principal component analysis in space forms

puoya tabaghi

may 31st, 2022

UC San Diego
JACOBS SCHOOL OF ENGINEERING
Electrical and Computer Engineering

overview

one-minute pitch

- *research goal: review affine subspaces on riemannian manifolds and derive close-form "optimal" solutions for pcas in spherical and hyperbolic spaces*

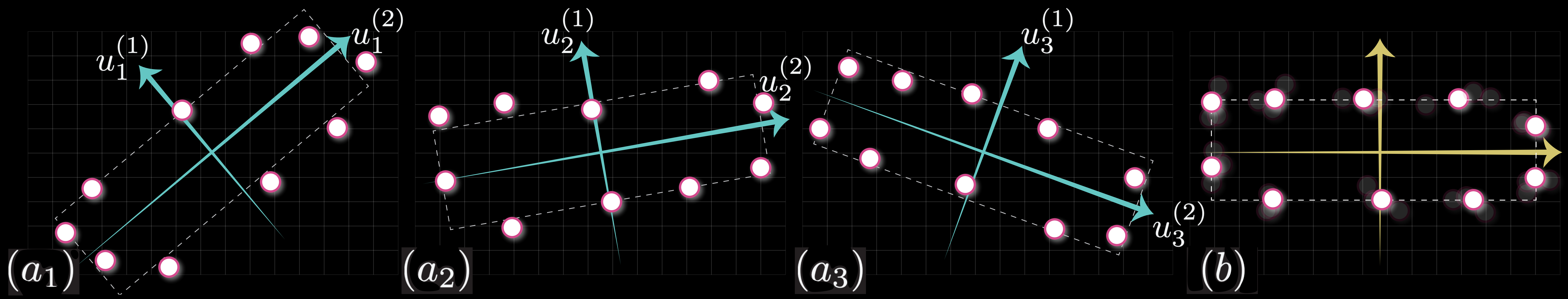
- *current work proposes an eigenequation approach to pcas in space forms (euclidean, spherical, and hyperbolic spaces)*

- *summary: we show what it means to have an (generalized) affine subspace in curved spaces, derive projection operators in spherical and hyperbolic spaces, compute the distance of a points to an affine subspace, and propose a distortion function to determine the best affine subspace via solving an appropriate eigenequation*

p. tabaghi, y. wang, and s. mirarab —- *an ongoing work*, 2022.

multiple co-inertia analysis

our motivation

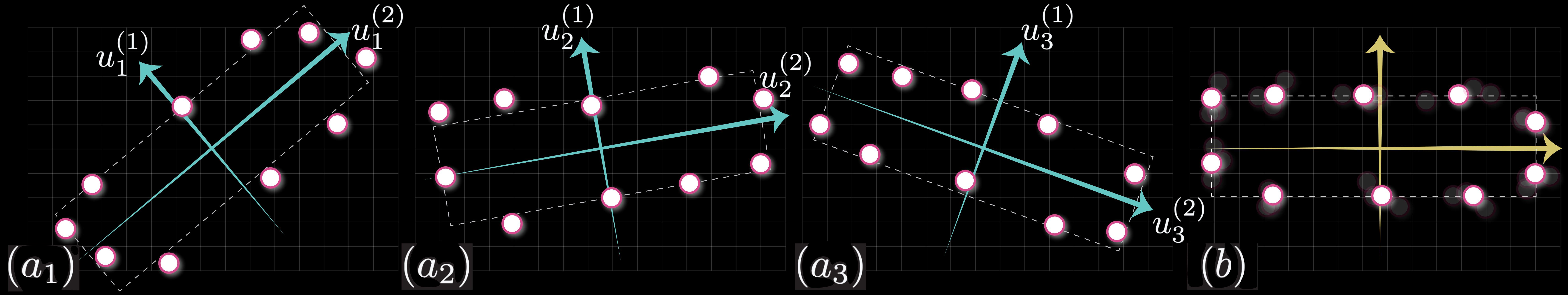


- $(a_1 - a_3)$ *find the co-inertia loadings* $u_k^{(1)}$ *and* $u_k^{(2)}$ *— where* $k \in [3]$ *— for three datasets*
- (b) *the estimated reference dataset*

tabaghi *et al.*, 2022.

multiple co-inertia analysis

our motivation



- solution is an eigenequation
- problem: align embedded trees in hyperbolic spaces

a simple way to estimate principal axis in hyperbolic spaces

tabaghi *et al.*, 2022.

principal component analysis

the minimum distortion affine subspace

- an affine map to model a set of N data points $x_1, \ldots, x_N \in \mathbb{R}^D$, i.e.,

$$\forall n \in [N] : x_n = Hy_n + p + \nu_n$$

- the columns of $H \in \mathbb{R}^{D \times d}$ $(D \gg d)$ forms a low-dimensional subspace, the feature vector $y_n \in \mathbb{R}^d$ corresponds to x_n, $p \in \mathbb{R}^D$ is the bias, and $\nu_n \in \mathbb{R}^D$ is the mismatch term

- pca assumes the following measure of distortion for an affine subspace:

$$\mathrm{dist.}(p + H \,|\, \mathscr{X}) = \sum_{n \in [N]} d\big(x_n, \mathscr{P}_{H,p}(x_n)\big)^2 : \quad \mathscr{P}_{H,p}(x_n) = \mathrm{argmin}_{x \in p+H} \, d(x, x_n),$$

where $\mathscr{X} = \{x_n\}_{n \in [N]}$, and $d(\,\cdot\,,\cdot\,)$ computes the euclidean distance

tabaghi *et al.*, 2022.

principal component analysis

the minimum distortion affine subspace

- an affine map to model a set of N data points $x_1, \ldots, x_N \in \mathbb{R}^D$, i.e.,

$$\forall n \in [N] : x_n = Hy_n + p + \nu_n$$

- the columns of $H \in \mathbb{R}^{D \times d}$ ($D \gg d$) forms a low-dimensional subspace, the feature vector $y_n \in \mathbb{R}^d$ corresponds to x_n, $p \in \mathbb{R}^D$ is the bias, and $\nu_n \in \mathbb{R}^D$ is the mismatch term

- pca assumes the following measure of distortion for an affine subspace:

$$\mathrm{dist.}(p + H \,|\, \mathscr{X}) = \sum_{n \in [N]} d\big(x_n, \mathscr{P}_{H,p}(x_n)\big)^2 : \quad \mathscr{P}_{H,p}(x_n) = \mathrm{argmin}_{x \in p+H}\ d(x, x_n),$$

where $\mathscr{X} = \{x_n\}_{n \in [N]}$, and $d(\,\cdot\,,\cdot\,)$ computes the euclidean distance

tabaghi *et al.*, 2022.

affine subspace — revisited

line-point formulation

- **definition:** the set $V \subseteq \mathbb{R}^D$ is an affine subspace if and only if

$$V = \{x \in \mathbb{R}^D : \langle x - p, h' \rangle = 0, \text{ for all } \; h' \in H^\perp\}$$

for a vector $p \in \mathbb{R}^D$ and a subspace H of $\mathbb{R}^D$

tabaghi *et al.*, 2022.

affine subspace — revisited

line-point formulation

- **definition:** $p + H = \{x \in \mathbb{R}^D : \langle x - p, h' \rangle = 0,$ for all $h' \in H^\perp\}$
- consider the following two parametric lines:

$$\gamma_{p,x}(t) = (1-t)p + tx, \text{ and } \gamma_{h'}(t) = p + th',$$

where $h' \in H^\perp$

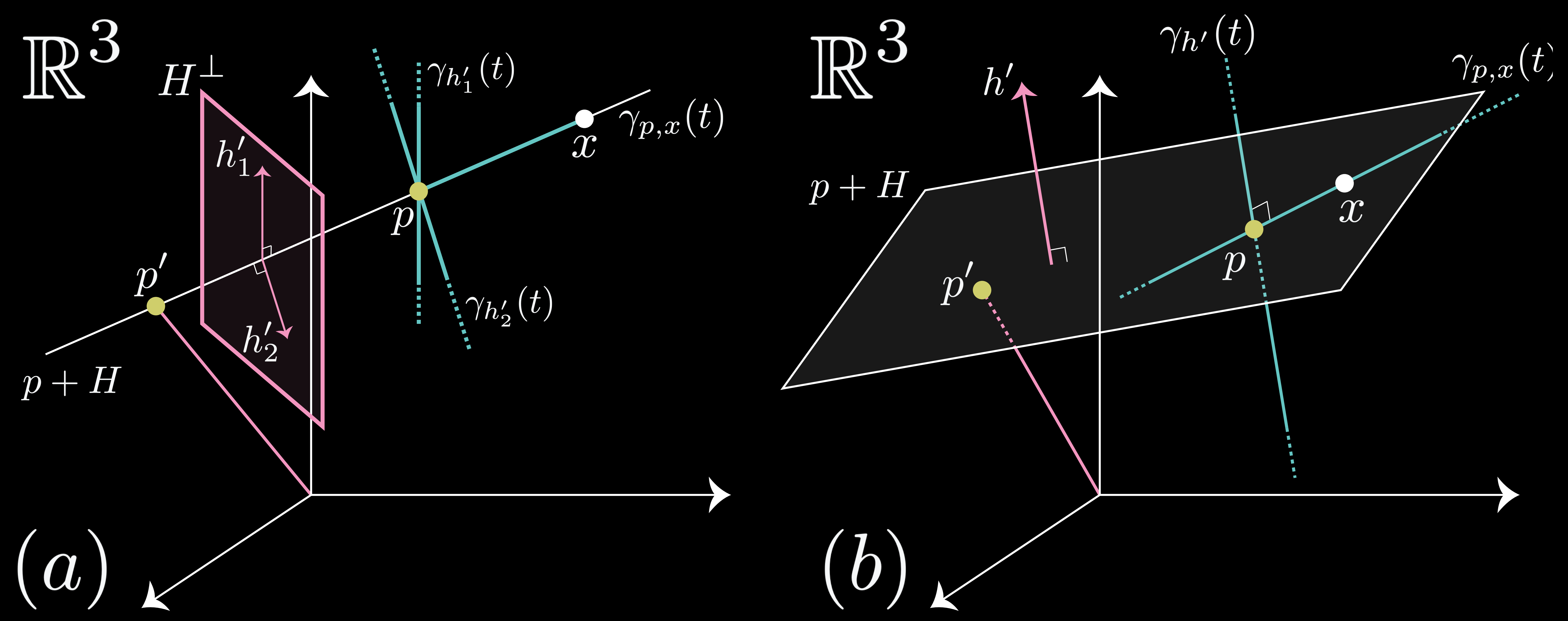
- **definition:** for a vector $p \in \mathbb{R}^D$ and a subspace H of $\mathbb{R}^D$, we have:

$$p + H = \{x \in \mathbb{R}^D : \langle \frac{d}{dt}\gamma_{p,x}(t)\,|_{t=0}, \frac{d}{dt}\gamma_{h'}(t)\,|_{t=0} \rangle = 0, \text{ for all } h' \in H^\perp\}$$

tabaghi *et al.*, 2022.

affine subspace — revisited

visualization



- (a, b) *one- and two-dimensional affine subspaces in a three-dimensional euclidean space*

tabaghi *et al.*, 2022.

generalized affine subspace

definition I

- **definition:** $p + H = \{x \in \mathbb{R}^D : \langle \frac{d}{dt}\gamma_{p,x}(t)\,|_{t=0}, \frac{d}{dt}\gamma_{h'}(t)\,|_{t=0} \rangle = 0, \text{ for all } h' \in H^{\perp}\}$

tabaghi *et al.*, 2022.

generalized affine subspace

definition

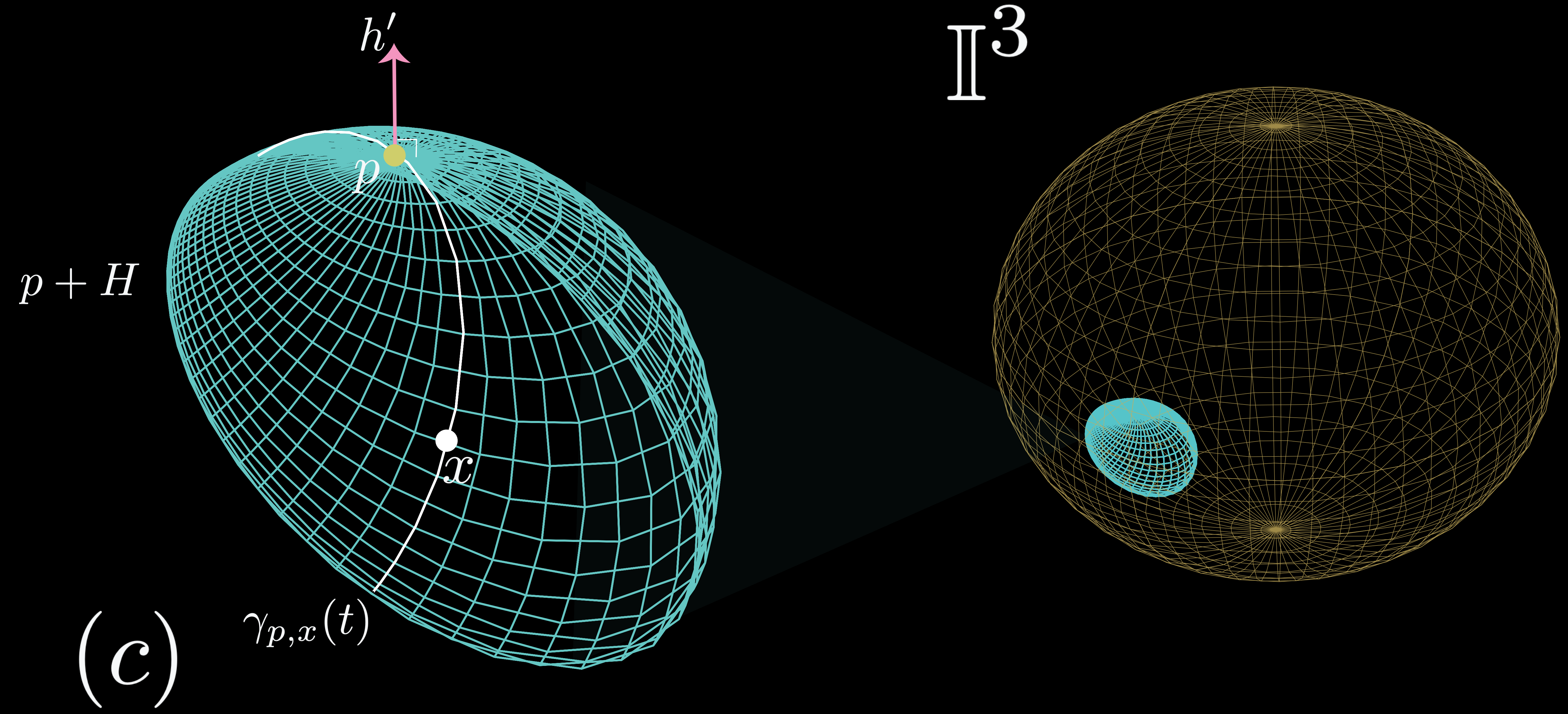
- **definition:** $p + H = \{x \in \mathbb{R}^D : \langle \frac{d}{dt}\gamma_{p,x}(t)\,|_{t=0}, h' \rangle = 0,$ for all $h' \in H^{\perp}\}$
- in manifold $\mathcal{M}$, $g_p(v, u)$ computes the inner products of $v, u \in T_p\mathcal{M}$ — the tangent space at p
- $\log_p(x)$ computes the initial velocity (in $T_p\mathcal{M}$) to move from p to x in one time-step
- **definition:** let $(\mathcal{M}, g)$ be a geodesically complete riemannian manifold and $p \in \mathcal{M}$, and H be a subspace of $T_p\mathcal{M}$, then we have

$$\mathcal{M}_H = \{x \in \mathcal{M} : g_p(\log_p(x), h') = 0, \forall h' \in H^{\perp} \subseteq T_p\mathcal{M}\}$$

tabaghi *et al.*, 2022.

generalized affine subspace

visualization



- (c) *two-dimensional affine subspace in a three-dimensional hyperbolic space (poincare model)*

Tabaghi *et al.*, 2022.

generalized affine subspace

definitions i, ii, iii

- **definition i:** let $(\mathscr{M}, g)$ be a geodesically complete riemannian manifold and $p \in \mathscr{M}$, and H be a subspace of $T_p\mathscr{M}$, then we have

$$\mathscr{M}_H = \{x \in \mathscr{M} : g_p(\log_p(x), h') = 0, \forall h' \in H^\perp \subseteq T_p\mathscr{M}\}$$

tabaghi *et al.*, 2022.

generalized affine subspace

definitions i, ii, iii, iv

- **definition i:** $\mathscr{M}_H = \{x \in \mathscr{M} : g_p(\log_p(x), h') = 0, \ \forall h' \in H^\perp \subseteq T_p\mathscr{M}\}$
- **definition ii:** $\mathscr{M}_H = \{x \in \mathscr{M} : \log_p(x) \in H\}$
- **definition iii:** $\mathscr{M}_H = \exp_p(H) :$ H is a subspace of $T_p\mathscr{M}$ (fletcher et al.)
- **definition iv:** there is a point $p \in \mathscr{M}_H$ such that
  - for any $v \in \mathscr{M}_H$, we have $\exp_p\left(\alpha \log_p(v)\right) \in \mathscr{M}_H$ where $\alpha \in \mathbb{R}$
  - for any $v_1, v_2 \in V$, we have $\exp_p\left(\log_p(v_1) + \log_p(v_2)\right) \in \mathscr{M}_H$

tabaghi *et al.*, 2022.

spherical space

subspace definition

- *the* d*-dimensional spherical space is a riemannian manifold* $(\mathbb{S}^d, g^{\mathbb{S}})$*, where*

$$\mathbb{S}^d = \{x \in \mathbb{R}^{d+1} : \langle x, x \rangle = 1\}, \text{ and } g_p^{\mathbb{S}}(u, v) = \langle u, v \rangle$$

where $u, v \in T_p\mathbb{S}^d = p^\perp$

- $\forall x, y \in \mathbb{S}^d : d(x, y) = \mathrm{acos}(\langle x, y \rangle)$
- $\mathbb{S}_H^d = \{x \in \mathbb{S}^d : \langle x, h' \rangle = 0, \forall h' \in H^\perp\} = \mathbb{S}^d \cap H$

tabaghi *et al.*, 2022.

spherical subspace

projection operator

- ***proposition:*** *the projection of* $x \in \mathbb{S}^d$ *onto* $\mathbb{S}^d_H$ *is given as follows*

$$\mathscr{P}_H(x) = \frac{1}{\|P_H(x)\|_2} P_H(x) \ \ : \ \ P_H(x) = \langle x, p \rangle p + \sum_{k \in [K]} \langle x, h_k \rangle h_k,$$

where $h_1, \ldots, h_K$ *are a complete set of orthonormal basis vectors for* H, *and*

$$d(x, \mathscr{P}_H(x)) = \operatorname{acos}\Big(\|P_H(x)\|_2 \Big) = \operatorname{acos}\left(\sqrt{\langle x, p \rangle^2 + \sum_{k \in [K]} \langle x, h_k \rangle^2} \right)$$

tabaghi *et al.*, 2022.

spherical subspace

isometry & low-dimensional representation

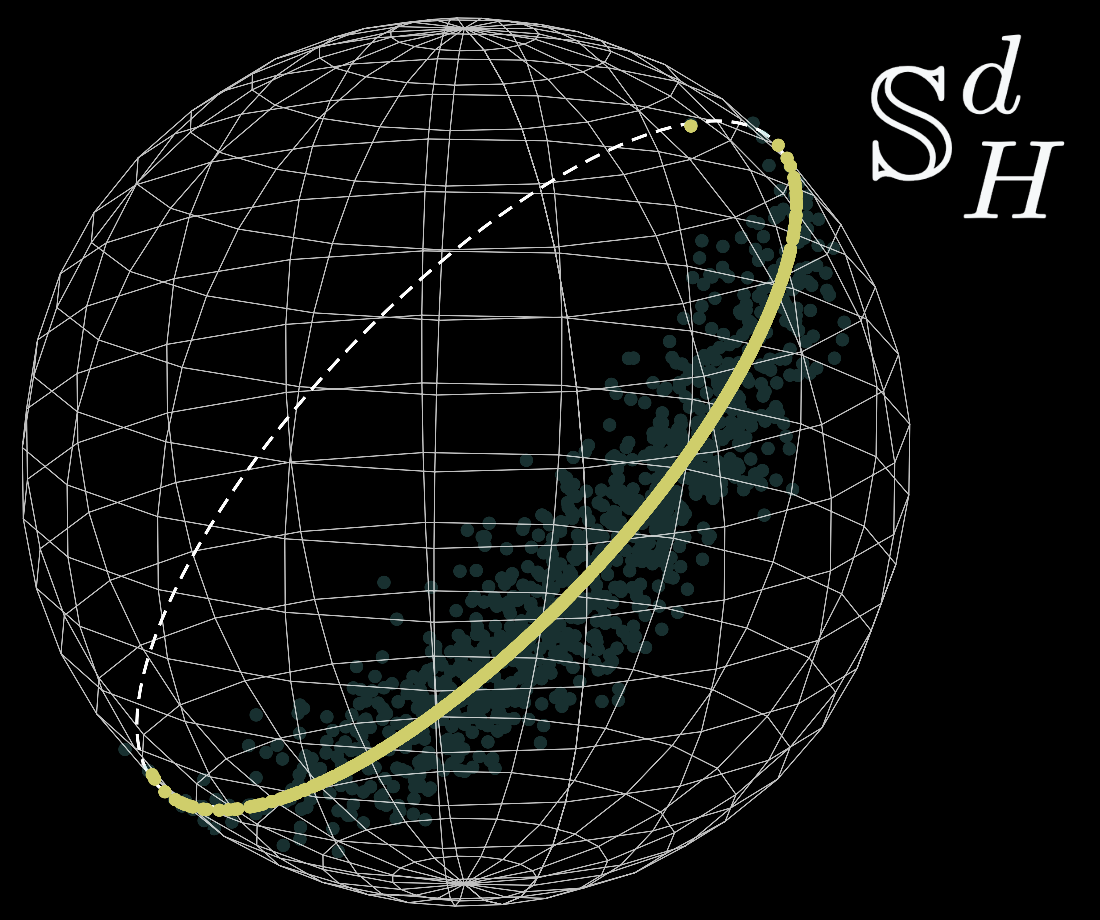
- ***theorem:*** $\mathbb{S}_H^d$ *and* $\mathbb{S}^K$ *are isometric, where* $K = \dim(H)$

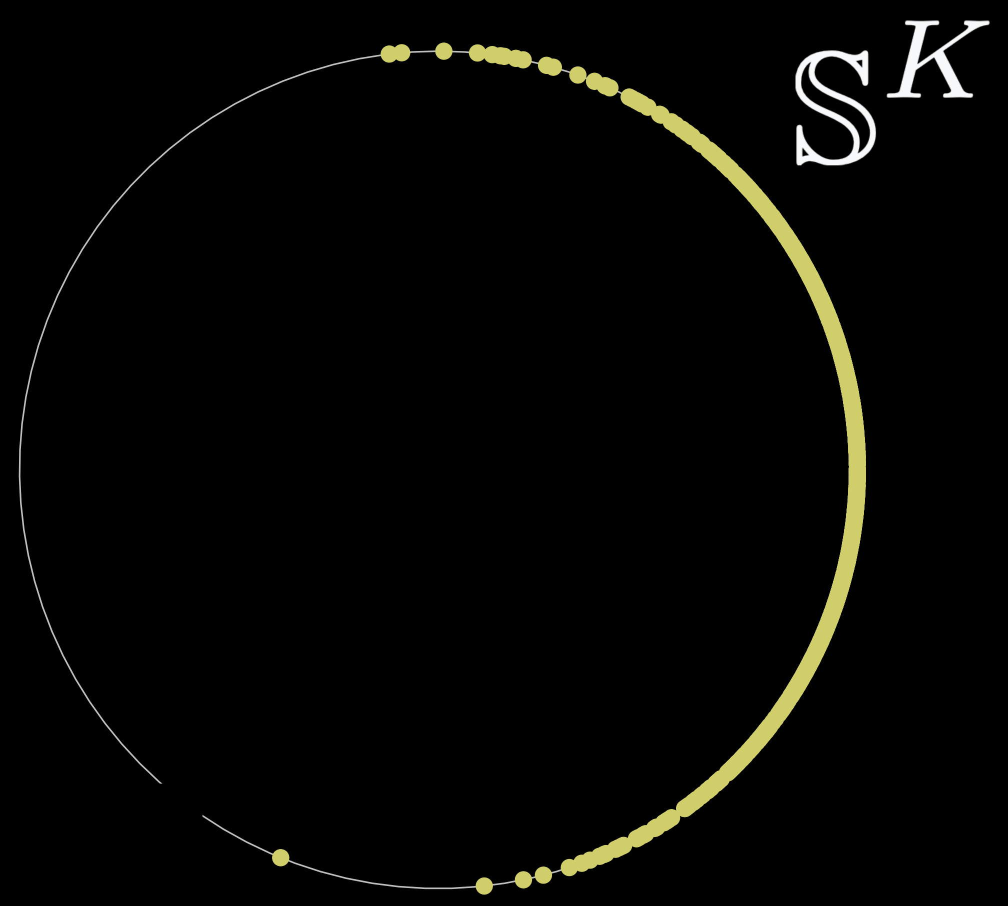
tabaghi *et al.*, 2022.

spherical subspace

isometry & low-dimensional representation

- ***theorem:*** $\mathbb{S}^d_H$ *and* $\mathbb{S}^K$ *are isometric*





- *the isometry* $\mathcal{Q} : \mathbb{S}^d_H \to \mathbb{S}^K$ *and its inverse are given as follows*

$$\mathcal{Q}(x) = \begin{bmatrix} \langle x, p \rangle \\ \langle x, h_1 \rangle \\ \vdots \\ \langle x, h_K \rangle \end{bmatrix} \in \mathbb{S}^K, \quad \mathcal{Q}^{-1}\left(\begin{bmatrix} y_0 \\ y_1 \\ \vdots \\ y_K \end{bmatrix} \right) = y_0 p + \sum_{k=1}^{K} y_k h_k \in \mathbb{S}^d_H,$$

tabaghi *et al.*, 2022.

spherical pca

definition & solution

- **definition:** for an affine subspace $\mathbb{S}^d_H$, we define the following measure of distortion:

$$\mathrm{dist}.(\mathbb{S}^d_H \mid \mathcal{X}) = -N^{-1} \sum_{n \in [N]} \cos^2\big(d(x_n, \mathcal{P}_H(x_n))\big)$$

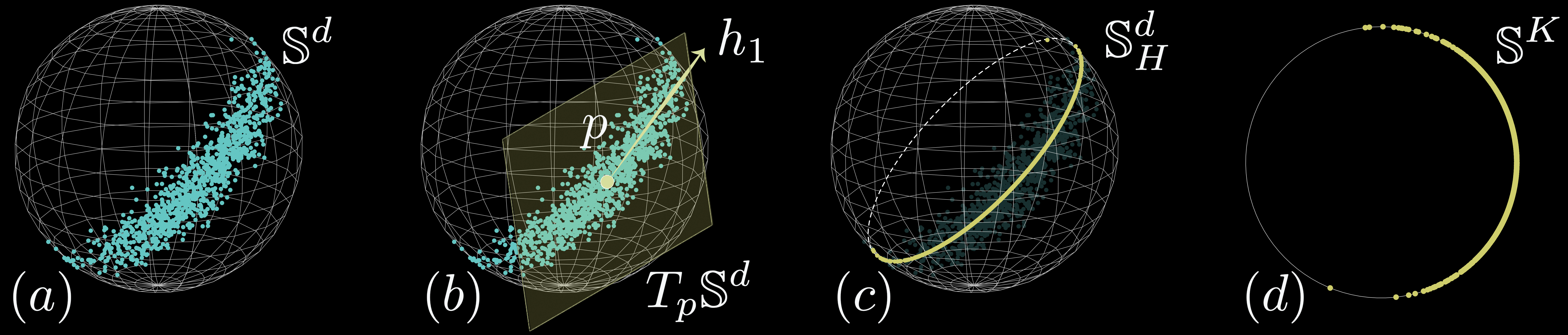
- let $x_1, \ldots, x_N \in \mathbb{S}^d$ and $C_x = N^{-1} \sum_{n \in [N]} x_n x_n^\top$ be a diagonalizable matrix

- **theorem:** p is the leading eigenvector of C_x and H is spanned by its remaining eigenvectors

tabaghi *et al.*, 2022.

spherical pca

summary



- (a) a set data points in $\mathbb{S}^2$
- (b) the best estimate for the base point p and the tangent vector $h_1 \in T_p\mathbb{S}^2$; these two define a spherical affine subspace $\mathbb{S}^2_H$ $(H = h_1)$
- (c) the projection of each point onto the affine subspace $\mathbb{S}^2_H$
- (d) the corresponding low-dimensional features in $\mathbb{S}^K$, where $K = \mathrm{affdim}(S^2_H) = 1$.

tabaghi *et al.*, 2022.

hyperbolic space

subspace definition

- *lorentzian inner product: for* $x, y \in \mathbb{R}^{d+1}$*, let* $[x, y] = x^\top J_d y$ *where* $J_d = \mathrm{diag}\left(-1,1,\ldots,1\right)$
- *the* d*-dimensional 'loid model is a riemannian manifold* $(\mathbb{L}^d, g^{\mathbb{L}})$*, where*

$$\mathbb{L}^d = \{x \in \mathbb{R}^{d+1} : [x, x] = -1,\ x_0 > 0\},\ \textit{and } g_p^{\mathbb{L}}(u, v) = [u, v]$$

where $u, v \in T_p\mathbb{L}^d = p^\perp$

- $\forall x, y \in \mathbb{L}^d : d(x, y) = \mathrm{acosh}(-[x, y])$
- $\mathbb{L}_H^d = \{x \in \mathbb{L}^d : \langle x, h' \rangle = 0,\ \forall h' \in H^\perp\} = \mathbb{L}^d \cap H$

tabaghi *et al.*, 2022.

hyperbolic subspace

projection operator

- ***proposition:*** *the projection of* $x \in \mathbb{L}^d$ *onto* $\mathbb{L}^d_H$ *is given as follows*

$$\mathscr{P}_H(x) = \frac{1}{\sqrt{-\|P_H(x)\|^2}} P_H(x) \quad : \quad P_H(x) = -[x, p]p + \sum_{k \in [K]} [x, h_k] h_k$$

where $h_1, \ldots, h_K$ *are a complete set of orthonormal basis vectors for* H*, and*

$$d(x, \mathscr{P}_H(x)) = \operatorname{acosh}\left(\sqrt{-\|P_H(x)\|^2}\right) = \operatorname{acosh}\left(\sqrt{[x, p]^2 - \sum_{k \in [K]} [x, h_k]^2}\right)$$

tabaghi *et al.*, 2022.

hyperbolic subspace

isometry & low-dimensional representation

- ***theorem:*** $\mathbb{L}^d_H$ *and* $\mathbb{L}^K$ *are isometric, where* $K = \dim(H)$
- *the isometry* $\mathcal{Q} : \mathbb{L}^d_H \to \mathbb{L}^K$ *and its inverse are given as follows*

$$\mathcal{Q}(x) = \begin{bmatrix} -[x, p] \\ [x, h_1] \\ \vdots \\ [x, h_K] \end{bmatrix} \in \mathbb{L}^K, \quad \mathcal{Q}^{-1}\left(\begin{bmatrix} y_0 \\ y_1 \\ \vdots \\ y_K \end{bmatrix} \right) = y_0 p + \sum_{k=1}^{K} y_k h_k \in \mathbb{L}^d_H,$$

tabaghi *et al.*, 2022.

hyperbolic pca

definition & problem statement

- **definition:** for an affine subspace $\mathbb{L}_H^d$, we define the following measure of distortion:

$$\text{dist}.(\mathbb{L}_H^d \,|\, \mathscr{X}) = N^{-1} \sum_{n \in [N]} \cosh^2\big(d(x_n, \mathscr{P}_H(x_n))\big)$$

- let $x_1, \ldots, x_N \in \mathbb{L}^d$ and $C_x = N^{-1} \sum_{n \in [N]} x_n x_n^\top$
- **problem:** hyperbolic pca aims to find $p \in \mathbb{L}^d$ and orthonormal vectors $h'_1, \ldots, h'_{K'} \in T_p\mathbb{L}^d$

that minimize $\sum_{k \in [K']} h_k'^\top J_d C_x J_d h'_k$

tabaghi *et al.*, 2022.

lorentzian space

preliminaries

- *the lorentzian $(d+1)$-space is a vector space $\mathbb{R}^{d+1}$ that is equipped with $[\,\cdot\,,\cdot\,]$*
- *example: J_d-adjoint operator*

$$[Ax, y] = [x, A^{[\top]}, y] \rightarrow A^{[\top]} = J_d A^\top J_d$$

- $A^{[-1]}$ *is the J_d-inverse of A if and only if $A^{[-1]} J_d A = A J_d A^{[-1]} = J_d$*
- *An invertible matrix A is called J_d-unitary if and only if $A^\top J_d A = J_d$*
- *the J_d-eigenequation:*

$$AJ_d v = \begin{cases} -\lambda v, & \textit{if } [v^*, v] = 1 \\ \lambda v, & \textit{if } [v^*, v] = -1 \end{cases}$$

tabaghi *et al.*, 2022.

hyperbolic pca

solution

- **problem:** hyperbolic pca aims to find $p \in \mathbb{L}^d$ and orthonormal vectors $h'_1, \ldots, h'_{K'} \in T_p\mathbb{L}^d$

that minimize $\sum_{k \in [K']} {h'_k}^\top J_d C_x J_d h'_k$

- **solution:** p is the negative J_d-eigenvector of C_x and $h'_1, \ldots, h'_{K'}$ are the positive J_d-eigenvectors

that correspond to the smallest K' J_d-eigenvalues of C_x

- **algorithm:** use modified power methods to compute J_d-eigenvectors

tabaghi *et al.*, 2022.

other formulations and applications

spherical and hyperbolic pca

- (liu et al.) spherical pca is the following matrix factorization problem:

$$\min_{\substack{H:H^\top H = I \\ \forall n \in [N]: \|y_n\|_2 = 1}} \sum_{n \in [N]} \|x_n - Hy_n\|_2^2, \quad H \in \mathbb{R}^{(d+1)\times(K+1)},$$

- (dai and muller) spherical pca solves the following problem (for functional data):

$$\mathrm{dist}\,.(\mathbb{S}_H^d \,|\, \mathscr{X}) = N^{-1} \sum_{n \in [N]} \left(d(x_n, \mathscr{P}_H(x_n))\right)^2$$

- (chami et al.) hyperbolic pca aims to maximize an intrinsic variance objective, i.e.,

$$\sigma_{\mathbb{H}}^2\left(\mathscr{X}\right) = N^{-2} \sum_{i,j \in [N]} d(x_i, x_j)^2$$

tabaghi *et al.*, 2022.

other formulations and applications

spherical and hyperbolic pca

- (liu et al.) spherical pca is the following matrix factorization problem:

$$\min_{\substack{H : H^\top H = I \\ \forall n \in [N] : \|y_n\|_2 = 1}} \sum_{n \in [N]} \|x_n - H y_n\|_2^2, \quad H \in \mathbb{R}^{(d+1)\times(K+1)},$$

- a new matrix factorization (hyperbolic pca):

$$\min_{\substack{H : H^\top J_d H = J_d \\ \forall n \in [N] : [y_n, y_n] = -1}} \sum_{n \in [N]} \|x_n - H y_n\|_2^2, \quad H \in \mathbb{R}^{(d+1)\times(K+1)},$$

- applications: data w/ cosine distance, hyperbolic embeddings, and (maybe?) hyperbolic matrix factorization of real data

tabaghi *et al.*, 2022.

thank you!