Spectral method for Gaussian tensor matching
Possible strategies and problems

Jiaxin Hu

May 27, 2022

This note aims to extend the spectral method in Fan et al. (2019) to the tensor case. We provide a possible algorithm and a high-level proof sketch. 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 $[n]$.
- Let $\boldsymbol{v}^{\otimes m}$ denote the outer product of m vectors $\boldsymbol{v} \in \mathbb{R}^n$; i.e., $\boldsymbol{v}^{\otimes m} = \boldsymbol{v} \otimes \cdots \otimes \boldsymbol{v} \in \mathbb{R}^{n^{\otimes m}}$.
- Let $\mathcal{C}_{n,m,r} \subset \mathbb{R}^{n^{\otimes m}}$ denote the collection of all symmetric orthogonally decomposable (odeco) tensors with dimension n, order m, and CP rank r. For an arbitrary $\mathcal{X} \in \mathcal{C}_{n,r}$, we have

$$\mathcal{X} = \sum_{i=1}^{r} \lambda_i \boldsymbol{v}_i^{\otimes m}, \quad \|\boldsymbol{v}_i\|_F = 1, \quad \boldsymbol{v}_i^T \boldsymbol{v}_j = 0 \text{ for all } i, j \in [r], i \neq j.$$

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).

In tensor case, I did not find many related literature about the spectral analysis for random tensors. 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. I also did not find the reference about the distribution of the decomposition component for tensor cases, which should be critical for the proof. More discussions about the decomposition are included in following questions.

2 Spectral algorithm via CP decomposition (Q5)

A direct extension of Fan et al. (2019) is to consider the CP tensor decomposition. Detailed procedures are in Algorithm 1

Algorithm 1 Spectral matching via CP decomposition

Input: Gaussian tensors $\mathcal{A}, \mathcal{B} \in \mathbb{R}^{n^{\otimes m}}$, rank of decomposition R (Q2) , bandwidth parameter η

1: Calculate the best rank-R symmetric odeco of $\mathcal{A}$ and $\mathcal{B}$

$$\hat{\mathcal{A}} = \underset{\mathcal{X}\in\mathcal{C}_{n,m,R}}{\arg\min} \|\mathcal{A}-\mathcal{X}\|_F^2, \quad \hat{\mathcal{B}} = \underset{\mathcal{X}\in\mathcal{C}_{n,m,R}}{\arg\min} \|\mathcal{B}-\mathcal{X}\|_F^2, (Q3, Q6)$$

where

$$\hat{\mathcal{A}} = \sum_{i=1}^{R} \lambda_i \boldsymbol{v}_i^{\otimes m}, \quad \hat{\mathcal{B}} = \sum_{i=1}^{R} \mu_i \boldsymbol{w}_i^{\otimes m}.$$

2: Construct the similarity matrix

$$\boldsymbol{S} = \sum_{i,j\in[R]} \frac{1}{(\lambda_i-\mu_j)^2+\eta^2} \boldsymbol{v}_i\boldsymbol{v}_i^T \boldsymbol{J} \boldsymbol{w}_j \boldsymbol{w}_j^T, (Q4)$$

where $\boldsymbol{J} \in \mathbb{R}^{n\times n}$ is the all-one matrix.

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

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

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

Question 2 (Why we need rank of decomposition as input?(Q2))**.** Given a tensor, it is hard to determine the CP rank of the tensor Kolda and Bader (2009). In matrix case, we can always implement the SVD with an n-by-n diagonal matrix and orthornomal singular vectors for any n-by-n matrix. In tensor case, the CP rank of a dimensional n tensor can exceed n. We only know the upper bound for CP rank is of order $\mathcal{O}(n^{m-1})$, which leads to computational inefficiency and many linear dependency among eigenvectors $\boldsymbol{v}_i$'s.

So, I set the rank of decomposition R as a tuning parameter currently. Because we do not which R is good enough for the CP decomposition to reveal most information in $\mathcal{A}$.

Question 3 (Should we really need odeco? Do other symmetric CP decompositions work? (Q3))**.** Two main reasons for me to use odeco now. First, the orthogonality among eigenvectors may bring technically convenience. In the proof of Fan et al. (2019), they take the advantages of the orthogonality among matrix eigenvectors. Second, there are some current results related to the signal estimation under the odeco setting; see Xia et al. (2022). In addition, the odeco tensors have CP rank at most n and then we can set $R = n$ to avoid the tuning of R. (To highlight the rank issue, I still keep R in the input even with odeco in Line 1.)

Question 4 (Can we avoid the projection $\boldsymbol{v}_i\boldsymbol{v}_i^T$? (Q4))**.** In Fan et al. (2019), they consider the projection $\boldsymbol{v}_i\boldsymbol{v}_i^T$ and $\boldsymbol{w}_j\boldsymbol{w}_j^T$ to avoid the multiplicity and sign issues. Though odeco tensor has a stronger uniqueness without multiplicity issue, we still have the sign issue. So, I keep using the projection for similarity matrix.

Question 5 (Can we use tucker decomposition than CP decomposition? (Q5))**.** The main reason for not using tucker decomposition is that tucker decomposition is not unique with orthogonal rotation while matrix SVD and CP decomposition are unique. With SVD and CP decomposition, we use the singular/eigenvalues to match the eigenvectors in two observations. The non-uniqueness of tucker decomposition makes it hard to find the correspondence of the eigenvectors in two observations and thereof hard to construct the similarity matrix $\boldsymbol{S}$.

There are also benefits to use tucker decomposition: tucker decomposition does not suffer the rank issue. Any tensor can be decomposed into the tucker form with a core tensor of dimension $\boldsymbol{r} = (n, \dots, n)$, and thus tucker decomposition avoids the approximation error occurs in the Line 1 of Algorithm 1.

Question 6 (What algorithm should we use to find the best approximation?(Q6))**.** Conditional on $\mathcal{A}$, the matching problem turns to a signal estimation problem with observation $\mathcal{B}$ and noise $\mathcal{Z}$. To achieve a good estimation performance, we may use the two-step power iteration algorithm as suggested by Xia et al. (2022), which shares the same spirit as the double-projection tensor SVD used in dTBM paper.

3 Proof sketch and difficulties

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

$$\boldsymbol{S} = \sum_{i,j \in [R]} \frac{1}{(\lambda_i - \mu_j)^2 + \eta^2} \boldsymbol{v}_i \boldsymbol{v}_i^T \boldsymbol{J} \boldsymbol{w}_j \boldsymbol{w}_j^T,$$

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

$$\boldsymbol{S}^* = \sum_{i,j \in [R]} \frac{1}{(\lambda_i - \lambda_j)^2 + \eta^2} \boldsymbol{v}_i \boldsymbol{v}_i^T \boldsymbol{J} \boldsymbol{v}_j \boldsymbol{v}_j^T.$$

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 R, η, with high probability*

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

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

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

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

where $C_3(n, m, R, \eta, \sigma)$ is a scalar related to parameters n, m, R, 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

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

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

Proof difficulties for Lemma 1. To prove Lemma 1, we need to know the (approximate) distributions of the decomposition component λ_i's and $\boldsymbol{v}_i$'s. Note that the decomposition component λ_i's and $\boldsymbol{v}_i$'s come from the approximation $\hat{\mathcal{A}}$ rather that $\mathcal{A}$, and $\hat{\mathcal{A}}$ may not have the Gaussian entries.

So we need the random tensor results for odeco and the approximation error of odeco for the random tensor. □

Proof difficulties for Lemma 2. The error between $\boldsymbol{S}$ and $\boldsymbol{S}^*$ should rely on the error between $\hat{\mathcal{A}}$ and $\hat{\mathcal{B}}$, which is upper bounded by the approximation errors between $\hat{\mathcal{A}}, \mathcal{A}$ and $\hat{\mathcal{B}}, \mathcal{B}$ and the estimation error between $\mathcal{A}$ and $\mathcal{B}$.

Again, we need the approximation error of odeco for the random tensor. In addition, current literature for estimation error are based on the low-rank tensors Xia et al. (2022); Zhang and Xia (2018), which may not be applicable for the $\mathcal{A}, \mathcal{B}$ in our case.

□

References

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

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.

Xia, D., Zhang, A. R., and Zhou, Y. (2022). Inference for low-rank tensors—no need to debias. *The Annals of Statistics*, 50(2):1220–1245.

Zhang, A. and Xia, D. (2018). Tensor SVD: Statistical and computational limits. *IEEE Transactions on Information Theory*, 64(11):7311–7338.