K-Tensors: Clustering Positive Semi-Definite Matrices

NEW YORK UNIVERSITY

Hanchao Zhang, Thaddeus Tarpey

Division of Biostatistics, Department of Population Health, Grossman School of Medicine, New York University

NYU Grossman School of Medicine

Abstract

This poster introduces a novel **self-consistency clustering algorithm (K-Tensors)** designed for positive-semidefinite matrices based on their **eigenstructures**. As positive semi-definite matrices can be represented as **ellipses or ellipsoids** in $\Re^p$, $p \geq 2$, it is critical to maintain their structural information to perform effective clustering. However, traditional clustering algorithms often vectorize the matrices, resulting in a loss of essential structural information. To address this issue, we propose a clustering algorithm involving the following concepts:

- Projection of Positive Semi-Definite Matrix
- Distance Metric Based on Eigenstructure of Positive Semi-Definite Matrices
- Self-Consistency Clustering Algorithms

This innovative approach to clustering positive semi-definite matrices has broad applications in several domains, including financial and biomedical research, such as analyzing functional connectivity data. By maintaining the structural information of positive semi-definite matrices, our proposed algorithm promises to cluster the positive semi-definite matrices in a more meaningful way, thereby facilitating deeper insights into the underlying data in various applications.

Preliminaries: Self-Consistency and Self-Consistency Algorithm

Hastie and Stuetzle [1989] introduced a self-consistent curve or principal curve to provide a curve summary of the data. Let $\boldsymbol{X} \in \Re^p$ be a random vector with density h and finite second moments assuming $\mathcal{E}(\boldsymbol{X}) = 0$. Let $\mathbf{f}$ denote a smooth C^∞ unit-speed curve in $\Re^p$. the projection index $\lambda_{\mathbf{f}} : \Re^p \to \Re^1$ is defined as:

$$\lambda_{\mathbf{f}}(\mathbf{x}) = \sup_{\lambda} \left\{ \lambda : \|\mathbf{x} - \mathbf{f}(\lambda)\| = \inf_{\mu} \|\mathbf{x} - \mathbf{f}(\mu)\| \right\}.$$

The projection index $\lambda_{\mathbf{f}}(\mathbf{x})$ of $\mathbf{x}$ is the value of λ for which $\mathbf{f}(\lambda)$ is closest to $\mathbf{x}$. Then $\mathbf{f}$ is called self-consistent or principal curve of h if $\mathcal{E}(\mathbf{X}|\lambda_{\mathbf{f}}(\mathbf{X})) = \mathbf{f}(\lambda)$ for a.e. λ.

Tarpey [1999] presented the self-consistency algorithm, which can be viewed as a generalization of the K-means algorithm. Let $\mathcal{S} \subset \Re^p$ be a measurable set and define the *domain of attraction* of a point $\mathbf{y} \in \mathcal{S}$, denoted by $\mathcal{D}_{\mathbf{y}}(\mathcal{S})$:

$$\mathcal{D}_{\mathbf{y}}(\mathcal{S}) := \left\{ \mathbf{x} \in \Re^p : \|\mathbf{x} - \mathbf{y}\| < \|\mathbf{x} - \mathbf{z}\|, \mathbf{z} \in \mathcal{S}, \mathbf{z} \neq \mathbf{y} \right\}.$$

This set represents the *domain of attraction* of $\mathbf{y}$ towards the points in $\mathcal{S}$, containing all the points in $\mathcal{S}$ that is closer to $\mathbf{y}$ than to any other point $\mathbf{z}$ in $\mathcal{S}$.

Preliminaries: Common Principal Components

Flury [1984] proposed the concept of common principal components as an extension to principal components analysis. This approach assumes that n groups share the same principal component axes, This method can be formulated as an optimization problem:

$$\underset{\mathbf{B}}{\text{minimize}} \quad \prod_{i=1}^{n} \left(\frac{\det\left(\operatorname{diag}\left(\mathbf{B}^T \boldsymbol{\Psi}_i \mathbf{B}\right)\right)}{\det\left(\mathbf{B}^T \boldsymbol{\Psi}_i \mathbf{B}\right)} \right)$$

$$\text{subject to} \quad \mathbf{B}^T\mathbf{B} = \mathbf{I},$$

Where $\boldsymbol{\Psi}_i$ is covariance matrix of each subpopulation. Different approaches for estimating the common principal components have been proposed by Flury and Gautschi [1986], Vollgraf and Obermayer [2006], and Hallin et al. [2014]. These methods use maximum likelihood estimation (MLE) and S-estimation to estimate the common principal components from the positive semi-definite matrices.

Some Notation

- $\mathcal{V}_q(\Re^p) = \{\mathcal{X} \in \Re^{p\times q} : \mathcal{X}^T\mathcal{X} = \mathbf{I}_q\}$: the set of all orthonormal q-frames in $\Re^p$
- $\mathbf{S}_+^p = \{\mathcal{X} \in \Re^{p\times p} | \mathcal{X} = \mathcal{X}^T, \mathcal{X} \succeq 0\}$: the set of all positive semi-definite matrices in $\Re^{p\times p}$
- $\mathscr{D}_+^p = \{\mathcal{X} \in \Re^{p\times p} | \mathcal{X} = (\mathbf{a}\mathbf{1}^T) \circ \mathbf{I}_p, \mathbf{a} \in \Re^p, \mathbf{a} \succeq 0\}$ the set of all diagonal matrices in $\Re^{p\times p}$ with only non-negative elements

Here, $\mathbf{I}$ is the identity matrix, $\mathbf{1}$ is the vector with all elements equal to 1, and $\circ$ represents Hadamard product.

Projections, Principal Positive Semi-Definite Tensors, and Principal Positive Semi-Definite Matrices

We assume that there exists a random positive semi-definite matrix $\boldsymbol{\Psi} \in \mathbf{S}_+^p$, with a probability density function denoted by $\mathbf{f}$. Additionally, we consider a p-frame orthonormal matrix $\mathbf{B} \in \mathcal{V}_p(\Re^p)$ in $\Re^p$ and define the projection of the random matrix $\boldsymbol{\Psi}$ onto $\mathbf{B}$ as follows:

$$\mathcal{P}_{\mathbf{B}}(\boldsymbol{\Psi}) = \mathbf{B}\boldsymbol{\Lambda}_{\mathbf{B}}(\boldsymbol{\Psi})\mathbf{B}^T,$$

where $\boldsymbol{\Lambda}_{\mathbf{B}}(\boldsymbol{\Psi}) = (\mathbf{B}^T\boldsymbol{\Psi}\mathbf{B}) \circ \mathbf{I} \in \mathscr{D}_+^p$ is a diagonal matrix that depends on the random matrix $\boldsymbol{\Psi}$, given a fixed $\mathbf{B}$. This projection allows us to determine the proportion of the random positive semi-definite matrix $\boldsymbol{\Psi}$ that can be explained by the orthonormal frame $\mathbf{B}$.

Domain of Attraction to an Orthonormal Basis

Let $\mathcal{A} \subset \mathbf{S}_+^p$ be a subset of all positive semi-definite matrices. We define $\mathcal{D}_{\mathbf{B}}(\mathcal{A})$ the *domain of attraction* of $\mathbf{B}$ with respect to the subset of positive semi-definite matrices $\mathcal{A}$ as follow:

$$\mathcal{D}_{\mathbf{B}}(\mathcal{A}) := \left\{ \boldsymbol{\Psi} \in \mathbf{S}_+^p, : \|\boldsymbol{\Psi} - \mathcal{P}_{\mathbf{B}}(\boldsymbol{\Psi})\|_F^2 \leq \inf_{\mathbf{A}} \|\boldsymbol{\Psi} - \mathcal{P}_{\mathbf{A}}(\boldsymbol{\Psi})\|_F^2, \mathbf{A} \neq \mathbf{B}, \mathbf{A} \in \mathcal{V}_p(\Re^p) \right\},$$

where $\|\cdot\|_F^2$ is the squared Frobenius norm. The *domain of attraction* toward an orthonormal basis matrix $\mathbf{B}$ is defined as the matrices that can be better diagonalized by orthonormal matrix $\mathbf{B}$ compared to any other orthonormal matrix $\mathbf{A}$. $\mathcal{P}_{\mathbf{B}}(\boldsymbol{\Psi})$ is another representation of principal or self-consistent positive semi-definite tensors. We are able to identify the *domain of attraction* of $\mathbf{B}$ by analyzing the differences between $\boldsymbol{\Psi}$ and its corresponding slice on the principal positive semi-definite tensor.

K-Tensors: Algorithm for Clustering Positive Semi-Definite Matrices

Algorithm 1: K-Tensors: Clustering Positive Semi-Definite Matrices

Set i $= 0$.
Start with an initial K partition of the data: $\mathcal{D}_{\mathbf{B}_k^0}(\mathcal{A})$
while $i > 1$ *and* ***Loss***$^i \neq$ ***Loss***$^{i-1}$ **do**
  for $1 \leq k \leq K$ **do**
    estimate common principal components for each group and update $\mathbf{B}_{k^i}$ by

$$\mathbf{B}_{k^i}^* = \sup_{\|\mathbf{B}_k\|_F^2} \left\{ \mathbf{B}_k : \left\|\boldsymbol{\Psi} - \mathcal{P}_{\mathbf{B}_k}(\boldsymbol{\Psi})\right\|_F^2 = \inf_{\mathbf{B}_k} \left\|\boldsymbol{\Psi} - \mathcal{P}_{\mathbf{B}_r}(\boldsymbol{\Psi})\right\|_F^2 \middle| \mathbf{B}_r \in \mathcal{V}_p(\Re^p), \mathbf{B}_k \neq \mathbf{B}_r \right\}$$

    obtain the new assignment for each observation and update $\mathcal{D}_{\mathbf{B}_{k^i}}(\mathcal{A})$ by

$$\mathcal{D}_{\mathbf{B}_{k^i}} = \left\{ \boldsymbol{\Psi} \in \mathbf{S}_+^p, : \left\|\boldsymbol{\Psi} - \mathcal{P}_{\mathbf{B}_k}(\boldsymbol{\Psi})\right\|_F^2 \leq \inf_{\mathbf{B}_r} \left\|\boldsymbol{\Psi} - \mathcal{P}_{\mathbf{B}_r}(\boldsymbol{\Psi})\right\|_F^2, \mathbf{B}_r \neq \mathbf{B}_k, \mathbf{B}_r \in \mathcal{V}_p(\Re^p) \right\}$$

    calculate the loss of this iteration by **Loss**$^i = \sum_{i=1}^n \sum_{k=1}^K \|\boldsymbol{\Psi}_i - \mathcal{P}_{\mathbf{B}_{k^i}^*} \mathbb{I}(i \in k)\|_F^2$
  end
end

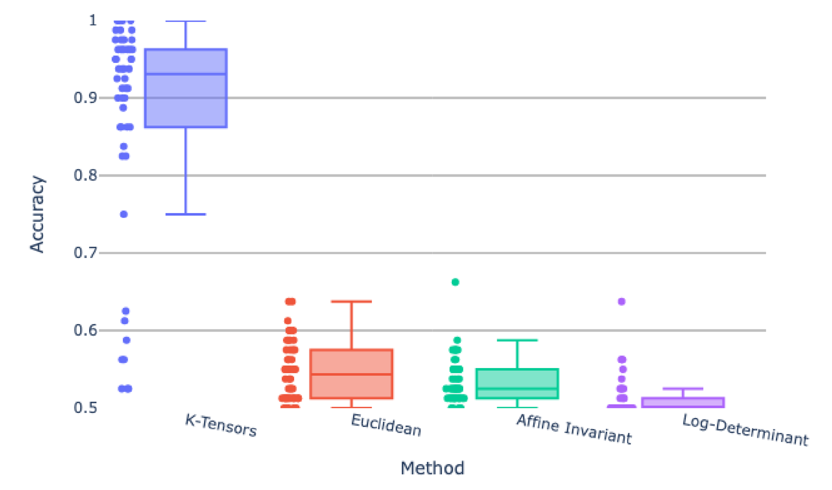
Definition: Principal Semi-Positive Tensors

Define a mapping from a diagonal matrix to a positive semi-definite matrix for a given $\mathbf{B} \in \mathcal{V}_p(\Re^p)$: $\mathcal{U}_{\mathbf{B}}(\boldsymbol{\Lambda}) = \mathbf{B}\boldsymbol{\Lambda}\mathbf{B}^T : \mathscr{D}_+^p \to \mathbf{S}_+^p$. We call $\mathcal{U}_{\mathbf{B}}(\boldsymbol{\Lambda})$ the principal, or self-consistency positive semi-definite tensors of $\mathbf{f}$ if $\mathcal{U}_{\mathbf{B}}(\boldsymbol{\Lambda}) = \mathcal{E}(\boldsymbol{\Psi}|\mathbf{B}(\boldsymbol{\Psi}) = \mathbf{B}, \boldsymbol{\Lambda}(\boldsymbol{\Psi}) = \boldsymbol{\Lambda})$ for a.e. $\boldsymbol{\Lambda}$.

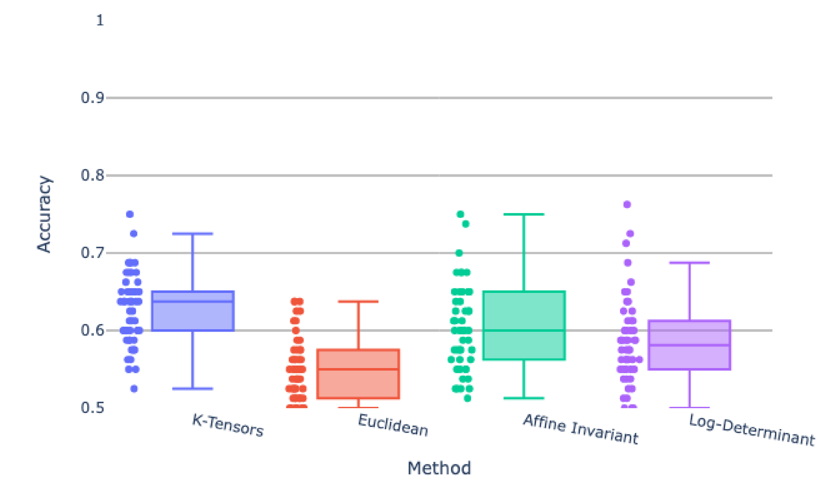
Simulation Studies

We evaluate the performance of our K-tensors algorithm in two simulation settings. In the first setting, we follow the structure proposed in Cook and Forzani [2008], where each functional connectivity matrix $\boldsymbol{\Psi}_i$ is modeled as $\boldsymbol{\Psi}_{i\in C_k} = \mathbf{U}_k\boldsymbol{\Lambda}_i\mathbf{U}_k^T + \mathbf{E}_i$. Here, $\boldsymbol{\Psi}_i$ and $\mathbf{E}_i$ are positive semi-definite matrices, $\boldsymbol{\Lambda}_i$ is a diagonal matrix, and $\mathbf{U}_k$ is an orthonormal matrix representing the latent subpopulations.

In the second simulation setting, we consider the Wishart distribution with degree of freedom from 10 to 45. In both settings, we assume 2 underlying true clusters, with each cluster consisting of 50 observations.



(a) Cook and Forzani [2008] Settings
$\mathcal{E}\|\mathbf{E}_i\|$ from 0.1 to 0.6



(b) Wishart Distribution Settings
df from 10 to 45

Acknowledgement

We would like to thanks Dr. Alessandro S. De Nadai (Harvard Medical School) and Dr. Emily R. Stern (NYU Grossman School of Medicine) for their valuable input on this project.

This work is supported by NIMH grants: R01 MH099003, R01 MH126981, R01 MH111794, R01 MH111794. and R61/R33 MH107589.

References

H. Abdi and L. J. Williams. Principal component analysis. *Wiley interdisciplinary reviews: computational statistics*, 2(4):433–459, 2010.

T. W. Anderson. Asymptotically efficient estimation of covariance matrices with linear structure. *The Annals of Statistics*, 1(1):135–141, 1973.

J. D. Banfield and A. E. Raftery. Model-based Gaussian and non-Gaussian clustering. *Biometrics*, pages 803–821, 1993.

D. A. Binder. Bayesian cluster analysis. *Biometrika*, 65(1):31–38, 1978.

H.-H. Bock. Convexity-based clustering criteria: theory, algorithms, and applications in statistics. *Statistical Methods & Applications*, 12: 293–317, 2003.

H.-H. Bock. Convexity-based clustering criteria: theory, algorithms, and applications in statistics. *Statistical Methods and Applications*, 12(3): 293–317, 2004.

R. P. Browne and P. D. McNicholas. Estimating common principal components in high dimensions. *Advances in Data Analysis and Classification*, 8:217–226, 2014.

E. C. Chi, G. I. Allen, and R. G. Baraniuk. Convex biclustering. *Biometrics*, 73(1):10–19, 2017.

P. Comon. Independent component analysis, a new concept? *Signal processing*, 36(3):287–314, 1994.

R. D. Cook and L. Forzani. Covariance reducing models: An alternative to spectral modelling of covariance matrices. *Biometrika*, 95(4): 799–812, 2008.

F. Edition et al. Diagnostic and statistical manual of mental disorders. *Am Psychiatric Assoc*, 21:591–643, 2013.

B. Flury. *A first course in multivariate statistics*. Springer Science & Business Media, 2013.

B. K. Flury. Two generalizations of the common principal component model. *Biometrika*, 74(1):59–69, 1987.

B. N. Flury. Common principal components in k groups. *Journal of the American Statistical Association*, 79(388):892–898, 1984.

B. N. Flury and W. Gautschi. An algorithm for simultaneous orthogonal transformation of several positive definite symmetric matrices to nearly diagonal form. *SIAM Journal on Scientific and Statistical Computing*, 7(1):169–184, 1986.

A. M. Franks and P. Hoff. Shared subspace models for multi-group covariance estimation. *J. Mach. Learn. Res.*, 20:171–1, 2019.