Learning the Low-dimensional Space via Rotations

…or doing PCA onto spheres!

Joint work with Jeremy E. Purvis, Didong Li

Luo, Hengrui, Jeremy E. Purvis, and Didong Li. "Spherical rotation dimension reduction with geometric loss functions." *Journal of Machine Learning Research* 25.175 (2024): 1-55.
https://www.jmlr.org/papers/v25/23-0547.html

Presenter: Hengrui Luo
Keywords: dimension reduction, principal component analysis, high-dimensional dataset.

Motivating example: two circles in $\mathbb{R}^3$

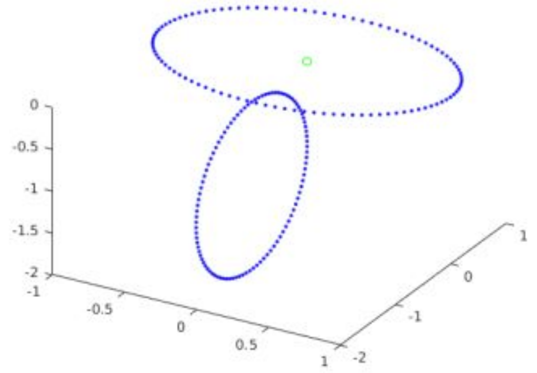
The circular structure is lost in the procedure of performing principal component analysis (PCA) and reduce one dimension: why?

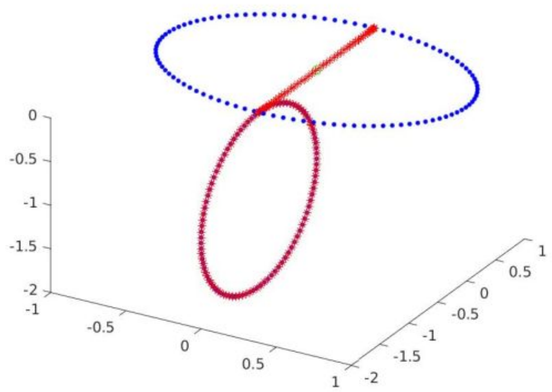
1. PCA maximizes global L^2 norm (i.e., variance)

AND

2. PCA assumes that the target space is also a plane.

Question 1: Can we perform dimension reduction by solving a new optimization problem?





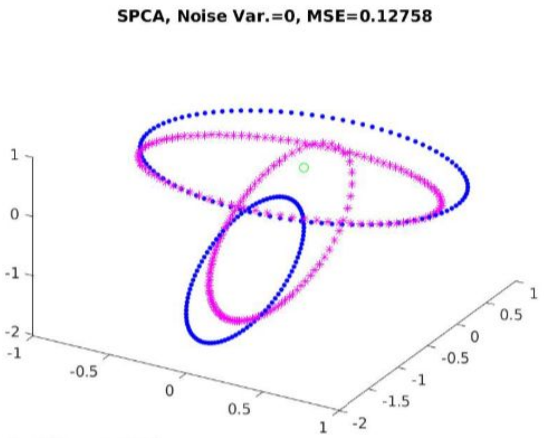
Motivating example: two circles in $\mathbb{R}^3$

The circular structure is lost in the procedure of performing principal component analysis (PCA) and reduce one dimension: why?

1. PCA maximizes global L^2 norm (i.e., variance)

AND

2. PCA assumes that the target space is also a plane.

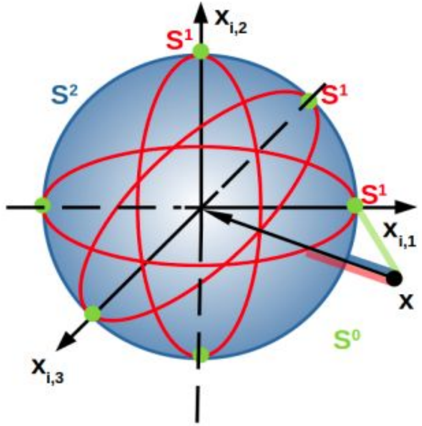


Question 2: Can we perform dimension reduction by projecting onto general manifolds/space instead of plane?

Projection onto spherical spaces $\mathbb{S}^d$

Suppose we want to perform the dimension reduction in such a way that, the reduced data falls onto a lower dimensional sphere. e.g., $\mathbb{R}^3$ to $\mathbb{S}^2$

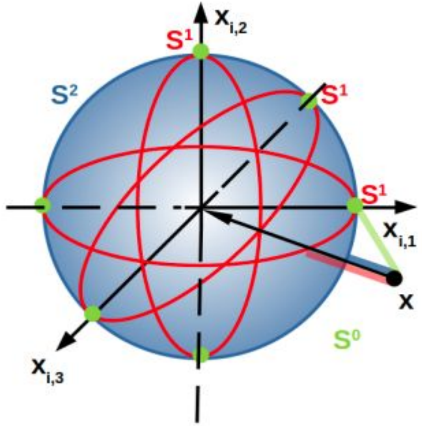
- **Solution 1:** Our SRCA develop new loss function, and works for sample size < target dimension.
- **Solution 2:** Our observation is that the projection can be achieved by solving a general optimization problem: target space (sphere with center and radius) and optimal target dimension can be determined simultaneously.



Design a loss function: point-to-sphere distance

Assume that the axes of the target space is aligned with axes of $\mathbb{R}^d$ (Eriksson and Jankowiak, 2021), the point-to-sphere distance (or point-to-ellipsoid distance with *W*) can be written out as:

$$d(x_i, S_{\mathcal{I}}(c,r))^2 = (x_i - c)^T W I_{\mathcal{I}^c}(x_i - c) + \left(\sqrt{(x_i - c)^T \sqrt{W}^T I_{\mathcal{I}} \sqrt{W}(x_i - c)} - r\right)^2$$
$$= (x_i - c)^T W (x_i - c) + r^2 - 2r\sqrt{(x_i - c)^T \sqrt{W}^T I_{\mathcal{I}} \sqrt{W}(x_i - c)}.$$



If we minimize this loss function for all data points x_i then the sphere $S_{\mathcal{I}}(c,r)$ (of axes ***I*** with axes in set $\mathcal{I} \subset \{1, 2, \cdots, d\}$ center *c*, radius *r*) is what we want.

Design a loss function: point-to-sphere distance

Strictly, this optimization problem is difficult, since we can choose 2^d different axes sets $\mathcal{I} \subset \{1, 2, \cdots, d\}$ to determine the target space.

We developed a relaxed problem constrained, with optional weight matrix W.

$$\min_{c \in \mathbb{R}^d, r \in \mathbb{R}^+} \sum_{i=1}^{n} \left((x_i - c)^T W (x_i - c) + r^2 - 2r \sqrt{(x_i - c)^T \sqrt{W}^T v^T I v \sqrt{W} (x_i - c)} \right),$$

$$\text{s.t.} \|v\|_{l_1} \leq d' + 1$$

We use different rotation methods to meet our assumption of being axis-aligned, it empirically/practically works well. However, learning optimal rotation is a long-standing problem (Arora, 2009).

We called our method Spherical Rotation Component Analysis (SRCA).

SRCA is better than PCA since…

1. Theoretically guaranteed pairwise MSE minimization

   i.e., the pairwise MSE between original points and the reduced point

2. Asymptotically guaranteed consistency and validity as an M-estimation problem
3. Empirically structural preserving
   a. Original geometrical structures in data sets
   b. Clustering preserving for most data sets
   c. Co-ranking superiority as a dimension reduction method
4. Algorithmic validity for arbitrary sample size and target dimension.

Example dataset: Ecoli (sample size > dimension)

SRCA is the best in cluster preserving

Index	Baseline	SRCA	SPCA	PCA	LLE	tSNE	UMAP
SC	0.257	0.267	0.260	0.200	0.209	0.293	0.290
CHI	133	192	190	215	46.6	376	376
DBI	1.49	1.59	1.58	2.56	2.40	1.37	1.32

SRCA is the best in MSE minimization

Dataset	Method/d' =	1	2	3	4
Ecoli	PCA	0.076693	0.035222	0.020522	**0.00756**
	SPCA	**0.047776**	0.032948	0.019648	0.01136
	SRCA	0.076660	**0.032799**	**0.018332**	**0.00756**

Example dataset: GTEx (sample size < dimension)

SRCA works for any target dimension

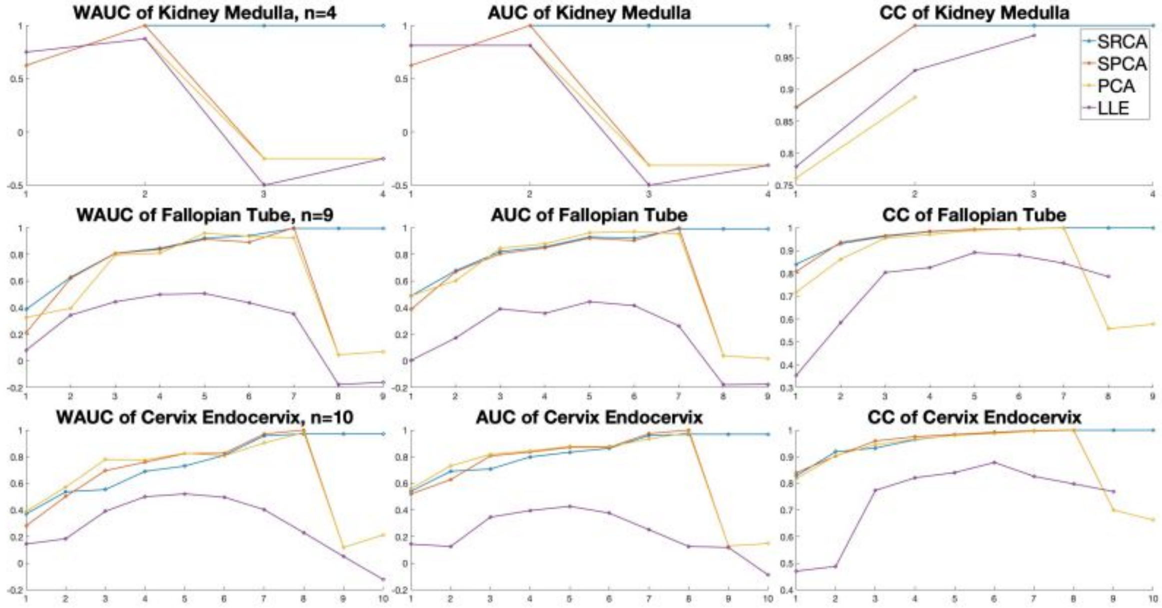


Figure 4.2: Coranking measurements of three GTEx tissues for different reduced dimension, the horizontal axes are retained dimension d', the vertical axes are score values.