Supplementary Materials

The purpose of this supplement is to provide a more detailed background and systematic development of the hPCA visualization method from section 2.5 in the main text. In keeping with the assumed linear algebra background of the main text, the materials in this supplement are focused on understanding the fundamental concepts needed to apply Riemannian geometry, rather than rigorous derivations of the relevant equations.

Section S1 qualitatively describes how overlapping local coordinate systems can be combined in order to navigate on curved spaces. Section S2 provides a brief overview of relevant concepts from differential geometry. Section S3 provides an overview of hyperbolic geometry, including models of hyperbolic space and the ambient spaces they are embedded within. Section S4 develops the machinery from Riemannian geometry needed to work in hyperbolic spaces.

Notation in this section is standard for differential geometry. Points on a manifold will be denoted with italics as p, q with coordinate representations $\boldsymbol{p}$, $\boldsymbol{q}$. Non-Euclidean functions will be distinguished from their Euclidean equivalents using subscripts (e.g., $\langle \cdot, \cdot \rangle_x$ and $||\cdot||_x$).

S1. Local PCAs and Local Coordinate Charts

A series of local coordinate systems that overlap at well-defined points, in combination with smooth mapping functions between respective coordinates, are referred to as *charts*. Using their shared points, overlapping charts can be glued together and projected onto an underlying curved surface. For example, in the field of

cartography, overlapping flat maps are aligned and projected onto a sphere, creating a globe that displays the entire surface of the earth. Notably, in this example, while individual flat maps are in terms of Euclidean coordinates, the curved surface of the earth is non-Euclidean.

In the current work, hPCA algorithms are a series of local PCAs with a local Euclidean coordinate system constructed at every level. Furthermore, since the output of the local PCA from a lower level is input to a subsequent local PCA at a higher level, these coordinate systems overlap at specific points. Within each local coordinate system, distances and displacement vectors between points are calculated as in Appendix A. These distances are mapped to an underlying curved manifold, such as the hyperbolic space $\mathbb{H}^n$, using the tools of differential geometry.

S2. Differential Geometry

Precise measurements on curved manifolds requires concepts and methods from differential geometry. Concepts relevant to the current work are highlighted below. Specific examples involving hyperbolic geometry will be provided in section S3. For an in-depth introduction to differential geometry, specifically for definitions without requiring an ambient space, we refer the reader to the texts by Lee [22,23] in the references.

A *manifold* $\mathcal{M}$ of dimension *n* is a set that can be locally approximated by the Euclidean space $\mathbb{R}^n$. For example, the surface of a sphere embedded in $\mathbb{R}^3$ can be locally approximated by the flat plane $\mathbb{R}^2$ and is therefore a 2-dimensional manifold. In this example, $\mathbb{R}^3$ is the ambient space in which the manifold is embedded.

The *tangent space* $T_p\mathcal{M}$ at point p from $\mathcal{M}$ is the (n+1)-dimensional vector space approximating $\mathcal{M}$ around p (Figure S1, left). Using the inner product from the ambient space, $T_p\mathcal{M}$ can be defined using an orthogonality condition. For instance, if the ambient space is Euclidean, vector $\boldsymbol{u}$ is an element of $T_p\mathcal{M}$ if $\langle \boldsymbol{u}, \boldsymbol{p} \rangle = 0$ where $\boldsymbol{p}$ is the coordinate representation of p. Note that $T_p\mathcal{M} \neq T_q\mathcal{M}$ if $p \neq q$.

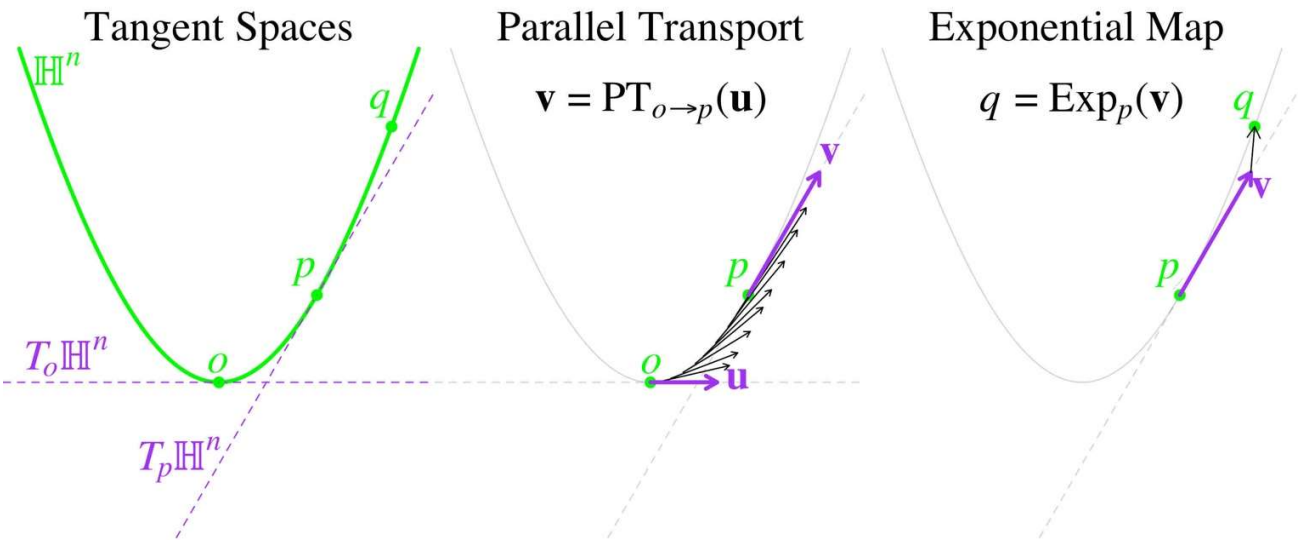


Figure S1: Riemannian geometry on a hyperbolic manifold

A *Riemannian metric* $g_p : T_p\mathcal{M} \times T_p\mathcal{M} \rightarrow \mathbb{R}$ is a collection of inner products on each $T_p\mathcal{M}$ that vary smoothly depending on point p. It defines the local geometry of $\mathcal{M}$ at each point and is used in calculating the lengths of curves on the manifold and distances between points on $\mathcal{M}$. If $\boldsymbol{G} = g_p$ is the matrix form of the Riemannian metric at point p, the inner product $\langle \cdot, \cdot \rangle_p$ and norm $||\cdot||_p$ are calculated as $\langle \boldsymbol{u}, \boldsymbol{v} \rangle_p = \boldsymbol{u}\boldsymbol{G}\boldsymbol{v}^T$ and

$||\boldsymbol{u}||_p = \langle \boldsymbol{u},\boldsymbol{u} \rangle_p^{1/2}$. If the curvature of $\mathcal{M}$ is constant at all points, such as on a sphere or hyperboloid, the matrix form of g_p is constant.

The ordered pair ($\mathcal{M}$, g_p) defines a *Riemannian manifold*. In the following sections, only Riemannian manifolds will be considered due to their metric properties.

A *geodesic* $\gamma : [0,1] \rightarrow \mathcal{M}$ is a generalization of a Euclidean straight line to Riemannian manifolds. The shortest path between points p and q on $\mathcal{M}$ is along a geodesic connecting the points. For example, on the sphere or the surface of the earth, geodesics are great circles passing through a pair of points. Distances between points along geodesics are denoted $\text{dist}_{\mathcal{M}}(\cdot,\cdot)$.

Tangent space vectors are transported along geodesics using the *Parallel Transport* function $\tau_{p \rightarrow q}$: $T_p\mathcal{M} \rightarrow T_q\mathcal{M}$ (Figure S1, center). Parallel transport is a norm-preserving map linking the tangent spaces of $\mathcal{M}$. More precisely, the Riemannian metric between vectors $\boldsymbol{v}$, $\boldsymbol{u}$ from $T_p\mathcal{M}$ is preserved, such that $\langle \boldsymbol{v}, \boldsymbol{u} \rangle_p = \langle \tau_{p \rightarrow q}(\boldsymbol{v}), \tau_{p \rightarrow q}(\boldsymbol{u}) \rangle_q$.

The *Exponential Map* function $\exp_p : T_p\mathcal{M} \rightarrow \mathcal{M}$ maps a small perturbation around p to a point on $\mathcal{M}$ along a geodesic determined by the direction of a tangent vector (Figure S1, right). Let $\gamma : [0,1] \rightarrow \mathcal{M}$ be a geodesic such that $\gamma(0) = p$, $\gamma(1) = q$, and $\gamma'(0) = \boldsymbol{u}$ is a vector from $T_p\mathcal{M}$. The Exponential Map is then defined as $\exp_p(\boldsymbol{u}) = \gamma(1)$. Notably, the distance along γ in $\mathcal{M}$ is preserved by $\exp_p$, such that $||\boldsymbol{u}||_p = \text{dist}_{\mathcal{M}}(\boldsymbol{p}, \boldsymbol{q})$.

The exact of form of the Exponential Map and Parallel Transport functions depend on the manifold and will be given below for hyperbolic manifolds.

S3. Hyperbolic Geometry

Hyperbolic geometry is the study of spaces with constant negative curvature. This is in contrast with Euclidean or spherical geometries with zero or positive curvature, respectively. Intuitively, the branching structure of a tree can be thought of as a discrete realization of a hyperbolic space. In this section and the following, this connection will be developed within a rigorous framework to enable precise mappings from the local coordinates output by hPCA to a latent negatively-curved manifold.

Hyperbolic spaces cannot be isometrically embedded within Euclidean space, leading to fundamental difficulties in understanding and visualizing negatively-curved space. Consequently, many models of hyperbolic geometry have been developed. These include the Poincaré disk $\mathbb{D}$, upper half-plane $\mathbb{U}$, and hyperboloid $\mathbb{H}$, among others. Each model emphasizes different aspects of hyperbolic geometry, but no model completely encapsulates all of its features. For instance, the hyperboloid model facilitates calculations within hyperbolic spaces, but its visualization is challenging. In contrast, the Poincaré disk and upper half-plane models allow for convenient visualization, but more complicated and numerically unstable calculations. Fortunately, all of these models are equivalent in the sense that there are isometric transformations between them. This allows each model to be used in its suitable domain and subsequently mapped to a different domain as needed. In the following, we develop an efficient method for projecting points and vectors within hyperbolic spaces, based on the

hyperboloid model of hyperbolic geometry. This method will be used to map the output of hPCA onto the manifold $\mathbb{H}^n$.

To construct the hyperboloid model of $\mathbb{H}^n$ embedded within an ambient space, we define the Lorentzian inner product $\langle\cdot,\cdot\rangle_H$ and associated norm $||\cdot||_H$ by

$$\langle \boldsymbol{x}, \boldsymbol{y} \rangle_H = -x_0 y_0 + \sum_{i=1}^{n} x_i y_i, \qquad \boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^{n+1},$$
$$\|\boldsymbol{x}\|_H = \sqrt{\langle \boldsymbol{x}, \boldsymbol{x} \rangle_H}, \qquad \|\boldsymbol{x}\|_H \in \mathbb{C}, \tag{S1}$$

where vector indices are counted starting from zero. The space $\mathbb{R}^{n+1}$ equipped with the above inner product is a (n+1)-dimensional Minkowski space. Note that, unlike the Euclidean inner product, the Lorentzian inner product can be negative.

The n-dimensional hyperboloid $\mathbb{H}^n$ is the collection of all points such that,

$$\mathbb{H}^n = \{\boldsymbol{x} \in \mathbb{R}^{n+1} \mid \langle \boldsymbol{x}, \boldsymbol{x} \rangle_H = -1, x_0 > 0\}. \tag{S2}$$

For n = 1, equation (S2) describes a standard hyperbola resulting from conic sections. That is, $\mathbb{H}^1$ is the intersection of a cone with a plane in $\mathbb{R}^3$, with the branches of the hyperbola curving around the x_0 axis. In higher dimensions, the equation describes a negatively-curved hypersurface centered at the point o = [1,0,…,0], located near the origin of $\mathbb{R}^{n+1}$.

Distances between points x and y on $\mathbb{H}^n$ with coordinates $\boldsymbol{x}$, $\boldsymbol{y}$ within the ambient Minkowski space, are calculated using hyperbolic trigonometry,

$$\mathrm{dist}_H(x, y) = \cosh^{-1}(\langle \boldsymbol{x}, \boldsymbol{y} \rangle_H), \qquad x, y \in \mathbb{H}^n. \tag{S3}$$

The shortest path between any pair of points on $\mathbb{H}^n$ is along a geodesic, with a total length calculated by equation (S3).

S4. Operations on Hyperbolic Manifolds

The tools of differential geometry outlined in section S2 can be applied to hyperbolic spaces. Since $\mathbb{H}^n$ is a locally-smooth embedding within $\mathbb{R}^{n+1}$ at all points, it is a manifold. The Riemannian metric on $\mathbb{H}^n$ is the diagonal matrix g_H = diag(-1,1,…,1). Consequently, ($\mathbb{H}^n$, g_H) is a Riemannian manifold and the tools of differential geometry can be applied.

The tangent space of $\mathbb{H}^n$ at point p with coordinate representation $\boldsymbol{p}$ is defined as,

$$T_p\mathbb{H}^n = \{\boldsymbol{u} \in \mathbb{R}^{n+1} \mid \langle \boldsymbol{u}, \boldsymbol{p} \rangle_H = 0\}. \tag{S4}$$

Notably, vectors in $T_p\mathbb{H}^n$ are not orthogonal to $\boldsymbol{p}$ in the Euclidean sense (e.g., $\langle \boldsymbol{u},\boldsymbol{p} \rangle = 0$), but instead follow the closely-related definition in equation (S4). Accordingly, the tangent space $T_p\mathbb{H}^n$ can be described as all vectors from the ambient space $\mathbb{R}^{n+1}$ that are hyperbolically-orthogonal to p. Each tangent space inherits the Lorentzian inner product (S1) from the ambient space. Furthermore, this inner product is positive-definite when restricted to vectors from a tangent space.

Vectors within a subspace of $\mathbb{R}^n$ can be easily embedded into $T_o\mathbb{H}^n$, the tangent space at the central point of the hyperbola o = [1,0,…,0]. Let $\boldsymbol{u'}$ be a subspace vector, formatted as a row vector to simplify notation. It is embedded within $\mathbb{R}^{n+1}$ using the inclusion function ι_o: $\mathbb{R}^k \rightarrow T_o\mathbb{H}^n$ defined by,

$$\iota_o(\boldsymbol{u}') = [0, \boldsymbol{u}', \boldsymbol{0}], \qquad \boldsymbol{u}' \in \mathbb{R}^k, \tag{S5}$$

where $\boldsymbol{0}$ is a vector of $n - k$ zeros. Note that this embedding preserves lengths within the ambient space. Specifically, if $\boldsymbol{u} = \iota_o(\boldsymbol{u'})$, then $||\boldsymbol{u'}|| = ||\boldsymbol{u}|| = ||\boldsymbol{u}||_H$.

The Parallel Transfer function $\tau_{p\rightarrow q}$: $T_p\mathbb{H}^n \rightarrow T_q\mathbb{H}^n$ is given by,

$$\tau_{p\to q}(\boldsymbol{u}) = \boldsymbol{u} + \frac{\langle \boldsymbol{q}, \boldsymbol{u} \rangle_H}{1 - \langle \boldsymbol{p}, \boldsymbol{q} \rangle_H}(\boldsymbol{p} + \boldsymbol{q}), \qquad \boldsymbol{u} \in T_p\mathbb{H}^n. \tag{S6}$$

Since Parallel Transfer preserves lengths, $||\boldsymbol{u}||_H = ||\tau_{p\to q}(\boldsymbol{u})||_H$. Additionally, equation (S6) gives a relatively simple method for calculating the coordinates of vectors in the tangent space of arbitrary points on $\mathbb{H}^n$. Specifically, the vector is first embedded into the tangent space $T_o\mathbb{H}^n$ using (S5), then parallel transported to the tangent space of any point on the manifold using (S6).

Lastly, the Exponential Map used to map $T_p\mathbb{H}^n$ to a point on $\mathbb{H}^n$ is given by

$$\exp_p(\boldsymbol{u}) = \cosh(\|\boldsymbol{u}\|_H)\,\boldsymbol{p} + \sinh(\|\boldsymbol{u}\|_H)\frac{\boldsymbol{u}}{\|\boldsymbol{u}\|_H}, \qquad \boldsymbol{u} \in T_p\mathbb{H}^n. \tag{S7}$$

Since the Exponential Map preserves lengths, $||\boldsymbol{u}||_H = \text{dist}_H(\exp_p(\boldsymbol{u}), \boldsymbol{p})$.

The above equations, in combination, facilitate calculations on $\mathbb{H}^n$. A vector from a locally Euclidean coordinate system is embedded into the tangent space at the center *o* using (S5). Subsequently, the embedded vector is then transported to the tangent space of an arbitrary point using (S6), preserving its hyperbolic length. Finally, the vector is projected onto the manifold using (S7), again preserving its hyperbolic length. The length of the original vector is preserved at every step in this process. Furthermore, since the Riemannian metric is unchanged, all steps locally preserve angles. Consequently, angles between variables and data points in the same tangent space are preserved by the embedding and projection into $\mathbb{H}^n$.

In the main text, the above functions were used to construct the composite function $\varphi_{i,p}$: SO(k) x $\mathbb{R}^k$ → $\mathbb{H}^k$,

$$\varphi_{i,p} = \exp_p \circ \tau_{o\to p} \circ \iota_o \circ \Delta_i, \tag{S8}$$

where Δ_i: SO(k) x $\mathbb{R}^k \rightarrow \mathbb{R}^k$ is the displacement function displacement coordinate function detailed in Appendix A of the main text. This composite function was used in Algorithm 2 to embed the matrix of eigenvectors from SO(k) and vector of eigenvalues from $\mathbb{R}^k$ from a local PCA onto the hyperboloid $\mathbb{H}^k$ and subsequently projected onto the Poincaré disk $\mathbb{D}$.

References:

1. Lee, J. M. (2006). *Riemannian manifolds: an introduction to curvature*. Springer.
2. Lee, J. M. (2012). *Introduction to Smooth Manifolds*. Springer.