, y \in \mathbb{H}^d$ that are of distances at least r from M. Let $d = d_{\mathbb{H}}(x, y)$ and $d' = d_{\mathbb{H}}(\pi^{\mathrm{G}}_M(x), \pi^{\mathrm{G}}_M(y))$ denote the original and projected distances of x and y. Then,*

$$\sinh(d'/2) \le \frac{\sinh(d/2)}{\cosh r}.$$

Proof. This is clearly true if $\pi^{\mathrm{G}}_M(x) = \pi^{\mathrm{G}}_M(y)$ or if both x and y belong to M. Thus, without loss of generality, we can assume that $\pi^{\mathrm{G}}_M(x) \ne \pi^{\mathrm{G}}_M(y)$ and $\pi^{\mathrm{G}}_M(x) \ne x$.

Let $L \subset M$ be the geodesic joining $\pi^{\mathrm{G}}_M(x)$ and $\pi^{\mathrm{G}}_M(y)$, so that $d_{\mathbb{H}}(x, M) = d_{\mathbb{H}}(x, L)$ and $d_{\mathbb{H}}(y, M) = d_{\mathbb{H}}(y, L)$. Let L_x be the half-plane of $\mathrm{GH}(L \cup \{x\})$ that is bounded by L and contains x. Note that if we rotate y around L until it hits L_x at a point $\bar{y}$ then

(i) $\pi^{\mathrm{G}}_L(\bar{y}) = \pi^{\mathrm{G}}_L(y)$,

(ii) $d_{\mathbb{H}}(\bar{y}, L) = d_{\mathbb{H}}(y, L)$,

(iii) $\pi^{\mathrm{H}}_{x,L}(y) = \bar{y}$ by the "open book interpretation" of horospherical projections (see the beginning of Appendix A.2).

Thus, using (i), (ii) and applying Proposition B.3 to $L, x, \bar{y}$ gives,

$$\sinh(d'/2) \le \frac{\sinh(d_{\mathbb{H}}(x, \bar{y})/2)}{\cosh r}.$$

By (iii) and Corollary A.12,

$$d_{\mathbb{H}}(x, \bar{y}) \le d_{\mathbb{H}}(x, y).$$

Combining these two inequalities gives the desired inequality. □

Since $\sinh(t) \approx t$ as $t \to 0$, applying Theorem B.4 as $y \to x$ gives

Corollary B.5. *Under the geodesic projection $\pi^{\mathrm{G}}_M(\cdot)$, every tangent vector $\vec{v}$ at x gets at least $\cosh(d_{\mathbb{H}}(x, M))$ times shorter.* □

This implies the main theorem of this section:

Theorem B.6. *Let $\gamma(t)$ be any smooth path in $\mathbb{H}^d$ that stays a distance at least r away from M. Then*

$$\mathrm{length}(\pi^{\mathrm{G}}_M(\gamma)) \le \frac{\mathrm{length}(\gamma)}{\cosh r}.$$

In particular, $\mathrm{length}(\pi^{\mathrm{G}}_M(\gamma)) < \mathrm{length}(\gamma)$ unless γ is entirely in M.

Proof. Note that the velocity vector of $\pi^{\mathrm{G}}_M(\gamma)$ is precisely $\pi^{\mathrm{G}}_M(\dot{\gamma})$. Then, by Corollary B.5,

$$\begin{aligned}\mathrm{length}(\pi^{\mathrm{G}}_M(\gamma)) &= \int \|\pi^{\mathrm{G}}_M(\dot{\gamma}(t))\|_{\mathbb{H}} dt\\ &\le \int \frac{\|\dot{\gamma}(t)\|_{\mathbb{H}}}{\cosh r} dt = \frac{\mathrm{length}(\gamma)}{\cosh r}.\end{aligned}$$

Since $\cosh r \ge 1$, this implies

$$\mathrm{length}(\pi^{\mathrm{G}}_M(\gamma)) \le \mathrm{length}(\gamma).$$

This can only be an equality if the $\cosh r$ factor is constantly 1, which means γ stays in M. □

When γ is a geodesic, this gives a proof of Proposition 3.5:

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:*

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

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

	Sarkar			PyTorch		
dimension	2	10	50	2	10	50
Balanced Tree	0.088	0.136	0.152	0.097	0.022	0.022
Phylo Tree	0.639	0.639	0.638	0.292	0.286	0.284
Diseases	0.247	0.212	0.210	0.126	0.054	0.055
CS Ph.D.	0.388	0.384	0.380	0.189	0.140	0.139

Table 4: Average distortion for embeddings learned with different methods. Sarkar refers to combinatorial embeddings learned with Sarkar's construction (Sarkar, 2011; Sala et al., 2018) while PyTorch refers to embeddings learned using the PyTorch code from (Gu et al., 2018).

Remark B.7. The above proposition does *not* imply that if x and y are of distances at least r from M then $d_{\mathbb{H}}(x, y)$ gets at least $\cosh r$ times shorter under the geodesic projection. This is because the geodesic segment between x and y can get closer to M in the middle. This is inevitable when x and y are extremely far apart: For example, if $d_{\mathbb{H}}(x, M) = d_{\mathbb{H}}(y, M) = r$, then the triangle inequality implies

$$d_{\mathbb{H}}(\pi^{\mathrm{G}}_M(x), \pi^{\mathrm{G}}_M(y)) \geq d_{\mathbb{H}}(x, y) - 2r.$$

Hence, if $d_{\mathbb{H}}(x, y) \gg r$ then the shrink factor cannot be much smaller than 1. Generally, the rule of thumb is that geodesic projections shrink distances by a big factor if the input points are not very far from each other but quite far from the target submanifold.

C. Additional Experimental Details

We first provide more details about our experimental setup, including a description of how we implemented different methods and then provide additional experimental results.

C.1. Datasets

Embedding The datasets we use are structured as graphs with nodes and edges. We map these graphs to the hyperbolic space using different embedding methods. The datasets from (Sala et al., 2018) can be found in the open-source implementation[7] and we compute hyperbolic embeddings for different dimensions using the PyTorch code from (Gu et al., 2018). We also consider embeddings computed with Sarkar's combinatorial construction (Sarkar, 2011) and report the average distortion measure with respect to the original graph distances in Table 4.

For classification experiments, we follow the experimental protocol in (Cho et al., 2019) and use the embeddings provided in the open-source implementation.[8]

[7] https://github.com/HazyResearch/hyperbolics

[8] https://github.com/hhcho/hyplinear

Centering For simplicity, we center the input embeddings so that they have a Fréchet mean of zero. This makes computations of projections more efficient. To do so, we compute the Fréchet mean of input hyperbolic points using gradient-descent, and then apply an isometric hyperbolic reflection that maps the mean to the origin.[9]

C.2. Implementation Details in the Poincaré Model

Appendix A.4 (in particular, Proposition A.24 and Algorithm 1) describes how geodesic projections and horospherical projections can be computed efficiently in the hyperboloid model. However, because the Poincaré ball model is useful for visualizations and is popular in the ML literature, here we also give high-level descriptions of simple alternative methods to compute these projections directly in the Poincaré model.

This subsection is organized as follows. We first describe an implementation of geodesic projections in the Poincaré model (Appendix C.2.1). We then describe our implementation of all baseline methods, which rely on geodesic projections (Appendix C.2.2). Finally, we describe how HoroPCA could also be implemented in the Poincaré model (Appendix C.2.3).

C.2.1. Computing Geodesic Projections

Recall that geodesic projections map points to a target submanifold such that the projection of a point is the point on the submanifold that is closest to it. In the Poincaré model, when the target submanifold goes through the origin, geodesic projections can be computed efficiently as follows.

Consider any geodesic submanifold P that contains the origin in the Poincaré Ball: this must be a linear space since any geodesic that goes through the origin in this model is a straight-line. The Euclidean reflection with respect to P is a hyperbolic isometry (i.e. preserves distances). Given any point x, we can compute its Euclidean (and hyperbolic) reflection $r_P(x)$ with respect to P. Then, the (hyperbolic geometry) midpoint between x and $r_P(x)$ belongs to P and is the geodesic projection of x onto P. There is a closed-form expression for this midpoint, which can be derived using Mobïus operations.

C.2.2. Implementation of Baselines

We now detail our baseline implementation.

PCA and tPCA We used the Singular Value Decomposition (SVD) PyTorch implementation to implement both PCA and tPCA. Before using the SVD algorithm, tPCA maps points to the tangent space at the Frćhet mean using

[9] This reflection can be computed in closed-form using circle inversions, see for instance Section 2 in (Sala et al., 2018).

	Balanced Tree		Phylo Tree		Diseases		CS Ph.D.	
	distortion (↓)	variance (↑)	distortion (↓)	variance (↑)	distortion (↓)	variance (↑)	distortion (↓)	variance (↑)
PCA	0.91	1.09	0.99	0.28	0.95	0.75	0.96	1.25
tPCA	0.83	4.01	0.27	62.54	0.64	35.45	0.55	86.10
PGA	0.67 ± 0.013	11.15 ± 0.941	$\underline{0.08} \pm 0.001$	$\underline{1110.30} \pm 0.577$	0.55 ± 0.006	56.82 ± 0.493	0.25 ± 0.002	270.28 ± 0.487
BSA	0.58 ± 0.018	16.74 ± 1.278	$\underline{0.08} \pm 0.000$	1110.05 ± 0.130	0.54 ± 0.008	57.73 ± 0.486	0.25 ± 0.001	270.98 ± 0.363
hMDS	<u>0.12</u>	<u>58.62</u>	<u>0.08</u>	498.05	<u>0.32</u>	**568.22**	<u>0.18</u>	**1241.34**
HoroPCA	**0.06** $\pm$ **0.010**	**62.72** $\pm$ 0.924	**0.01** $\pm$ 0.000	**1190.27** $\pm$ 2.974	**0.09** $\pm$ 0.010	<u>161.32</u> $\pm$ 1.185	**0.05** $\pm$ 0.004	<u>994.69</u> $\pm$ 321.359

Table 5: Dimensionality reduction results on 10-dimensional combinatorial embeddings (Sarkar's construction) reduced to two dimensions. Results are averaged over 5 runs for non-deterministic methods. Best in **bold** and second best <u>underlined</u>.

the logarithmic map. Having centered the data, this mapping can be done using the logarithmic map at the origin which is simply:

$$\log_{\mathbf{o}}(x) = \tanh^{-1}(||x||)\frac{x}{||x||}.$$

PGA and BSA Both PGA and BSA rely on geodesic projections. When the data is centered, the target submanifolds in BSA and PGA are simply linear spaces going through the origin. We can therefore compute the geodesic projections in these methods using the computations described above. To test their sensitivity to the base point, we also implemented two baselines that perturb the base point, by adding Gaussian noise.

hMDS To implement hMDS, we simply implemented the formulas in the original paper (see Algorithm 2 in (Sala et al., 2018)).

hAE For hyperbolic autoencoders, we used two fully-connected layers following the approach of (Ganea et al., 2018). One layer was used for encoding into the reduced number of dimensions, and one layer was used for decoding into the original number of dimensions. We trained these networks by minimizing the reconstruction error, and used the intermediate hidden representations as low-dimensional representations.

C.2.3. HoroPCA Implementation

The definition of horospherical projections onto $K > 1$ directions involves taking the intersection of K horospheres $S(p_1, x) \cap \cdots \cap S(p_K, x)$ (Section 3.2.2). Although Appendix A.4 shows that horospherical projections can be computed efficiently in the hyperboloid model without actually computing these intersections, it is also possible to implement HoroPCA by directly computing these intersections. This can be done in a simple iterative fashion.

For example, in the Poincaré ball, horospheres can be represented as Euclidean spheres. Note that the intersection of two d-dimensional Euclidean hyperspheres $S(p_0, r_0)$ and $S(p_1, r_1)$ is a $d-1$-dimensional hypersphere which can be represented by a center, a radius, and $d-1$-dimensional subspace. The radius is uniquely determined by the radii of the original spheres r_0, r_1 and the distance between their centers $\|p_0 - p_1\|$ and analyzing the 2-dimensional case, for which there is a simple closed formula. The center can similarly be found as a weighted combination of the centers p_0, p_1 based on the formula for the 2D case. Lastly, this $d-1$-dimensional subspace can be represented by noting it is the orthogonal space to the vector $p_1 - p_0$. Thus the intersection of these d-dimensional hyperspheres can be easily computed, and this process can be iterated to find the intersection of K spheres $S(p_i, r_i)$.

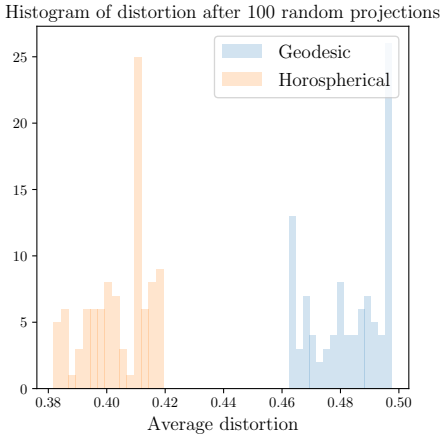


Figure 6: Histogram of average distortion computed for the projections of 1000 random hyperbolic points onto 100 random directions. Best in **bold** and second best <u>underlined</u>.

C.3. Additional Experimental Results

C.3.1. Distortion Analysis

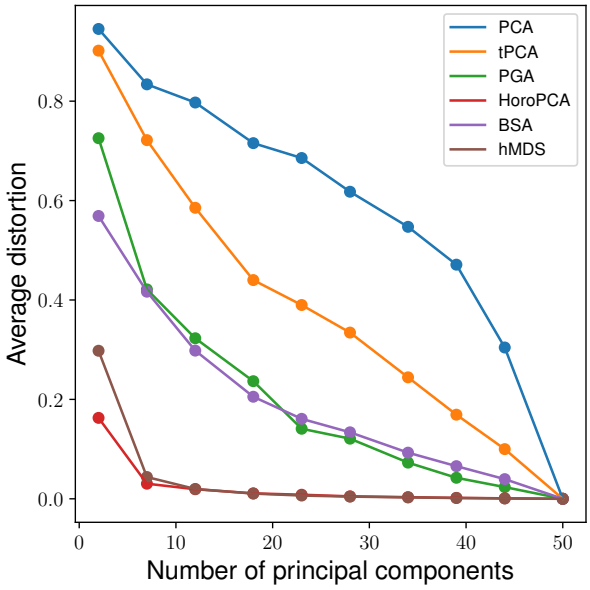
We first analyze the average distortion incurred by horospherical and geodesic projections on a toy example with synthetically-generated data. We generate 1000 points the Poincaré disk by sampling tangent vectors from a multivariate Gaussian, and mapping them to the disk using the exponential map at the origin. We then sample 100 random ideal points (directions) in the disk, and consider the corresponding straight-line geodesics pointing towards these

directions. We project the datapoints onto these geodesics and measure the average distortion from before and after projection and visualize the results in a histogram (Fig. 6). As we can see, Horospherical projections achieve much lower distortion than geodesic projections on average, suggesting that these projections may better preserve information such as distances in higher-dimensional datasets.

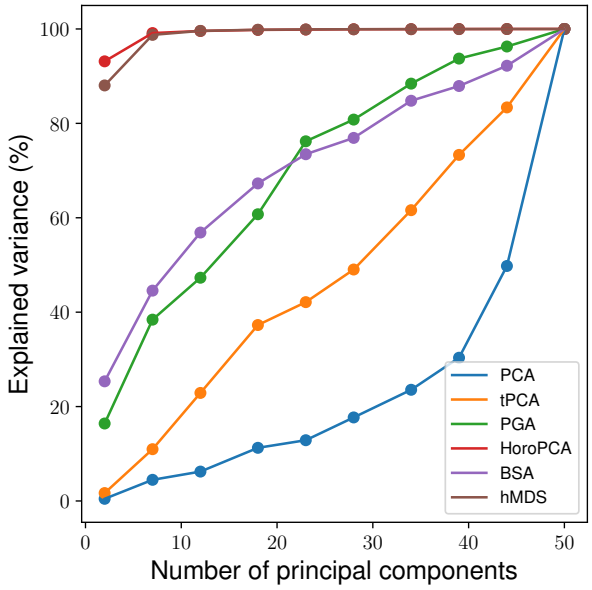
C.3.2. Dimensionality Reduction Results

Sarkar embeddings Table 2 showed dimensionality reduction results for embeddings learned with optimization. Here, we consider the same reduction experiment on combinatorial embeddings and report the results in Table 5. The results confirm the trends observed in Table 2: HoroPCA outperforms baseline methods, with significant improvements on distance preservation.

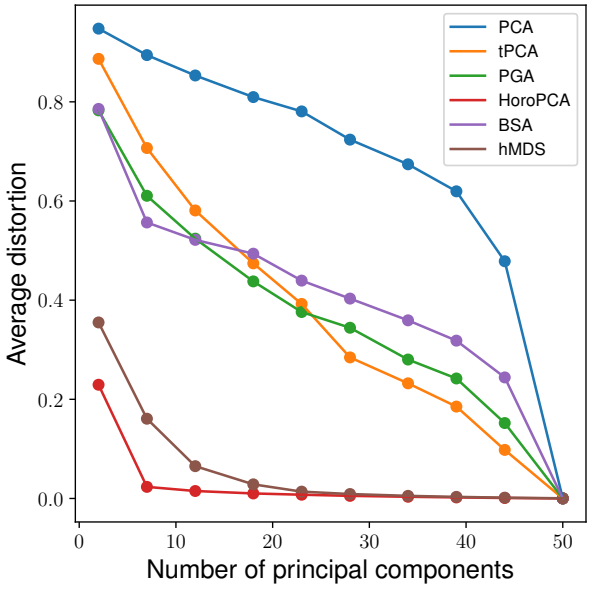
More dimension/component configurations We consider the reduction of 50-dimensional PyTorch embeddings of the Diseases and CS Ph.D. datasets. We plot average distortion and explained Fréchet variance for different number of components in Fig. 7. HoroPCA significantly outperforms all previous generalizations of PCA. HoroPCA also outperforms hMDS, which is a competitive baseline, but not a PCA method as discussed before.



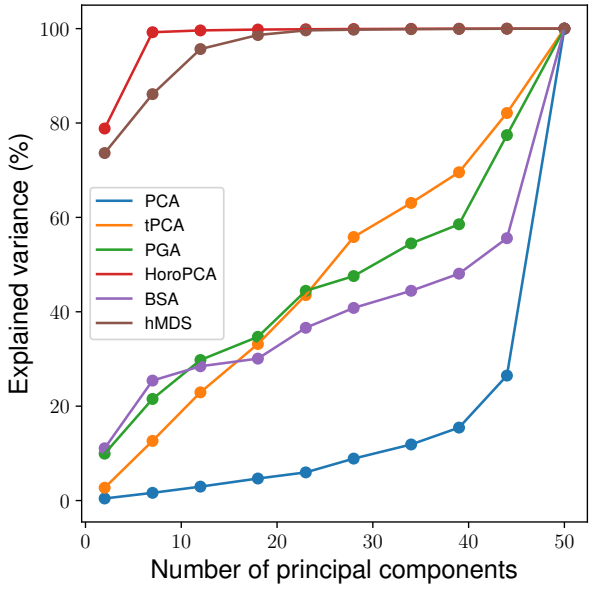
(a) Diseases average distortion.



(b) Diseases explained Fréchet variance.

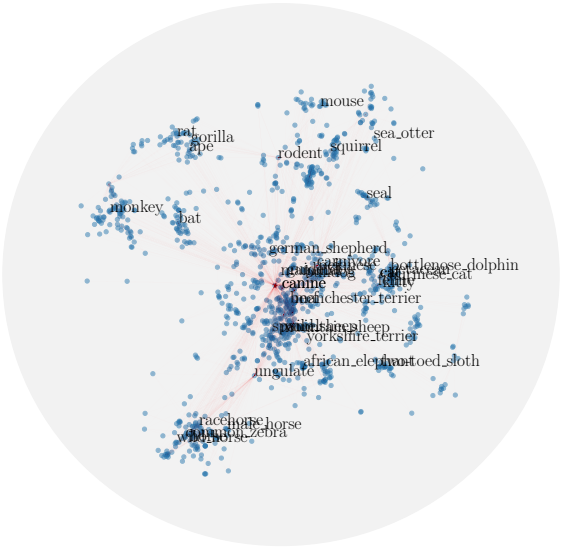


(c) CS Ph.D. average distortion.

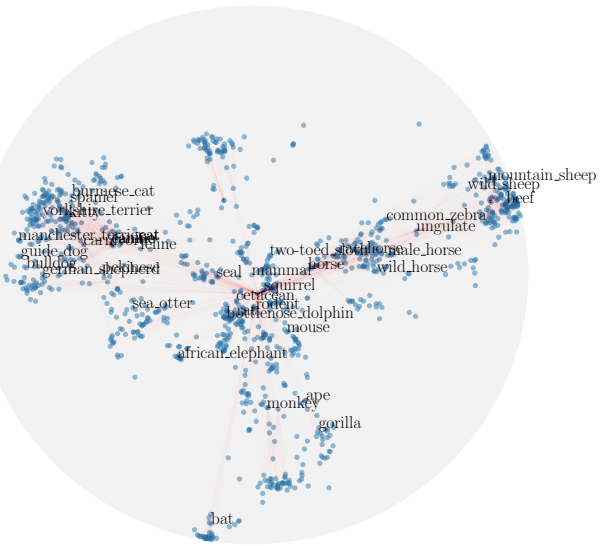


(d) CS Ph.D. explained Fréchet variance.

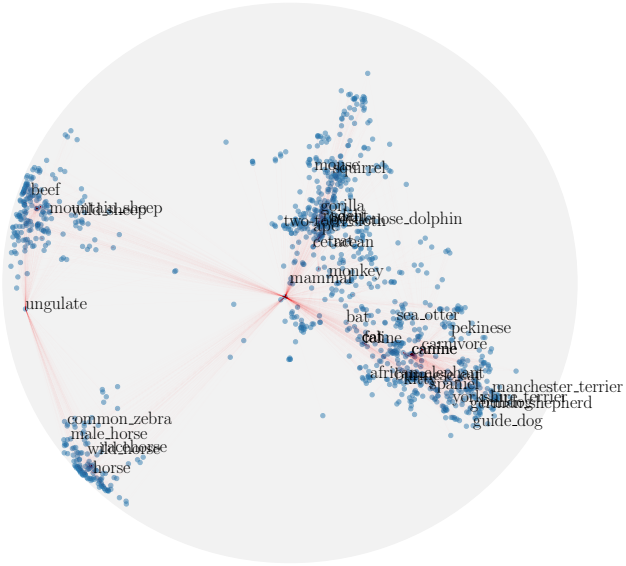
Figure 7: Average distortion and explained Fréchet variance for different numbers of principal components.



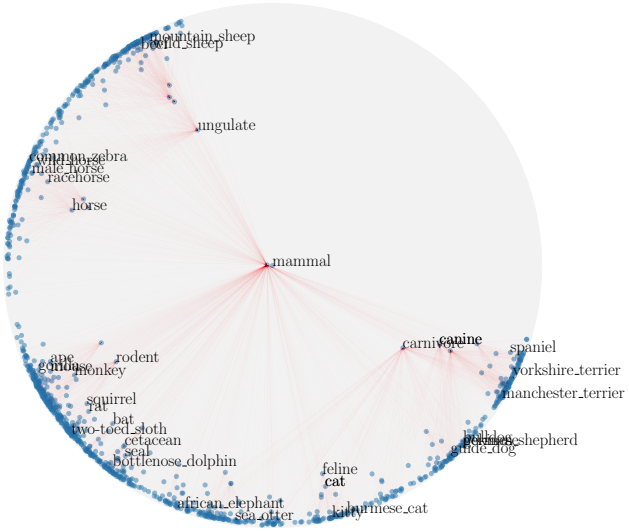
(a) PCA (average distortion: 0.942)



(b) PGA (average distortion: 0.534)



(c) BSA (average distortion: 0.532)



(d) HoroPCA (average distortion: 0.078)

Figure 8: Visualization of embeddings of the WordNet mammal subtree computed by reducing 10-dimensional Poincaré embeddings (Nickel & Kiela, 2017).