Principal Decomposition with Nested Submanifolds

Su Jiaji

College of Mathematics and Statistics
Chongqing University

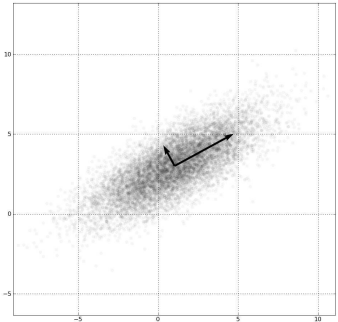
Shanghai Institute for Mathematics and Interdisciplinary Sciences
May 28, 2026

Principal Component Analysis

- $\{x_1, \ldots, x_n\} \subset \mathbb{R}^D$: points of interest
- $\mathcal{A}_{D-1} = \{z \in \mathbb{R}^D : v_D^\top (z - p) = 0\}$ with $\|v_D\| = 1$
- minimize $\sum_i \mathsf{d}(x_i, \mathcal{A}_{D-1})^2 = \sum_i \big(v_D^\top (x_i - p)\big)^2$
- One solution: $p = \bar{x}$,

$$v_D = \underset{v \in \mathbb{R}^D,\, \|v\|=1}{\arg\min}\; v^\top \sum_i (x_i - \bar{x})(x_i - \bar{x})^\top v$$

- Remaining components: $x_i' = x_i - v_D v_D^\top (x_i - \bar{x})$.
- Repeat with $\{x_i'\} \subset \mathcal{A}_{D-1}$ ...



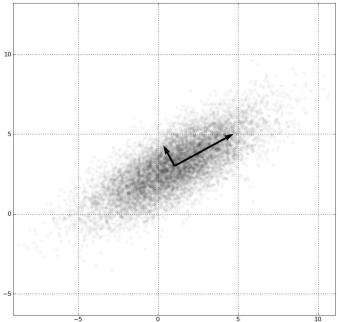
Principal Component Analysis

- $\{x_1, \ldots, x_n\} \subset \mathbb{R}^D$: points of interest
- $\mathcal{A}_{D-1} = \{z \in \mathbb{R}^D : v_D^\top (z - p) = 0\}$ with $\|v_D\| = 1$
- minimize $\sum_i \mathsf{d}(x_i, \mathcal{A}_{D-1})^2 = \sum_i \big(v_D^\top (x_i - p)\big)^2$
- One solution: $p = \bar{x}$,

$$v_D = \underset{v \in \mathbb{R}^D,\, \|v\|=1}{\arg\min}\; v^\top \sum_i (x_i - \bar{x})(x_i - \bar{x})^\top v$$

- Remaining components: $x_i' = x_i - v_D v_D^\top (x_i - \bar{x})$.
- Repeat with $\{x_i'\} \subset \mathcal{A}_{D-1}$ ...

$\hookrightarrow \mathcal{A}_d = \bar{x} + \mathsf{span}(v_1, \ldots, v_d)$



$\{\bar{x}\} \subset \mathcal{A}_1 \subset \cdots \mathcal{A}_{D-1} \subset \mathbb{R}^D$: a *flag*.

Why did PCA become so successful?

- **Geometry:** closest fitting lines and planes [Pearson, 1901].
- **Statistical decomposition:** principal components as uncorrelated variables explaining maximal variance [Hotelling, 1933, Hotelling, 1936].
- **Population/sample link:** sampling properties of principal components [Girshick, 1936, Girshick, 1939].
- **Asymptotic theory:** limiting behavior of sample PCs [Anderson, 1963].
- **Computation:** practical SVD algorithms made PCA routine in data analysis.

PCA became influential because geometry, statistics, computation, and asymptotics all aligned.

PCA as a nested geometric decomposition

PCA gives more than one low-dimensional display:

$$\{\bar{x}\} \subset \mathcal{A}_1 \subset \mathcal{A}_2 \subset \cdots \subset \mathcal{A}_D = \mathbb{R}^D.$$

- One decomposition, many compatible summaries.
- The one-, two-, ..., d-dimensional summaries are mutually compatible.
- Each $\mathcal{A}_d = \bar{x} + \text{span}(v_1, \ldots, v_d)$ is a linear subspace.
- This flag structure gives a dimension-by-dimension decomposition, not just a single embedding.
- hope to keep this nested idea, but replaces linear subspaces by smooth nonlinear submanifolds.

Linear Decomposition can be Inadequate

- Data can take values on non-Euclidean spaces:
  - resides in known non-linear spaces:
    $\hookrightarrow$ compositional data; semi-positive definite matrices ...
  - has non-Euclidean topology:
    $\hookrightarrow$ directions; molecular geometry ...
- or distributed around lower dimensional structures
  $\hookrightarrow$ strongly nonlinear covariance structures ...

Many ways to generalize PCA

Method	What is generalized?	Typical limitation
Additive PCs	Linear combinations $\rightarrow$ additive nonlinear transformations	Often gives codimension-one additive constraints
Principal curves/ surfaces	Linear subspaces $\rightarrow$ self-consistent curves/surfaces	Nested hierarchy is not automatic
Principal flows/ submanifolds	Local principal directions on known manifolds	Usually dimension-specific
FPCA	Vectors $\rightarrow$ functions; Covariance matrix $\rightarrow$ covariance operator	Still mainly linear in a Hilbert space
PNS / torus PCA	PCA on special non-Euclidean spaces	Subspaces have special geometric restrictions
PNSM	Linear subspaces $\rightarrow$ nested smooth submanifolds	Requires local covariance and numerical fitting

Additive Principal Component*

For $X = (X_1, \ldots, X_D)^\top$, consider $\phi_i = \phi_i(X_i)$ s.t.

$$\phi_i \in H_i \subset \{\phi : \mathbb{E}[\phi(X)] = 0,\ \mathsf{Var}[\phi(X)] < \infty\},$$

$$\Phi := (\phi_1, \ldots, \phi_D) \in H_1 \times \cdots \times H_D.$$

Target: minimize $\mathsf{Var}(\sum_i \phi_i)$ subject to $\sum_i \mathsf{Var}(\phi_i) = 1$.

- Use $\phi_1(X_1) + \cdots + \phi_D(X_D)$ instead of $a_1 X_1 + \cdots + a_D X_D$
- Natural setting: L^2-type Hilbert spaces of centered finite-variance transformations
- Focus on the smallest PC/low-variance relationships
- Find $\sum \phi_i(X_i) \approx 0$
- $\{x = (x_1, \ldots, x_D) \in \mathbb{R}^D : \sum \phi_i(x_i) = 0\}$: an *additive manifold with co-dimensional 1*

*Donnell, D. J., Buja, A., & Stuetzle, W. (1994). *Analysis of additive dependencies and concurvities using smallest additive principal components*. AoS

Principal Curves[†]

Find a smooth curve passing the 'middle' of the data set/minimizing variation orthogonal to the curve

- $\gamma(t)$: a smooth curve with parameter t
- $\gamma^{-1}(x) = \arg\min_t \|x - \gamma(t)\|$
- Principal curve: $\mathbb{E}[X \mid \gamma^{-1}(X) = t] = \gamma(t)$
- Estimator:
  Let $\gamma_0(t) = \bar{x} + vt$, define
  $$\gamma_j(t) = \mathbb{E}[X \mid \gamma_{j-1}^{-1}(X) = t]$$

[†] Hastie, T., & Stuetzle, W. (1989). *Principal curves*. JASA.

Functional PCA[‡]

Treat functional observations as random elements in $L^2(\mathcal{T})$:

$$\langle f, g \rangle = \int_{\mathcal{T}} f(t) g(t) \, \mathrm{d}t.$$

The covariance matrix becomes a covariance operator

$$(\mathcal{C}f)(t) = \int_{\mathcal{T}} C(s,t) f(s) \, \mathrm{d}s,$$

and FPCA solves

$$\mathcal{C}\phi_k = \lambda_k \phi_k, \qquad \xi_{ik} = \langle X_i - \mu, \phi_k \rangle.$$

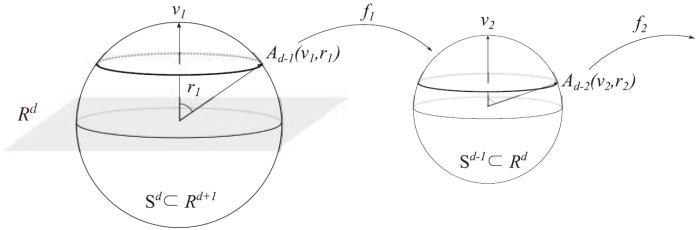
- FPCA is linear PCA in an infinite-dimensional Hilbert space.
- In practice, smoothing, basis expansion, and spectral truncation give finitely many scores.
- Truncation is usually an approximation, not necessarily a true finite-rank assumption.

[‡]Starting from JO Ramsay et al, 1990s

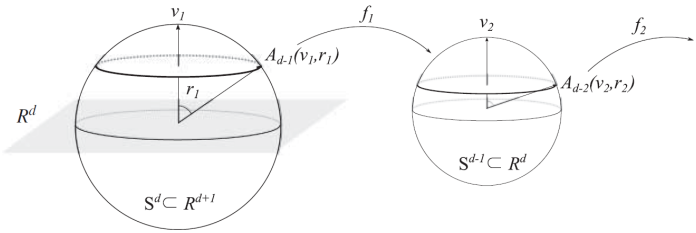
Principal Nested Spheres[§]

- $\{x_i^{(0)}\}_{i=1}^n \subset \mathcal{S}^D \subset \mathbb{R}^{D+1}$
- For $p, q \in \mathcal{S}^D$, $\mathsf{d}(p, q) = \arccos(p^\top q)$
- With $v \in \mathcal{S}^D$ and $r \in (0, \pi/2]$, a sub-sphere of $\mathcal{S}^D$:

$$A_{D-1}(v, r) = \{x \in \mathcal{S}^D : \mathsf{d}(v, x) = r\} = \mathcal{S}^D \cap \{x \in \mathbb{R}^{D+1} : v^\top x = \cos(r)\}$$



[§] Jung, S., Dryden, I. L., & Marron, J. S. (2012). *Analysis of principal nested spheres.* Biometrika.

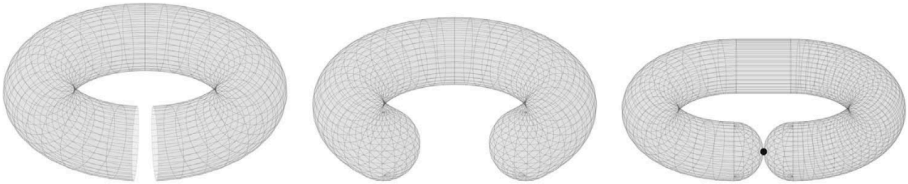


- Let $\mathcal{L}_j(v,r) = \sum_{i=1}^n \left(\cos^{-1}(v^\top x_i^{(j-1)}) - r\right)^2$, and

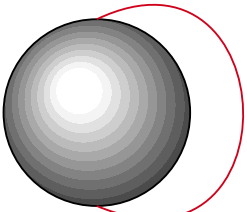
$$(\widehat{v}_j, \widehat{r}_j) = \underset{v \in \mathcal{S}^{D-j+1}, r \in (0,2\pi]}{\arg\min} \mathcal{L}_{j-1}(v,r).$$

- $\widehat{A}_{D-j} = A_{D-j}(\widehat{v}_j, \widehat{r}_j)$
- project $x_i^{(j-1)}$ along the geodesic to $\widehat{A}_{D-j}$
- re-scale and rewrite coordinates to make $\{x_i^{(j)}\} \subset \mathcal{S}^{D-j}$

Torus PCA[¶]

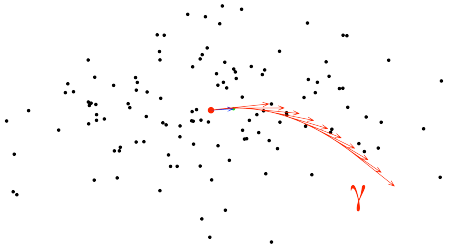


- Data on torus: $T^D = \mathcal{S}^1 \times \cdots \times \mathcal{S}^1$
- Have gaps on most dimensions
- Cut into special spheres and perform PNS



[¶] Eltzner, B., Huckemann, S., & Mardia, K. V. (2018). *Torus principal component analysis with applications to RNA structure*. AoAS

Principal Flows[‖]

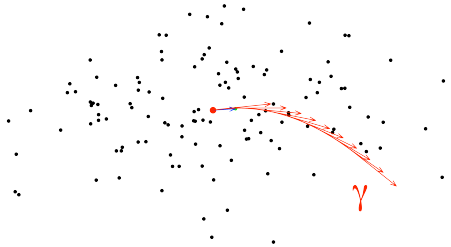


Find a smooth curve $\gamma(t)$ s.t.

- residing on known Riemannian manifold $(\mathcal{M}, g)$
- passing the 'middle' of data

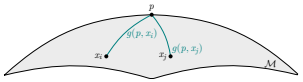
[‖] Panaretos, V. M., Pham, T., & Yao, Z. (2014). *Principal flows*. JASA.

Principal Flows[‖]



Find a smooth curve $\gamma(t)$ s.t.

- residing on known Riemannian manifold $(\mathcal{M}, g)$
- passing the 'middle' of data
- start from $\gamma(0)$, the *Fréchet mean*



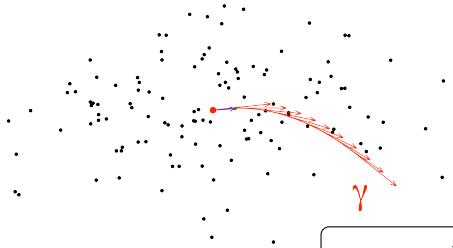
$$\Psi(p) = \sum g^2(p, x_i)/N$$

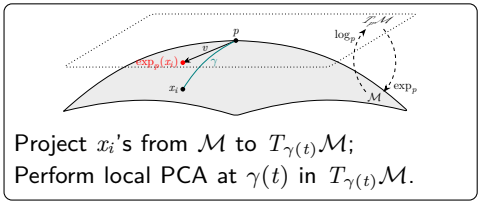
Fréchet mean: $\arg\min_{p\in\mathcal{M}} \Psi(p)$

Fréchet variance: $\min_{p\in\mathcal{M}} \Psi(p)$

[‖] Panaretos, V. M., Pham, T., & Yao, Z. (2014). *Principal flows*. JASA.

Principal Flows[‖]





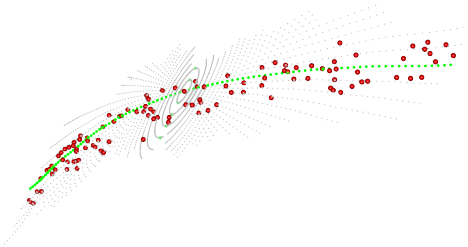
Project x_i's from $\mathcal{M}$ to $T_{\gamma(t)}\mathcal{M}$;
Perform local PCA at $\gamma(t)$ in $T_{\gamma(t)}\mathcal{M}$.

Find a smooth curve $\gamma(t)$ s.t.

- residing on known Riemannian manifold $(\mathcal{M}, g)$
- passing the 'middle' of data
- start from $\gamma(0)$, the *Fréchet mean*
- $\dot{\gamma}$ fits the first principal direction of the *tangent space PCA* at $\gamma(t)$.

[‖] Panaretos, V. M., Pham, T., & Yao, Z. (2014). *Principal flows*. JASA.

Principal Sub-manifold**



Find a smooth sub-manifold s.t.

- residing on known Riemannian manifolds; passing the 'middle' of data
- start from the Fréchet mean
- the ray $\dot{\gamma}$ fits the principal direction derived from the tangent space PCA at $\gamma(t)$
- the collection of all such rays is the principal sub-manifold.

**Yao, Z., Eltzner, B., & Pham, T. (2016–2024). *Principal sub-manifolds*. Statistica Sinica, to appear.

Some Issues

Nonlinear embedding methods:

- often give one representation for one chosen target dimension

Principal curves, flows, and submanifolds:

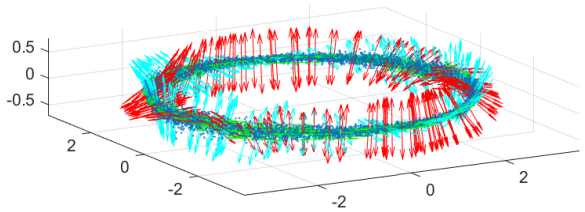
- capture nonlinear geometry
- do not automatically produce a dimension-by-dimension nested hierarchy

Principal nested spheres and torus PCA:

- provide nested decompositions
- are restricted to special geometric subspaces

$\hookrightarrow$ The missing object is not just a nonlinear representation, but a nonlinear analogue of the PCA flag.

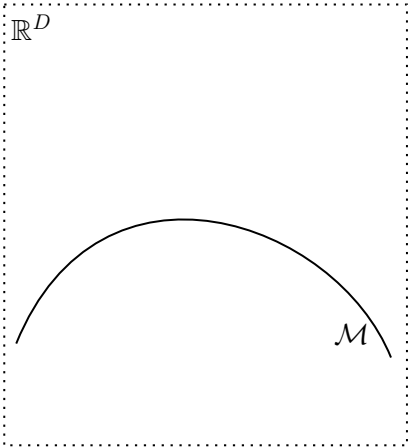
Intuitions: Decompose based on Smooth Covariance Structures



Fitting submanifolds such that

- tangent spaces approximate the linear space generated by principal directions
- passing through the 'middle' of the data clouds
- with different dimensionalities and a nested structure

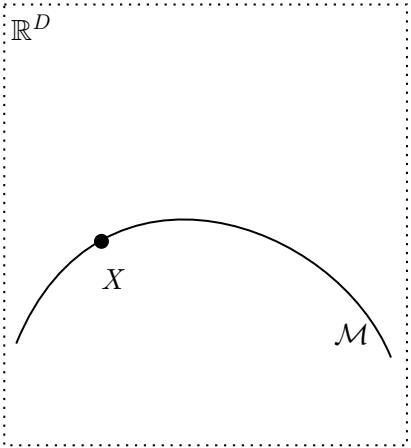
Manifold Fitting Setup*



- $\mathcal{M} \subset \mathbb{R}^D$: smooth latent manifold
- $\dim(\mathcal{M}) = d$, $d < D$, $\text{reach}(\mathcal{M}) \geq \tau$

*Fefferman, C., et al. (2018). *Fitting a putative manifold to noisy data.* PMLR.
Yao, Z., & Xia, Y. (2019 – 2025). *Manifold fitting under unbounded noise.* JMLR, to appear.
Yao, Z., Su, J., Li, B., & Yau, S.T. (2023). *Manifold fitting.* arXiv preprint arXiv:2304.07680. ...

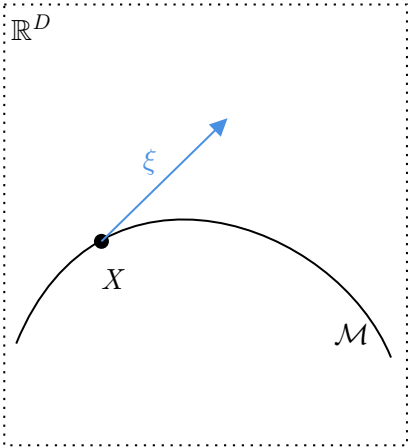
Manifold Fitting Setup*



- $\mathcal{M} \subset \mathbb{R}^D$: smooth latent manifold
- $\dim(\mathcal{M}) = d$, $d < D$, $\text{reach}(\mathcal{M}) \geq \tau$
- $X \in \mathcal{M} \subset \mathbb{R}^D$: unobserved sample from $\mu(\mathcal{M})$

*Fefferman, C., et al. (2018). *Fitting a putative manifold to noisy data.* PMLR.
Yao, Z., & Xia, Y. (2019 – 2025). *Manifold fitting under unbounded noise.* JMLR, to appear.
Yao, Z., Su, J., Li, B., & Yau, S.T. (2023). *Manifold fitting.* arXiv preprint arXiv:2304.07680. ...

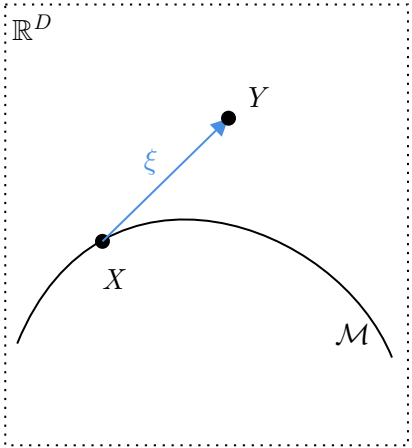
Manifold Fitting Setup*



- $\mathcal{M} \subset \mathbb{R}^D$: smooth latent manifold
- $\dim(\mathcal{M}) = d$, $d < D$, $\text{reach}(\mathcal{M}) \geq \tau$
- $X \in \mathcal{M} \subset \mathbb{R}^D$: unobserved sample from $\mu(\mathcal{M})$
- $\xi \in \mathbb{R}^D$, $\xi \sim \phi_\sigma^{(D)}$: ambient space noise

*Fefferman, C., et al. (2018). *Fitting a putative manifold to noisy data.* PMLR.
Yao, Z., & Xia, Y. (2019 – 2025). *Manifold fitting under unbounded noise.* JMLR, to appear.
Yao, Z., Su, J., Li, B., & Yau, S.T. (2023). *Manifold fitting.* arXiv preprint arXiv:2304.07680. ...

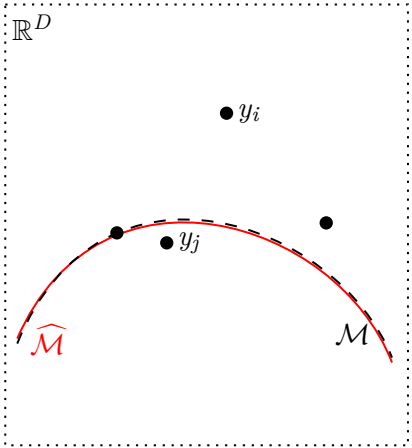
Manifold Fitting Setup*



- $\mathcal{M} \subset \mathbb{R}^D$: smooth latent manifold
- $\dim(\mathcal{M}) = d$, $d < D$, $\text{reach}(\mathcal{M}) \geq \tau$
- $X \in \mathcal{M} \subset \mathbb{R}^D$: unobserved sample from $\mu(\mathcal{M})$
- $\xi \in \mathbb{R}^D$, $\xi \sim \phi_\sigma^{(D)}$: ambient space noise
- $Y = X + \xi \sim \nu = \mu \star \phi_\sigma^{(D)}$: observation

*Fefferman, C., et al. (2018). *Fitting a putative manifold to noisy data.* PMLR.
Yao, Z., & Xia, Y. (2019 – 2025). *Manifold fitting under unbounded noise.* JMLR, to appear.
Yao, Z., Su, J., Li, B., & Yau, S.T. (2023). *Manifold fitting.* arXiv preprint arXiv:2304.07680. ...

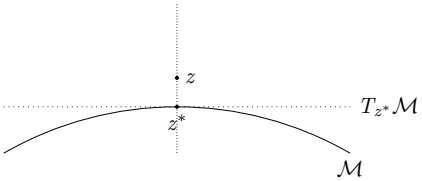
Manifold Fitting Setup*



- $\mathcal{M} \subset \mathbb{R}^D$: smooth latent manifold
- $\dim(\mathcal{M}) = d$, $d < D$, $\text{reach}(\mathcal{M}) \geq \tau$
- $X \in \mathcal{M} \subset \mathbb{R}^D$: unobserved sample from $\mu(\mathcal{M})$
- $\xi \in \mathbb{R}^D$, $\xi \sim \phi_\sigma^{(D)}$: ambient space noise
- $Y = X + \xi \sim \nu = \mu \star \phi_\sigma^{(D)}$: observation
- Estimate $\mathcal{M}$ with $\widehat{\mathcal{M}}$ based on $\{y_i\}_{i=1}^N$

*Fefferman, C., et al. (2018). *Fitting a putative manifold to noisy data.* PMLR.
Yao, Z., & Xia, Y. (2019 – 2025). *Manifold fitting under unbounded noise.* JMLR, to appear.
Yao, Z., Su, J., Li, B., & Yau, S.T. (2023). *Manifold fitting.* arXiv preprint arXiv:2304.07680. ...

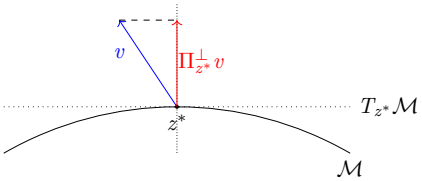
Fit the Latent Manifold†



- z: a point of interest
- $z^* = \arg\min_{z' \in \mathcal{M}} \mathsf{d}(z, z')$: projection of z on $\mathcal{M}$
- $T_{z^*}\mathcal{M}$: tangent space of $\mathcal{M}$ at z^*

† Yao, Z., & Xia, Y. (2019 – 2025). *Manifold fitting under unbounded noise.* JMLR, to appear.

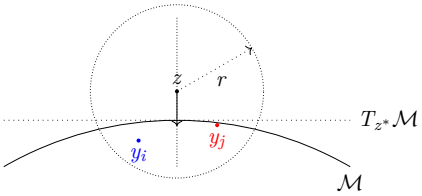
Fit the Latent Manifold†



- z: a point of interest
- $z^* = \arg\min_{z' \in \mathcal{M}} \mathsf{d}(z, z')$: projection of z on $\mathcal{M}$
- $T_{z^*}\mathcal{M}$: tangent space of $\mathcal{M}$ at z^*
- $\Pi^{\perp}_{z^*}$: projection matrix onto the normal space of $T_{z^*}\mathcal{M}$
- $\widehat{\Pi}^{\perp}_{z}$: estimator of $\Pi^{\perp}_{z^*}$

† Yao, Z., & Xia, Y. (2019 – 2025). *Manifold fitting under unbounded noise.* JMLR, to appear.

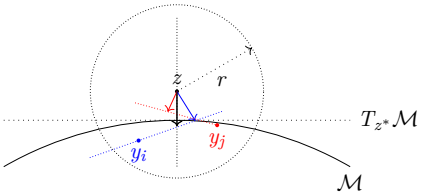
Fit the Latent Manifold†



- $r = \mathcal{O}(\sqrt{\sigma})$

†Yao, Z., & Xia, Y. (2019 – 2025). *Manifold fitting under unbounded noise.* JMLR, to appear.

Fit the Latent Manifold†



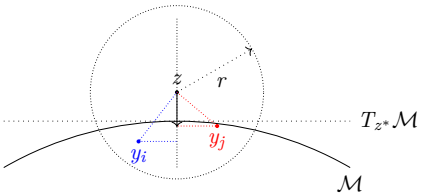
$$\tilde{\alpha}_i(z) = \left(1 - \frac{\|z - y_i\|^2}{r^2}\right)^{\beta} \mathbb{I}(\|z - y_i\| \le r)$$

$$\alpha_i(z) = \frac{\tilde{\alpha}_i(z)}{\sum_{i=1}^{n} \tilde{\alpha}_i(z)}, \quad \beta > 2,$$

- $r = \mathcal{O}(\sqrt{\sigma})$
- $\widehat{\Pi}_{y_i}^{\perp}$: estimator of $\Pi_{y_i^*}^{\perp}$
- $\Psi_z = \mathbb{P}_{D-d}\left(\sum_i \alpha_i(z)\widehat{\Pi}_{y_i}^{\perp}\right)$: estimator of $\Pi_{z^*}^{\perp}$
- $\mathbb{P}_d(A)$: projection of matrix A on its leading d eigenvectors

† Yao, Z., & Xia, Y. (2019 – 2025). *Manifold fitting under unbounded noise.* JMLR, to appear.

Fit the Latent Manifold†



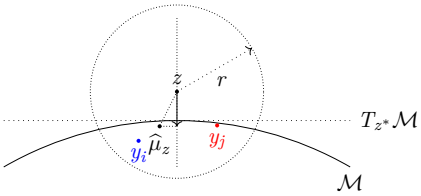
- $\Psi_z = \mathbb{P}_{D-d}\left(\sum_i \alpha_i(z)\widehat{\Pi}^{\perp}_{y_i}\right)$
- $f(z) = \sum_i \alpha_i(z)\Psi_z(z - y_i)$

$$\tilde{\alpha}_i(z) = \left(1 - \frac{\|z - y_i\|^2}{r^2}\right)^{\beta} \mathbb{I}(\|z - y_i\| \le r)$$

$$\alpha_i(z) = \frac{\tilde{\alpha}_i(z)}{\sum_{i=1}^{n} \tilde{\alpha}_i(z)}, \quad \beta > 2,$$

†Yao, Z., & Xia, Y. (2019 – 2025). *Manifold fitting under unbounded noise.* JMLR, to appear.

Fit the Latent Manifold†



$$\widehat{\mathcal{M}} = \{z \in \mathbb{R}^D : \mathsf{d}(z, \mathcal{M}) \le cr, c < 1, f(z) = 0\}$$

$$\Rightarrow \mathsf{d}(z, \mathcal{M}) \le Cr^2 \text{ for any } z \in \widehat{\mathcal{M}}$$

with probability

$$1 - d\exp\{-cNr^{d+2}\}.$$

- $\Psi_z = \mathbb{P}_{D-d}\left(\sum_i \alpha_i(z)\widehat{\Pi}^{\perp}_{y_i}\right)$
- $\widehat{\mu}_z = \sum_i \alpha_i(z) y_i$: weighted mean of y_i
- $f(z) = \sum_i \alpha_i(z)\Psi_z(z - y_i) = \Psi_z(z - \widehat{\mu}_z)$

†Yao, Z., & Xia, Y. (2019 – 2025). *Manifold fitting under unbounded noise.* JMLR, to appear.

How Does This Work?

With properly designed weighting functions:

$$\tilde{\alpha}_i(z) = \left(1 - \frac{\|z - y_i\|^2}{r^2}\right)^{\beta} \mathbb{I}(\|z - y_i\| \leq r), \quad \alpha_i(z) = \frac{\tilde{\alpha}_i(z)}{\sum_{i=1}^{n} \tilde{\alpha}_i(z)}, \quad \beta > 2,$$

- Ψ_z is second order smooth, and estimates $\Pi_{z^*}^{\perp}$ well;
- the Jacobian matrix $J_f(x)$ satisfies $\|J_f(x) - \Psi_z\|_F \leq C\sigma/r$

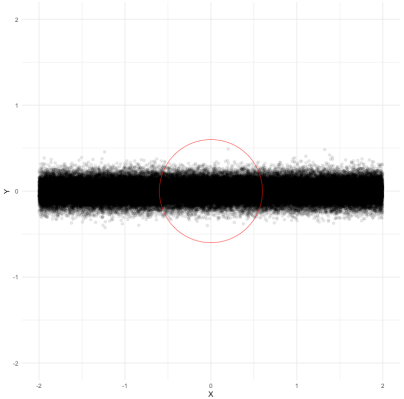
Together with the smoothness of $\mathcal{M}$, for $x \in \widehat{\mathcal{M}}$, in its neighbourhood

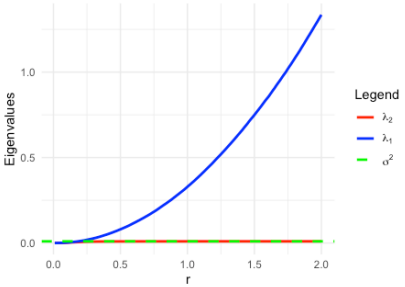
- construct $h(z) = V_x^{\top} f(z)$, which is rank $D - d$
- $f(z) = 0$ iff $h(z) = 0$

$\hookrightarrow \widehat{\mathcal{M}}$ is d-dimensional locally. Reach can be derived from Hessian of f.

Kernel methods and the remaining geometric question

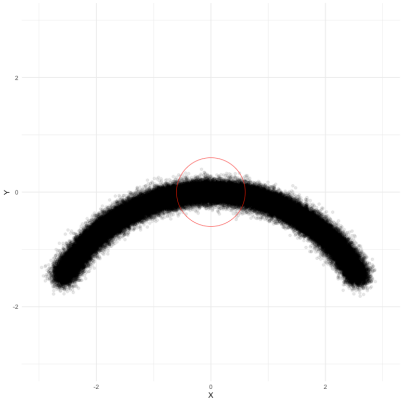
Kernel PCA and kernel embeddings can provide nonlinear representations.

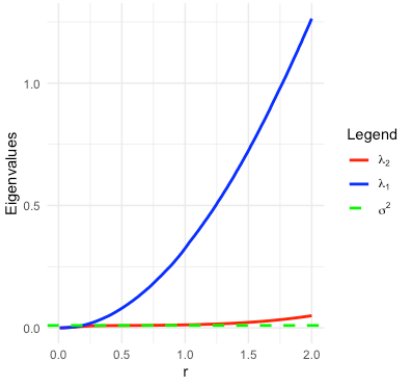




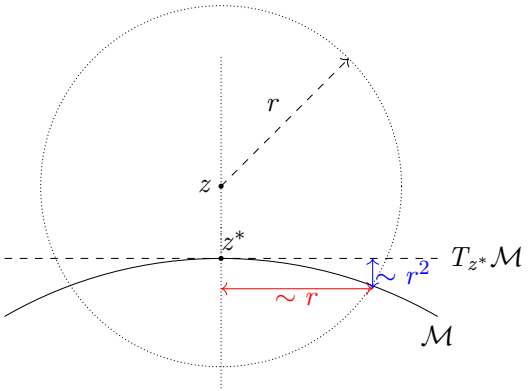
Kernel methods and the remaining geometric question

Kernel PCA and kernel embeddings can provide nonlinear representations.





Kernel methods and the remaining geometric question



- local covariance identifies small-variance directions
- $P(Y \in \mathcal{B}_D(z, r)) \propto r^d / \mathrm{vol}(\mathcal{M})$
- Given $Y \in \mathcal{B}_D(z, r)$, the variance of Y:
  $\leftrightarrow \sim r^2 + \sigma^2$
  $\updownarrow \sim r^4 + \sigma^2$

  $\hookrightarrow$ zero sets of bias fields define nested smooth submanifolds

Principal Nested Submanifolds – Population

- $X \in \mathbb{R}^D$: random vector of interest
- $\mathcal{X}$: support of X
- r: a radius parameter
- z: point of interest, with $\mathrm{d}(z, \mathcal{X}) < cr$
- Local average: $\mu_z = E\{X \mid X \in \mathcal{B}(z, r)\}$
- Local covariance matrix centered at z:

$$\Sigma_z = E\{(X - \mu_z)(X - \mu_z)^\top \mid X \in \mathcal{B}(z, r)\}$$

- SVD:
  - $\lambda_{z,1} < \cdots < \lambda_{z,D}$
  - $v_{z,1}, \ldots, v_{z,D}$, with $\|v_{z,j}\| = 1$

Principal Nested Submanifolds – Population

- $\Pi_{z,j} = v_{z,j} v_{z,j}^\top$: projection matrix onto the j-th principal direction at z

Consider the jth bias vector

$$f_j(z) = \Pi_{z,j}(\mu_z - z),$$

and let

$$\mathcal{M}_d = \{z \in \mathbb{R}^D : \ \mathrm{d}(z, \mathcal{X}) \leq cr, \ \sum_{j=1}^{D-d} f_j(z) = 0\}.$$

Then, if X satisfies concentration, identifiability, and smoothness assumptions, there is

- $\dim(\mathcal{M}_d) = d$
- $\mathcal{M}_1 \subset \mathcal{M}_2 \subset \cdots \subset \mathcal{M}_{D-1} \subset \mathbb{R}^D$

Role of the fixed radius r

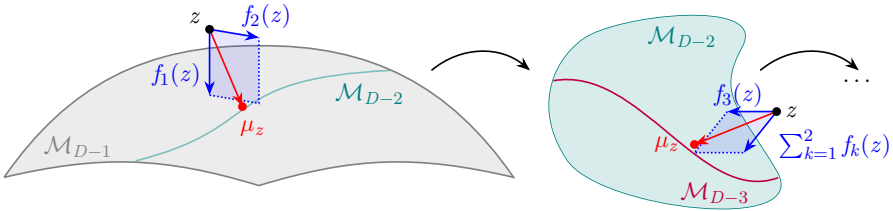
PNSM targets a geometric hierarchy:
The radius parameter r controls the locality, or smoothing scale, of the construction.

- Local means and covariances are computed inside neighborhoods of size r.
- Using the same r across dimensions gives a common geometric scale for

$$\mathcal{M}_{r,1} \subset \mathcal{M}_{r,2} \subset \cdots \subset \mathcal{M}_{r,D-1}.$$

- Different fixed choices of r produce different-resolution nested hierarchies.
- Allowing r to vary by dimension may be useful, but weakens the interpretation of one coherent hierarchy.

Intuition of Principal Nested Submanifolds



Empirical

- $\mathcal{X}_D = \{x_i\}_{i=1}^N \subset \mathbb{R}^D$
- Local covariance matrices:

$$\widehat{\Sigma}_i = \frac{\sum_j (x_i - x_j)(x_i - x_j)^\top \mathbb{I}(\|x_i - x_j\| < r)}{\sum_j \mathbb{I}(\|x_i - x_j\| < r)}$$

- $\widehat{\Pi}_{i,j}$: projection matrix onto the j-th smallest principal direction of $\widehat{\Sigma}_i$
- Weighting functions:

$$\tilde{\alpha}_i(z) = \left(1 - \frac{\|z - x_i\|^2}{r^2}\right)^\beta \mathbb{I}(\|z - x_i\| \le r), \quad \alpha_i(z) = \frac{\tilde{\alpha}_i(z)}{\sum_{i=1}^n \tilde{\alpha}_i(z)}$$

with $\beta > 2$.

Empirical

Estimate

- $\widehat{\Pi}_{z,j} = \mathbb{P}_1\left(\sum_i \alpha_i(z)\widehat{\Pi}_{i,j}\right)$
- $\widehat{\mu}_z = \sum_i \alpha_i(z) x_i$
- $\widehat{f}_j(z) = \widehat{\Pi}_{z,j}\left(\widehat{\mu}_z - z\right)$

and define the estimator for the nested d-dimensional submanifold as

$$\widehat{\mathcal{M}}_d = \{z \in \mathbb{R}^D : \ \mathsf{d}(z, \mathcal{X}_D) \leq cr, \ \sum_{j=1}^{D-d} \widehat{f}_j(z) = 0\}.$$

Then, $\dim(\widehat{\mathcal{M}}_d) = d$, $\widehat{\mathcal{M}}_1 \subset \cdots \subset \widehat{\mathcal{M}}_{D-1} \subset \mathbb{R}^D$, and $\widehat{\mathcal{M}}_d$ is close to $\mathcal{M}_d$ with high probability.

Estimation & Projection

- Embed data on $\mathcal{M}$ in $\mathbb{R}^D$ if needed
- Start with $\mathcal{X}_D$ and calculate all $\widehat{\Pi}_{i,j}$'s
- for $d = \dim(\mathcal{M}) - 1 : 1$
  - for each element in $\mathcal{X}_{d+1}$, update with rule

$$z' = z + \sum_{j=1}^{D-d} \widehat{f}_j(z)$$

  - stop when $\| \sum_{k=1}^{j} \widehat{f}_j(z) \| < \epsilon$
  - set the converge points as $\mathcal{X}_d$
  - project $\mathcal{X}_d$ on $\mathcal{M}$ if needed.
- Return $\mathcal{X}_d$ for $d = 1, \ldots, \dim(\mathcal{M}) - 1$.

Estimator consistency

Write the population and empirical bias fields as

$$g_r(z) = \sum_{k=1}^{D-d} b_{r,k}(z), \qquad \widehat{g}_r(z) = \sum_{k=1}^{D-d} \widehat{b}_{r,k}(z).$$

Under mass, smoothness, eigengap, and transversality assumptions, for fixed r,

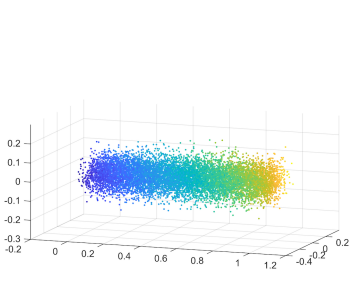
$$\sup_{z \in \mathcal{Z}} \left\| \widehat{g}_r(z) - g_r(z) \right\| = \mathcal{O}_p\left(\sqrt{\frac{\log n}{n r^D}} \right) + B_r.$$

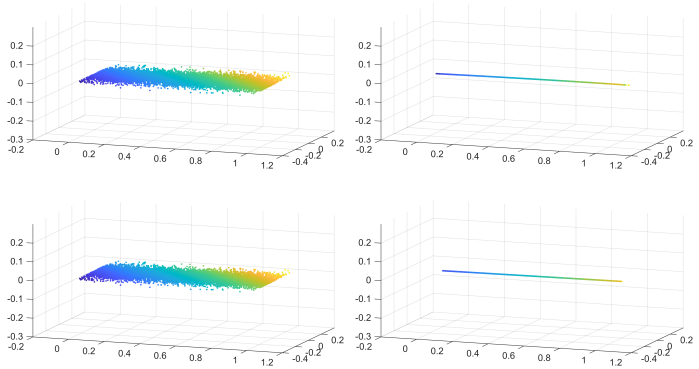
Consequently,

$$\mathsf{d_H}(\widehat{\mathcal{M}}_{r,d}, \mathcal{M}_{r,d}) \leq C \left\{ \mathcal{O}_p\left(\sqrt{\frac{\log n}{n r^D}} \right) + B_r \right\}.$$

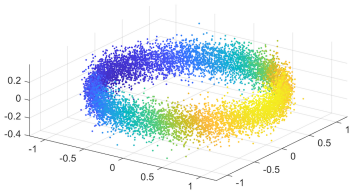
- stochastic term: local moment estimation and eigenspace perturbation
- B_r: deterministic mismatch between hard-ball population definition and smooth empirical weighting
- empirical zero sets are close to population zero sets

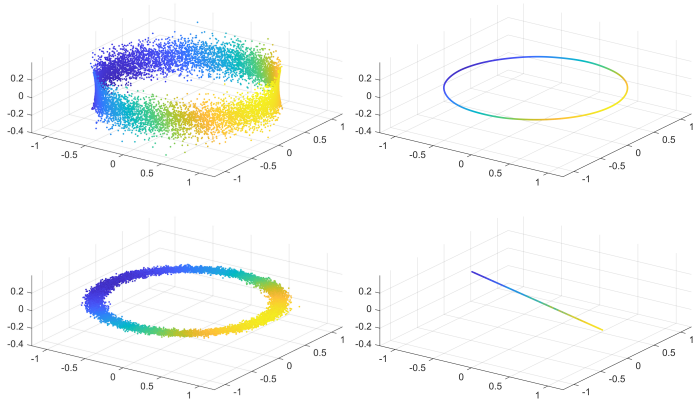
Simulation with Euclidean Space I



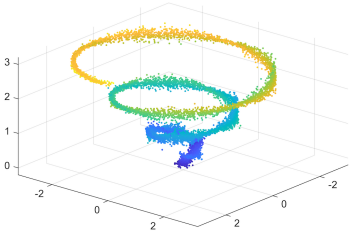


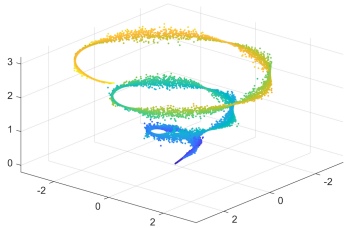
Simulation with Euclidean Space II

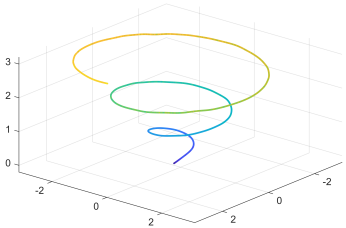


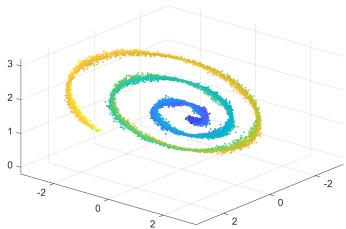


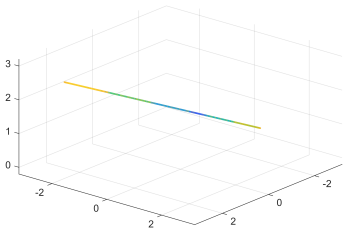
Simulation with Euclidean Space III



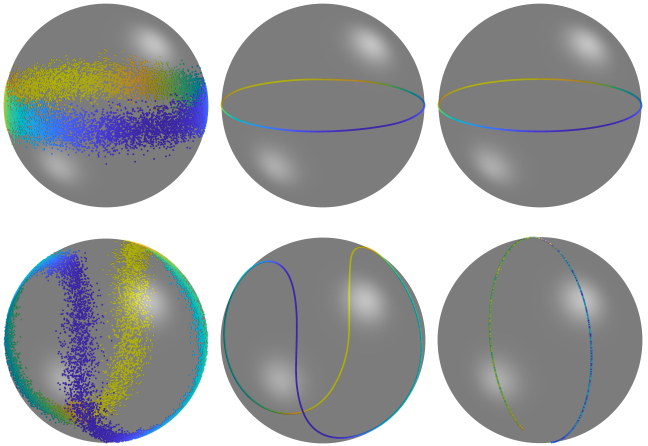


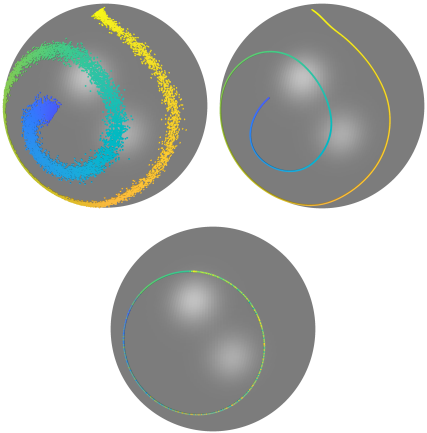




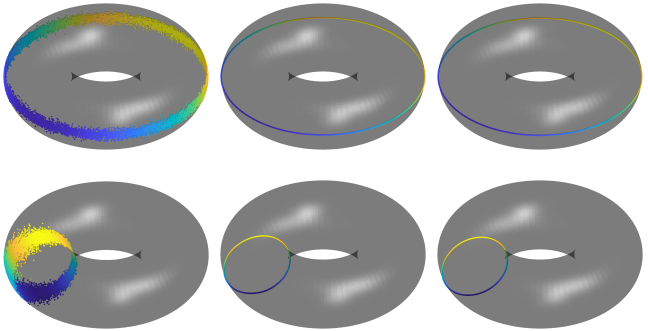


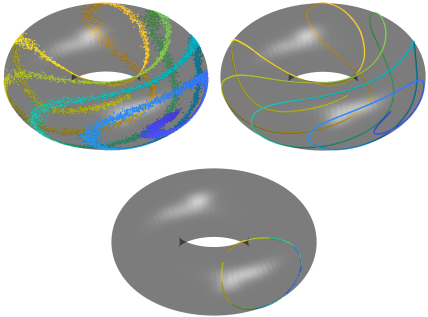
Simulation with Spheres



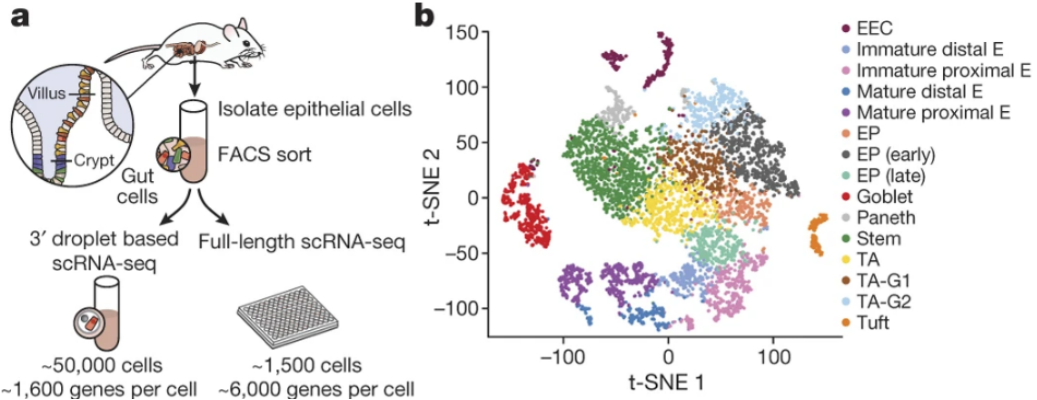


Simulation with Tori





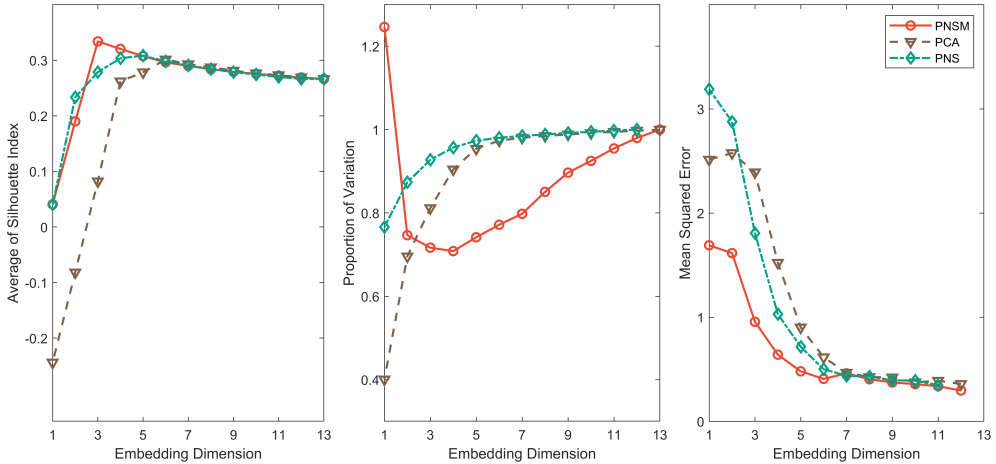
scRNA Data Set‡



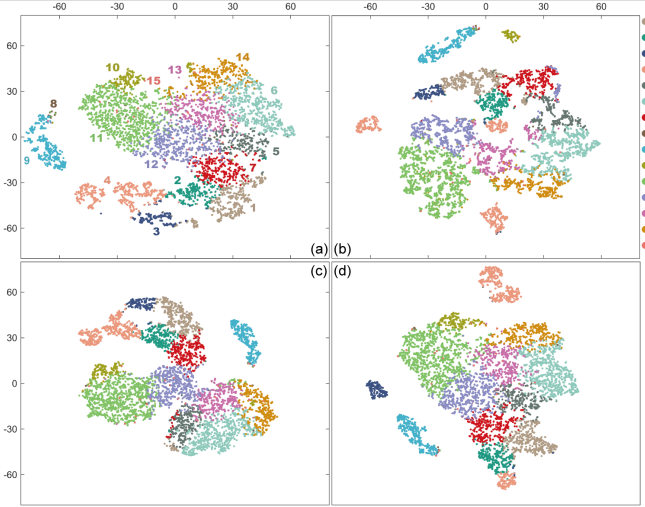
Totally 7,216 small intestinal epithelium cells from 6 mice, with the first 13 PCs.

‡Haber, Adam L., et al. (2017). *A single-cell survey of the small intestinal epithelium.* Nature.

Compare

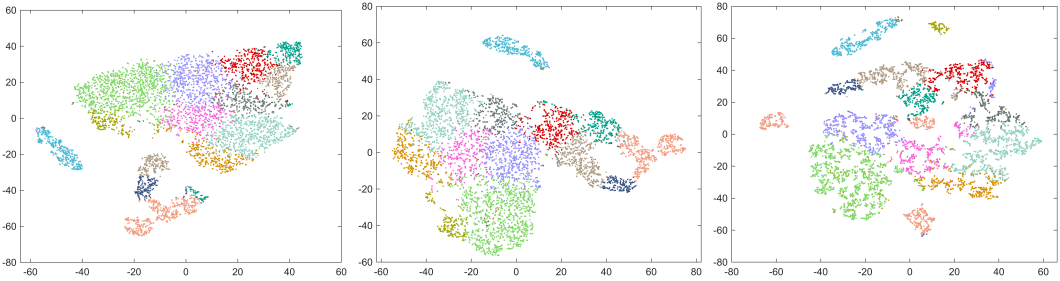

Average of Silhouette Index
Embedding Dimension
Proportion of Variation
Embedding Dimension
Mean Squared Error
Embedding Dimension
PNSM
PCA
PNS

tSNE Visualization§



§(a)Orignial Data; (b) PNSM with $d = 3$; (c) PCA with $d = 6$; (d) PNS with $d = 5$

tSNE Visualizations: $d \in \{13, 8, 3\}$



Conclusion & Outlook

- **Principal Nested Submanifolds**: a nonlinear generalization of PCA's nested geometric decomposition.
  - Replaces PCA's linear flag by nested smooth submanifolds
  - Applies to general spaces
  - Achieves superior performance in simulation and real single-cell data
- Future directions:
  - Scalable versions for large-scale omics data.
  - Integration into neural networks like generative models.

Thank you!
Questions are welcome.

Slides and preprint:

sujiaji.cn/nav

Reference

Anderson, T. W. (1963).
Asymptotic theory for principal component analysis.
The Annals of Mathematical Statistics, 34(1):122–148.

Girshick, M. (1936).
Principal components.
Journal of the American Statistical Association, 31(195):519–528.

Girshick, M. (1939).
On the sampling theory of roots of determinantal equations.
The Annals of Mathematical Statistics, 10(3):203–224.

Hotelling, H. (1933).
Analysis of a complex of statistical variables into principal components.
Journal of educational psychology, 24(6):417.

Hotelling, H. (1936).
Simplified calculation of principal components.
Psychometrika, 1(1):27–35.

Pearson, K. (1901).
Liii. on lines and planes of closest fit to systems of points in space.
The London, Edinburgh, and Dublin philosophical magazine and journal of science, 2(11):559–572.