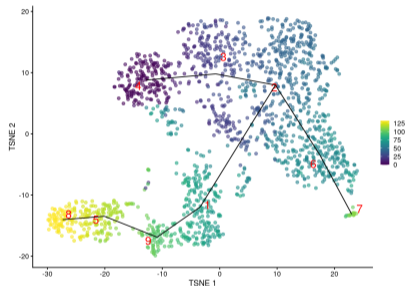
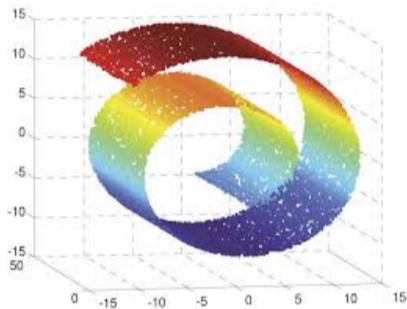


Principal Decomposition with Nested Submanifolds

Su Jiaji

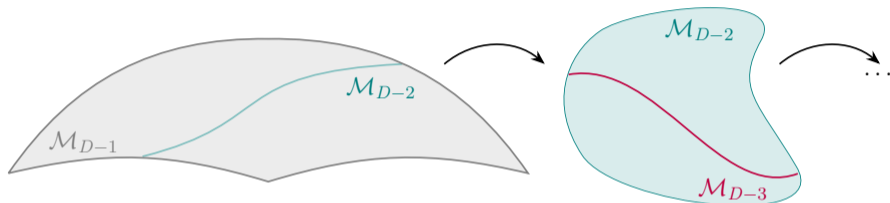
November 7, 2025
School of Statistics
Renmin University of China

Motivation: Nonlinear Structure & Nested Representation



- 真实数据常呈非线性低维结构
- 现有非线性降维缺少 **PCA** 式的层级解释
- 目标: 在一般流形上实现可解释的嵌套分解

Core Idea: Nested Dimensional Reduction



$$\mathbb{R}^D \supset \mathcal{M}_{D-1} \supset \mathcal{M}_{D-2} \supset \cdots \supset \mathcal{M}_1$$

- 在局部估计主变化方向
- 去除最小变化方向, 构造低一维子流形
- 逐层重复, 结构可解释: 每一层都有几何含义

Positioning Among Existing Methods

方法	非线性	可逐层解释	适用范围
PCA	否	是	仅近似线性结构
t-SNE/UMAP/...	是	否	仅可视化，无层级结构
Principal Flows*	是	否	拟合一维流形
Principal Nested Spheres [†]	是	是	仅适用于球面
Torus PCA [‡]	是	是	仅适用于环面
PNSM(本方法)	是	是	一般流形，多尺度嵌套结构

- 非线性方法可以降维，但缺乏 PCA 那样的逐层解释能力
- 现有嵌套方法受限于特定几何结构（如球面、单纯形等）
- PNSM = 非线性 + 可嵌套解释 + 一般流形适用性

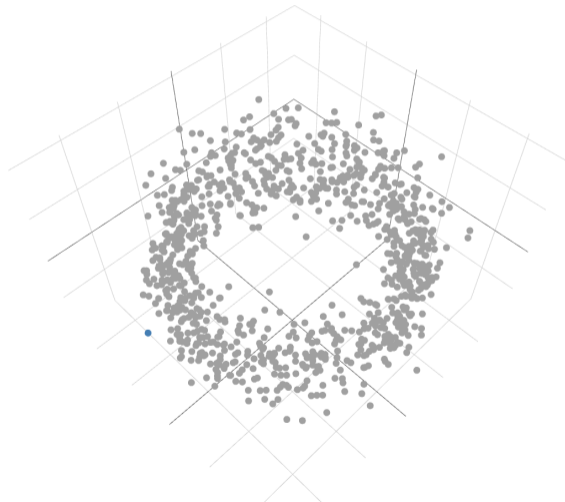
* Panaretos, V. M., Pham, T., & Yao, Z. (2014). *Principal flows*. JASA.

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

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

Methods Overview: Local Geometry \rightarrow Nested Submanifolds

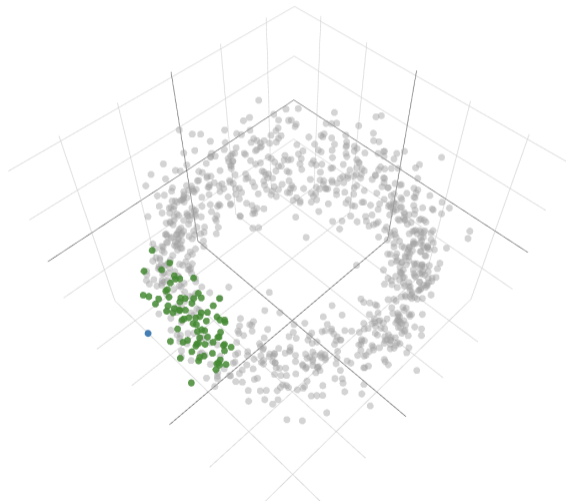
对任意 $z \in \{x_i\}$:



Methods Overview: Local Geometry \rightarrow Nested Submanifolds

对任意 $z \in \{x_i\}$:

- 确定其局部邻域

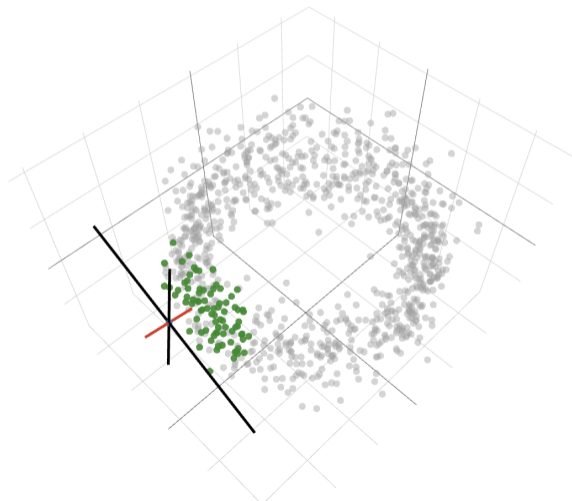


Methods Overview: Local Geometry \rightarrow Nested Submanifolds

对任意 $z \in \{x_i\}$:

- 确定其局部邻域
- 在邻域内计算 μ_z, Σ_z ,
及 Σ_z 最小特征值对应的特征向量 $v_{z,1}$
- 沿 $v_{z,1}$ 将 z 迭代投影回主子流形:

$$z' = z - (v_{z,1} v_{z,1}^\top)(z - \mu_z)$$



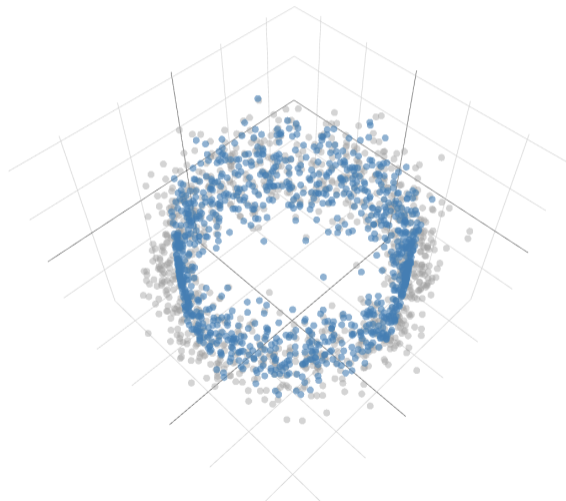
Methods Overview: Local Geometry \rightarrow Nested Submanifolds

对任意 $z \in \{x_i\}$:

- 确定其局部邻域
- 在邻域内计算 μ_z, Σ_z ,
及 Σ_z 最小特征值对应的特征向量 $v_{z,1}$
- 沿 $v_{z,1}$ 将 z 迭代投影回主子流形:

$$z' = z - (v_{z,1} v_{z,1}^\top)(z - \mu_z)$$

遍历所有数据点, 得到新的点集 $\{x'_i\}$;



Methods Overview: Local Geometry \rightarrow Nested Submanifolds

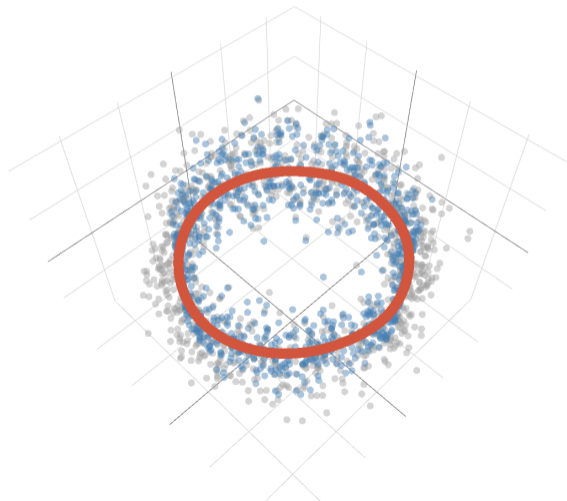
对任意 $z \in \{x_i\}$:

- 确定其局部邻域
- 在邻域内计算 μ_z, Σ_z ,
及 Σ_z 最小特征值对应的特征向量 $v_{z,1}$
- 沿 $v_{z,1}$ 将 z 迭代投影回主子流形:

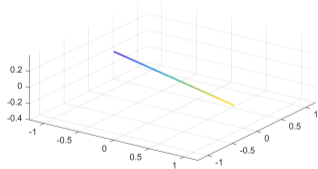
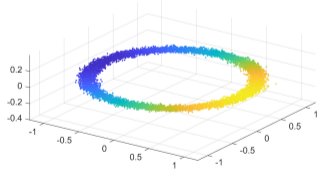
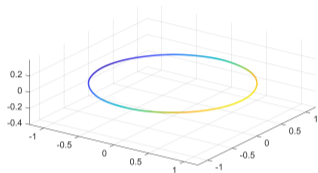
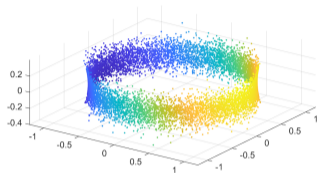
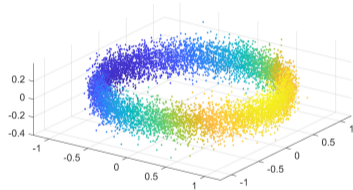
$$z' = z - (v_{z,1} v_{z,1}^\top)(z - \mu_z)$$

遍历所有数据点, 得到新的点集 $\{x'_i\}$;
在 $\{x'_i\}$ 上重复, 形成嵌套结构:

$$\mathcal{M}_{D-1} \supset \mathcal{M}_{D-2} \supset \cdots \supset \mathcal{M}_1$$

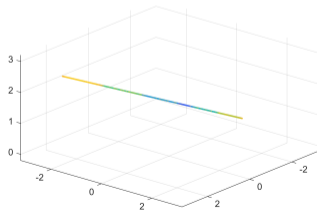
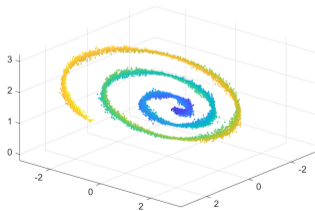
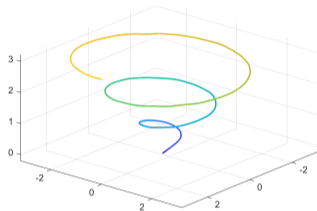
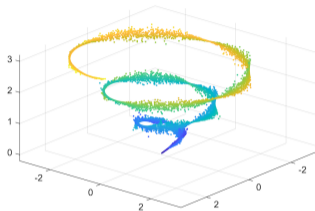
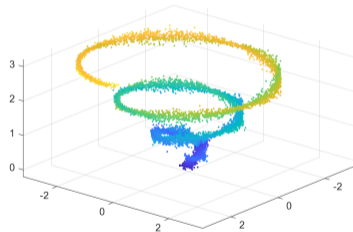


Simulations: Euclidean Data (I)



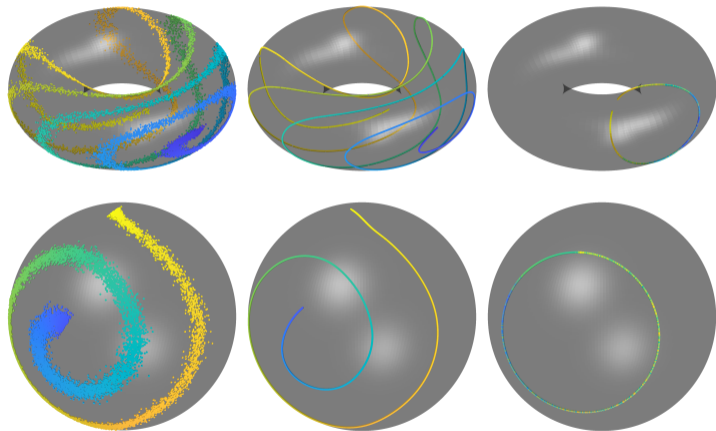
PNSM 恢复圆形主结构与层级 PCA 在线性子空间中扁平化结构

Simulations: Euclidean Data (II)

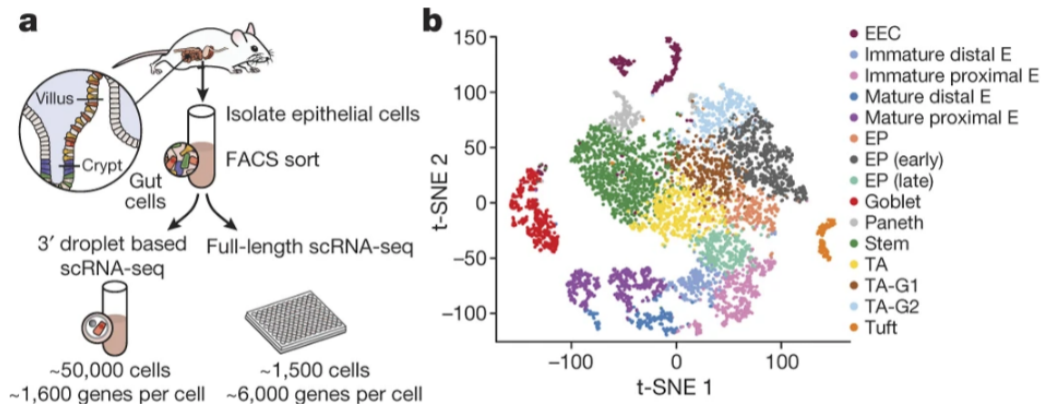


非线性曲率下仍能逐层展开主干与分支, 噪声与弯曲不破坏层级重构

Simulations: Manifolds Beyond Euclidean



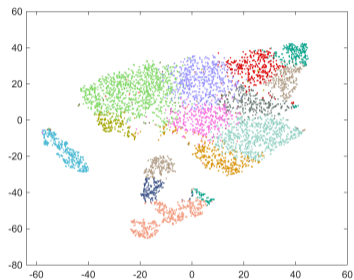
同样适用于球面环面, 与 tPCA/PNS 相比展现更好结构保持能力

Real Data: scRNA-seq[§]

共有 7216 个小肠上皮细胞, 来自 6 只小鼠, 使用前 13 个主成分。

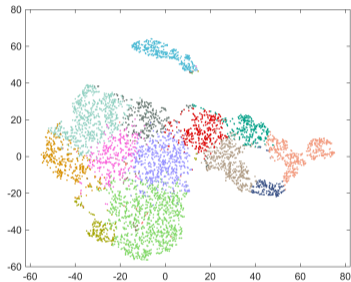
[§]Haber, Adam L., et al. (2017). *A single-cell survey of the small intestinal epithelium*. Nature.

Visualization Across Scales



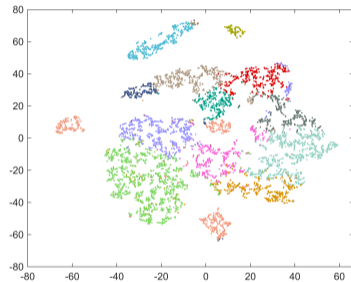
细胞混合、形状复杂

$d = 13$



主干分化轨迹显现

$d = 8$



亚群层级结构进一步展开

$d = 3$

Summary & Outlook

- **Principal Nested Submanifolds** 将 PCA 的逐层嵌套分解思想推广到一般的流形上
 - 能从数据中恢复多尺度的低维嵌套结构
 - 不依赖特定坐标系或球面几何假设, 适用范围广
 - 在模拟数据与单细胞数据中均展现出结构保持与可解释性优势
- 未来研究方向:
 - 规模化算: 近邻检索/并行加速/带宽自适应
 - 与生成模型融合: 作为几何先验嵌入表示学习与生成

谢谢聆听, 欢迎讨论!

Slides & Preprint:



sujiaji.cn/nav