Spectral method for Gaussian tensor matching with Tucker decomposition

Jiaxin Hu

May 30, 2022

This note aims to extend the spectral method in Fan et al. (2019) to the tensor case. Previous note uses CP decomposition which is not suitable and we use Tucker decomposition in this note. We use red color to mark the problems and then discuss the problem in the end of each section.

1 Preliminary

1.1 Notation

- For the index $\boldsymbol{\omega} = (\omega_1, \ldots, \omega_m) \in [n]^m$ and permutation π on $[n]$, let $\pi \circ \boldsymbol{\omega} = (\pi(\omega_1), \ldots, \pi(\omega_m))$ denote the permuted index.
- Let $\mathcal{P}_n$ be the collection of all possible permutations on $[n]$.
- For a matrix $\boldsymbol{A} \in \mathbb{R}^{n \times n}$ and $r \in [n]$, let $\mathrm{SVD}_r(\boldsymbol{A}) \in \mathbb{R}^{n \times r}$ denote the top r left singular vectors corresponding to the largest r singular values.
- For a tensor $\mathcal{A} \in \mathbb{R}^{n_1 \times \cdots \times n_m}$ and $k \in [m]$, let $\mathrm{Mat}_k(\mathcal{A}) \in \mathbb{R}^{n_k \times \prod_{i \neq k} n_i}$ denote the mode-k matrcization of $\mathcal{A}$.

1.2 Model

Consider the order-m tensors $\mathcal{A} \in \mathbb{R}^{n^{\otimes m}}$ and $\mathcal{B} \in\in \mathbb{R}^{n^{\otimes m}}$, and the true permutation $\pi^* : [n] \mapsto [n]$ with corresponding permutation matrix $\Pi^* \in \mathbb{R}^{n \times n}$. The tensors $\mathcal{A}$ and $\mathcal{B}$ follow the super-symmetric m-d Gaussian Correlated model with parameters (π^*, ρ^*) for $\rho^* \in [0,1]$; i.e., for all indices $\boldsymbol{\omega} = (\omega_1, \ldots, \omega_m)$ such that $1 \leq \omega_1 \leq \cdots \leq \omega_m \leq n$, we assume

$$\mathcal{A}_{\boldsymbol{\omega}} \sim_{i.i.d.} N(0, 1/n)\text{(Q1)}, \quad \mathcal{A}_{\omega_{\pi'(1)}, \ldots, \omega_{\pi'(m)}} = \mathcal{A}_{\boldsymbol{\omega}} \text{ for all } \pi' \in \mathcal{P}_m, \quad \mathcal{B} = \mathcal{A} \times_1 \Pi^* \times_2 \cdots \times_m \Pi^* + \sigma \mathcal{Z},$$

where $\sigma = \frac{\sqrt{1-\rho^*}}{\rho^*}$, $\mathcal{Z}_{\boldsymbol{\omega}} \sim_{i.i.d.} N(0, 1/n)$(Q1), $\mathcal{A}_{\boldsymbol{\omega}} \perp \mathcal{Z}_{\boldsymbol{\omega}}$. Parameter ρ^* is the correlation coefficient between variables $\mathcal{A}_{\boldsymbol{\omega}}$ and $\mathcal{B}_{\pi^* \circ \boldsymbol{\omega}}$.

Question 1 (Gaussian orthogonal ensemble for tensor? (Q1) **).** In Fan et al. (2019), they assume the random matrix $\boldsymbol{A} \in \mathbb{R}^{n \times n}$ follows the Gaussian Orthogonal Ensemble (GOE) where

$\boldsymbol{A}_{ij} \sim N(0, 1/n)$ for $i < j$ and $\boldsymbol{A}_{ii} = N(0, 2/n)$. The GOE assumption provides critical theoretical results for the spectral analysis on the random matrix $\boldsymbol{A}$: the coordinates of the eigenvectors are approximately independent and follow $N(0, 1/n)$, and the eigenvectors are distributed independently with eigenvalues. See Proposition 3.1 in Fan et al. (2019).

One possible extension of GOE to tensor case is provided in de Morais Goulart et al. (2021). The variances of the entries are related to the repetitive indices in the index vector $\boldsymbol{\omega}$, and the situation goes complicate as order m increases. The tensor GOE and random tensor theory should be the key component of our proof.

2 Spectral algorithm via Tucker decomposition (Q2)

We extend the spectral method in Fan et al. (2019) by considering Tucker decomposition of tensor observations. Detailed procedures are in Algorithm 1.

Algorithm 1 Spectral matching via Tucker decomposition

Input: Gaussian tensors $\mathcal{A}, \mathcal{B} \in \mathbb{R}^{n^{\otimes m}}$, bandwidth parameter η

1: Calculate the singular space of $\mathcal{A}$ and $\mathcal{B}$

$$\boldsymbol{U} = \mathrm{SVD}_n(\mathrm{Mat}_1(\mathcal{A})), \quad \boldsymbol{V} = \mathrm{SVD}_n(\mathrm{Mat}_1(\mathcal{B})), (Q4, Q5)$$

and the core tensors

$$\mathcal{C} = \mathcal{A} \times_1 \boldsymbol{U}^T \times_2 \cdots \times_m \boldsymbol{U}^T, \quad \mathcal{D} = \mathcal{B} \times_1 \boldsymbol{V}^T \times_2 \cdots \times_m \boldsymbol{V}^T.$$

2: Let $\mathcal{C}_{i:}, \mathcal{D}_{i:} \in \mathbb{R}^{n^{\otimes m-1}}$ be the i-th first mode slices of tensors $\mathcal{C}, \mathcal{D}$, and $\boldsymbol{U}_i, \boldsymbol{V}_i \in \mathbb{R}^n$ denote the i-th columns of matrices $\boldsymbol{U}, \boldsymbol{V}$, for all $i \in [n]$. Construct the similarity matrix

$$\boldsymbol{S} = \sum_{i,j \in [n]} \frac{1}{(\|\mathcal{C}_{i:}\|_F - \|\mathcal{D}_{j:}\|_F)^2 + \eta^2} \boldsymbol{U}_i \boldsymbol{U}_i^T \boldsymbol{J} \boldsymbol{V}_j \boldsymbol{V}_j^T, (Q3)$$

where $\|\cdot\|$ is the tensor Frobenius norm and $\boldsymbol{J} \in \mathbb{R}^{n \times n}$ is the all-one matrix.

3: Find the estimate permutation by solving the linear assignment problem (LAP) with $\boldsymbol{S}$

$$\hat{\pi} = \arg\max_{\pi \mathcal{P}_n} \sum_{k \in [n]} \boldsymbol{S}_{k,\pi(k)}.$$

Output: Estimated permutation $\hat{\pi}$.

Question 2 (Why we use Tucker rather than CP decomposition? (Q2))**.** Given a tensor, it is hard to determine the CP rank of the tensor Kolda and Bader (2009) and the best low rank approximation is provably shown to be an ill-posed problem De Silva and Lim (2008). Therefore, we use Tucker decomposition which does not suffer from the approximation issue and is closely related to matrix SVD. In particularly, we consider the Tucker decomposition $\mathcal{A} = \mathcal{C} \times_1 \boldsymbol{U} \times_2 \cdots \times_m \boldsymbol{U}$ in which $\boldsymbol{U}$ is the left singular matrix in the SVD of the first mode matricization $\mathcal{M}_1(\mathcal{A}) \in \mathbb{R}^{n \times n^{m-1}}$. The matrix $\boldsymbol{U}$ can be uniquely determined upto sign and permutation once the $\mathcal{M}_1(\mathcal{A})$ has distinct singular values. Therefore, Tucker decomposition enjoys the same uniqueness property as matrix

SVD. The column correspondence between $\boldsymbol{U}_i$ and $\boldsymbol{V}_j$ can be determined by the singular values of $\mathcal{M}_1(\mathcal{A}), \mathcal{M}_1(\mathcal{B})$, which are equal to the slice Frobenius norms $\|\mathcal{C}_{i:}\|_F$ and $\|\mathcal{D}_{j:}\|_F$.

Question 3 (Can we avoid the projection $\boldsymbol{U}_i\boldsymbol{U}_i^T$? (Q3))**.** In Fan et al. (2019), they consider the projection $\boldsymbol{U}_i\boldsymbol{U}_i^T$ and $\boldsymbol{V}_j\boldsymbol{V}_j^T$ to avoid the multiplicity and sign issues. Notice that Tucker decomposition also has the same multiplicity and sign issues as matrix SVD. We keep using projection in this step.

Question 4 (Is $\boldsymbol{S}$ invariant to the rotation of $\boldsymbol{U}, \boldsymbol{V}$? (Q4))**.** No, the $\boldsymbol{S}$ is not invariant to the orthogonal rotation of $\boldsymbol{U}, \boldsymbol{V}$. The high-level idea of spectral method is to compare the weights (or loadings) of n nodes on the same singular space. The weights are the entries of the singular vectors, and the success of spectral algorithm relies on the correct correspondence of the singular vectors $\boldsymbol{U}_i$ and $\boldsymbol{V}_j$. Unlike singular space which is robust to rotation, the pairwise comparison of the singular vectors is sensitive to rotation. Consider the following counterexample.

Example 1 (Counterexample for rotation invariance.)**.** Consider the noiseless case $\sigma = 0, n = 2, m = 2$, and true permutation π^* is identity mapping. Let

$$\boldsymbol{A} = \boldsymbol{U}\boldsymbol{C}\boldsymbol{U}^T = \boldsymbol{U}'\boldsymbol{C}'\boldsymbol{U}^T, \quad \text{where } \boldsymbol{U} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \boldsymbol{C} = \begin{pmatrix} 2 & 0 \\ 0 & 1 \end{pmatrix}, \boldsymbol{U}' = \begin{pmatrix} \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} & -\frac{\sqrt{2}}{2} \end{pmatrix}, \boldsymbol{C}' = \begin{pmatrix} \sqrt{2} & \frac{\sqrt{2}}{2} \\ \sqrt{2} & -\frac{\sqrt{2}}{2} \end{pmatrix}.$$

In our case, the $\boldsymbol{S}$ with $\boldsymbol{U}, \mathcal{C}, \boldsymbol{V}, \mathcal{D}$ in Line 2 is equal to

$$\begin{aligned} \boldsymbol{S}(\boldsymbol{U}, \boldsymbol{C}, \boldsymbol{U}, \boldsymbol{C}) &= \sum_{i,j \in [n]} \frac{1}{(\|\boldsymbol{C}_{i:}\|_F - \|\boldsymbol{C}_{j:}\|_F)^2 + \eta^2} \boldsymbol{U}_i\boldsymbol{U}_i^T \boldsymbol{J} \boldsymbol{U}_j \boldsymbol{U}_j^T, \\ &= \frac{1}{\eta^2}\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} + \frac{1}{1+\eta^2}\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}. \end{aligned}$$

With the rotated $\boldsymbol{U}'$ and $\boldsymbol{C}'$, the matrix $\boldsymbol{S}$ becomes

$$\begin{aligned} \boldsymbol{S}(\boldsymbol{U}', \boldsymbol{C}', \boldsymbol{U}, \boldsymbol{C}) &= \sum_{i,j \in [n]} \frac{1}{(\|\boldsymbol{C}'_{i:}\|_F - \|\boldsymbol{C}_{j:}\|_F)^2 + \eta^2} \boldsymbol{U}'_i(\boldsymbol{U}'_i)^T \boldsymbol{J} \boldsymbol{U}_j \boldsymbol{U}_j^T, \\ &= \frac{\sqrt{2}}{(\sqrt{5/2}-2)^2 + \eta^2}\begin{pmatrix} \frac{\sqrt{2}}{2} & 0 \\ \frac{\sqrt{2}}{2} & 0 \end{pmatrix} + \frac{\sqrt{2}}{(\sqrt{5/2}-1)^2 + \eta^2}\begin{pmatrix} 0 & \frac{\sqrt{2}}{2} \\ 0 & \frac{\sqrt{2}}{2} \end{pmatrix}. \end{aligned}$$

The matrix $\boldsymbol{S}(\boldsymbol{U}, \boldsymbol{C}, \boldsymbol{U}, \boldsymbol{C}) \neq \boldsymbol{S}(\boldsymbol{U}', \boldsymbol{C}', \boldsymbol{U}, \boldsymbol{C})$. The matrix $\boldsymbol{S}(\boldsymbol{U}', \boldsymbol{C}', \boldsymbol{U}, \boldsymbol{C})$ has dominate elements in the first column, and the implementing the LAP on $\boldsymbol{S}(\boldsymbol{U}', \boldsymbol{C}', \boldsymbol{U}, \boldsymbol{C})$ can not reveal the true permutation.

In contrast, the $\boldsymbol{S}$ is invariant the permutation of its columns because we consider the sum of all possible $i, j \in [n]$.

Question 5 (Sanity check of spectral algorithm. (Q5))**.** Consider the noiseless case $\sigma = 0$ and true permutation π^* is identity mapping. We have

$$\mathrm{Mat}_1(\mathcal{B}) = \mathrm{Mat}_1(\mathcal{A}), \text{ and thus } \boldsymbol{V} = \boldsymbol{U}, \mathcal{D} = \mathcal{C}.$$

With a proper choice of η, the weights of the sum are dominated by the term $\frac{1}{(\|\mathcal{C}_{i:}\|_F - \|\mathcal{D}_{j:}\|_F)^2 + \eta^2}$ over than the term $\boldsymbol{U}_i^T \boldsymbol{J} \boldsymbol{V}_j$. Therefore, we have

$$\boldsymbol{S} \approx \sum_{i \in [n]} \frac{1}{\eta^2} \boldsymbol{U}_i \boldsymbol{U}_i^T,$$

and $\pi^* = \arg\max_{\pi \mathcal{P}_n} \sum_{k \in [n]} \boldsymbol{S}_{k,\pi(k)}$.

3 Proof sketch and difficulties

Recall the similarity matrix constructed by $\mathcal{A}$ and $\mathcal{B}$

$$\boldsymbol{S} = \sum_{i,j\in[n]} \frac{1}{(\|\mathcal{C}_{i:}\|_F - \|\mathcal{D}_{j:}\|_F)^2 + \eta^2} \boldsymbol{U}_i \boldsymbol{U}_i^T \boldsymbol{J} \boldsymbol{V}_j \boldsymbol{V}_j^T,$$

where $\boldsymbol{J} \in \mathbb{R}^{n\times n}$ is the all-one matrix. We now consider the similarity matrix replacing $\mathcal{B}$ by $\mathcal{A}$,

$$\boldsymbol{S}^* = \sum_{i,j\in[n]} \frac{1}{(\|\mathcal{C}_{i:}\|_F - \|\mathcal{C}_{j:}\|_F)^2 + \eta^2} \boldsymbol{U}_i \boldsymbol{U}_i^T \boldsymbol{J} \boldsymbol{U}_j \boldsymbol{U}_j^T,$$

To recover the true permutation by solving the linear assignment problem with $\boldsymbol{S}$, we need to show following lemmas.

Lemma 1 (Dominance of diagonal elements in $\boldsymbol{S}^*$)**.** *Under a proper range for the parameters η, with high probability*

$$\min_{i\in[n]} \boldsymbol{S}^*_{ii} > C_1(n, m, \eta), \quad \max_{i\neq j\in[n]} \boldsymbol{S}^*_{ij} < C_2(n, m, \eta),$$

where $C_1(n, m, \eta), C_2(n, m, \eta)$ are scalars related to parameters n, m, η.

Lemma 2 (Estimation and approximation errors)**.** *Under a proper range for the parameters η, with high probability*

$$\max_{i,j\in[n]} |\boldsymbol{S}^*_{ij} - \boldsymbol{S}_{ij}| < C_3(n, m, \eta, \sigma),$$

where $C_3(n, m, \eta, \sigma)$ is a scalar related to parameters n, m, η and noise level σ.

We then need to find the range of noise level to make the estimation and approximation error negligible compared with the permutation error; i.e., find σ such that

$$2C_3(n, m, \eta, \sigma) < C_1(n, m, \eta) - C_2(n, m, \eta).$$

Then, we are able to show the exact recovery of the Algorithm 1.

References

de Morais Goulart, J. H., Couillet, R., and Comon, P. (2021). A random matrix perspective on random tensors. *stat*, 1050:2.

De Silva, V. and Lim, L.-H. (2008). Tensor rank and the ill-posedness of the best low-rank approximation problem. *SIAM Journal on Matrix Analysis and Applications*, 30(3):1084–1127.

Fan, Z., Mao, C., Wu, Y., and Xu, J. (2019). Spectral graph matching and regularized quadratic relaxations i: The gaussian model. *arXiv preprint arXiv:1907.08880.*

Kolda, T. G. and Bader, B. W. (2009). Tensor decompositions and applications. *SIAM Review*, 51(3):455–500.