Estimating smooth tensors with unknown permutations

Chanwoo Lee [1] Miaoyan Wang [1]

[1]Department of Statistics, University of Wisconsin-Madison

Main Problem

We consider permuted tensor model,

$$\mathcal{Y} = \Theta \circ \sigma + \mathcal{E},$$

where $\mathcal{Y} \in \mathbb{R}^{d\times\cdots\times d}$ is an observed data tensor, Θ is an unknown smooth signal tensor, σ is an unknown permutation, and $\mathcal{E} \in \mathbb{R}^{d\times\cdots\times d}$ is a noise tensor consisting of zero-mean sub-Gaussian entries.

Main problem: how to estimate $\Theta \circ \sigma$ from the observed data tensor $\mathcal{Y}$?

Limitations of low-rank assumption

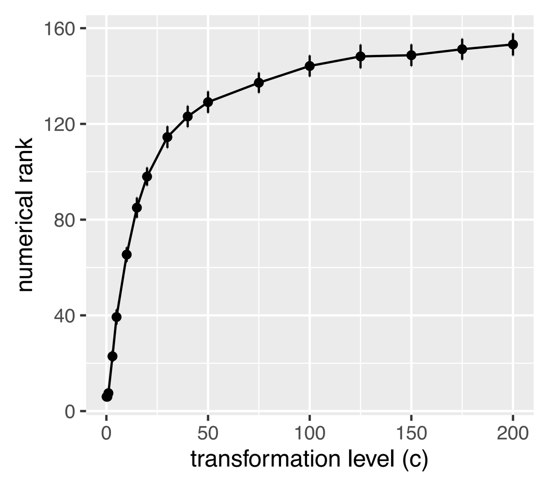
Low-rank models assume

$$\Theta \circ \sigma = \sum_{\ell=1}^{r} \lambda_\ell \boldsymbol{a}_1^{(\ell)} \otimes \cdots \otimes \boldsymbol{a}_m^{(\ell)},$$

where λ_ℓ is a scalar and $\boldsymbol{a}_k^{(\ell)} \in \mathbb{R}^d$ for all $(k,\ell) \in [m]\times[r]$.

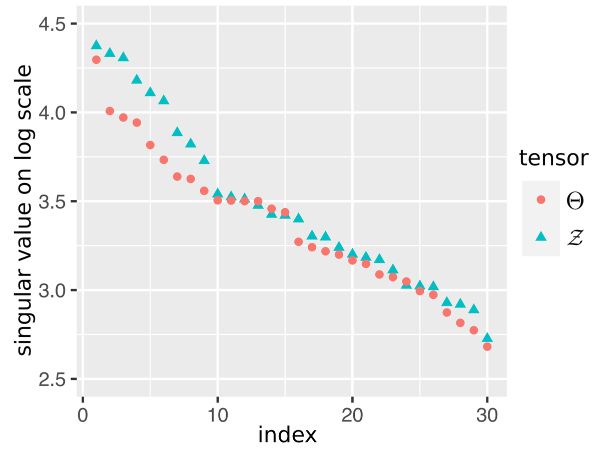
However, low rank models are

- Sensitive to order-preserving transformation

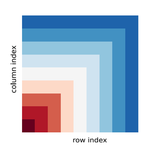


$$\Theta = \frac{1}{1+\exp(-c(\mathcal{Z}))}, \quad \text{where} \quad \mathcal{Z} = \boldsymbol{a}^{\otimes 3} + \boldsymbol{b}^{\otimes 3} + \boldsymbol{c}^{\otimes 3}.$$

- Inadequate for special structures.



$$\Theta = \log(1+\mathcal{Z}), \quad \text{where} \quad \mathcal{Z}(i,j,k) = \frac{1}{d}\max(i,j,k).$$



Our main assumption

Instead, we assume there exists $f\colon [0,1]^m \to \mathbb{R}$ that satisfies

- Representation : $\Theta(\omega) = f(\omega/d)$ for all $\omega \in [d]^m$,
- α-Hölder continuity : $|f(\boldsymbol{x}) - f(\boldsymbol{y})| \leq L\|\boldsymbol{x}-\boldsymbol{y}\|_\alpha^\alpha$ for all $\boldsymbol{x},\boldsymbol{y} \in [0,1]^m$, where the norm $\|\boldsymbol{x}\|_p^p := \sum_{i=1}^m |x_i|^p$ for $\boldsymbol{x} \in \mathbb{R}^m$.

Stochastic block approximation (SBA) to smooth tensor

Lemma (Block approximation). For any $k < d$, there exists k-membership function $z\colon [d] \to [k]$, and $\mathcal{G} \in \mathbb{R}^{k\times\cdots\times k}$ such that

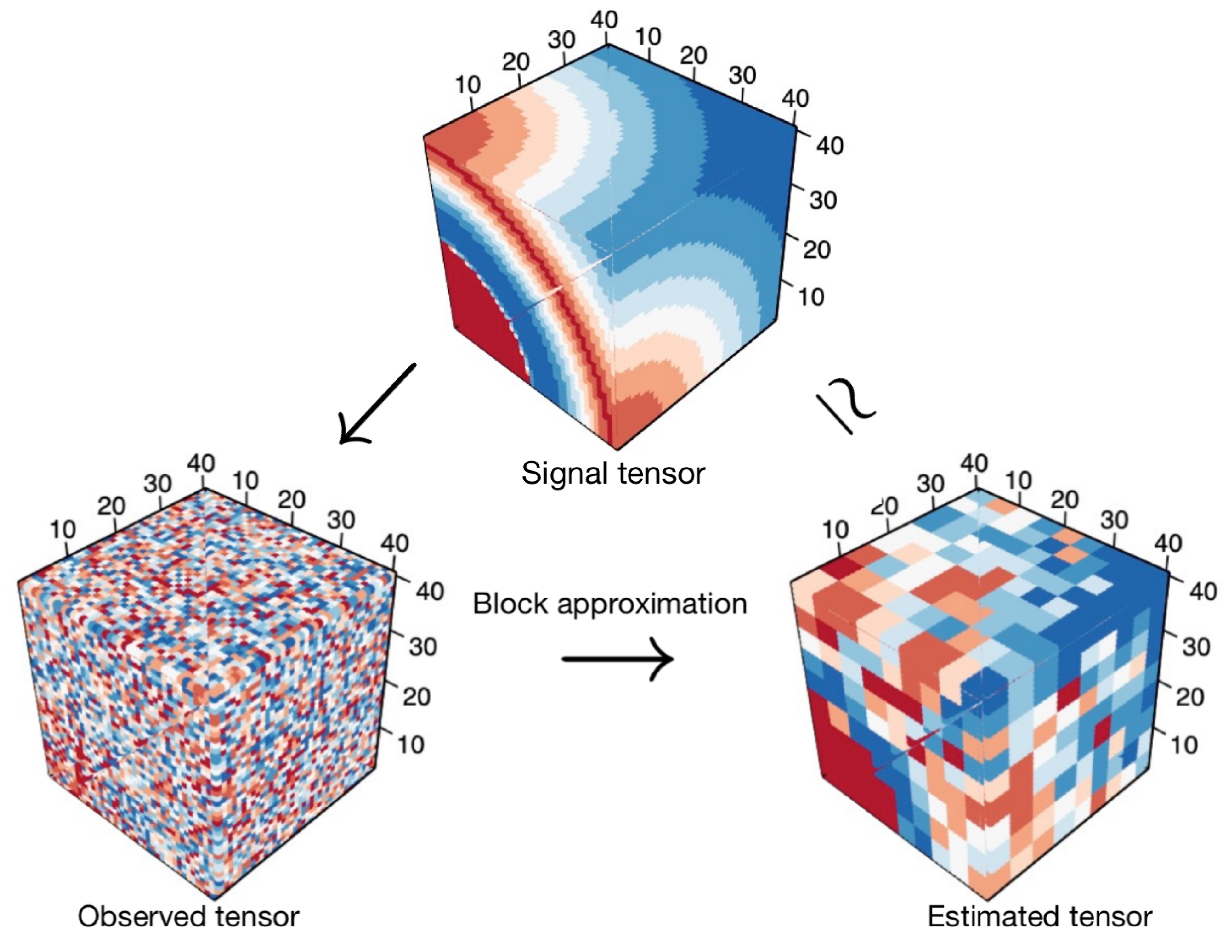
$$\frac{1}{d^m}\sum_{\omega\in[d]^m} |(\Theta\circ\sigma)(\omega) - \mathcal{G}(z(\omega))|^2 \leq \frac{m^2L^2}{k^{2\alpha}}.$$

Based on this lemma and algorithms in [2], we find the optimizer,

$$(\hat{z},\hat{\mathcal{G}}) = \mathop{\arg\min}_{z\colon[d]\to[k],\mathcal{G}\in\mathbb{R}^{k\times\cdots\times k}} \sum_{\omega\in[d]^m} |\mathcal{Y}(\omega) - \mathcal{G}(z(\omega))|^2. \quad (1)$$

We estimate the $\Theta\circ\sigma$ by

$$(\widehat{\Theta\circ\sigma})(\omega) = \hat{\mathcal{G}}(\hat{z}(\omega)), \text{ for all } \omega \in [d]^m. \quad (2)$$



Theoretical guarantees

Theorem (Mean square error). Let $\hat{\Theta}$ be the estimator from (2) with the choice of $k = \lceil d^{\frac{m}{m+2\alpha}} \rceil$. Then,

$$\frac{1}{d^m}\|\widehat{\Theta\circ\sigma} - \Theta\circ\sigma\|_F^2 \lesssim \underbrace{d^{-\frac{2m\alpha}{m+2\alpha}}}_{\text{Nonparametric rate}} + \underbrace{\frac{\log d}{d^{m-1}}}_{\text{Clustering rate}}, \quad (3)$$

with high probability.

Remark: Depending on constants m and α, convergence rate becomes

$$\text{RHS of (3)} \asymp \begin{cases} d^{-\frac{2\alpha}{1+\alpha}} & m=2, \alpha\in(0,1), \\ \log d/d & m=2, \alpha=1, \\ d^{-\frac{2m\alpha}{m+2\alpha}} & m>2. \end{cases}$$

Though SBA guarantees fast convergence rate, polynomial complexity algorithms for (1) are unknown.

Ongoing work: polynomial algorithms

Sort-And-Smooth (SAS) method extended from [1]

Under the monotonically increasing degree assumption on signal Θ,

$$\frac{1}{d^{m-1}}\sum_{\ell=2}^{m}\sum_{i_\ell\in[d]} \Theta(i,i_2,\ldots,i_m) > \frac{1}{d^{m-1}}\sum_{\ell=2}^{m}\sum_{i_\ell\in[d]} \Theta(j,i_2,\ldots,i_m), \text{ for all } i>j.$$

Step 1 (sorting): Find $\hat{\sigma}$ so that the degree of $\mathcal{Y}\circ\hat{\sigma}^{-1}$ is increasing.
Step 2 (smoothing): Estimate signal matrix $\hat{\Theta} = \text{Block}_k(\mathcal{Y}\circ\hat{\sigma}^{-1})$, where $\text{Block}_k(\Theta) := \text{Average}\{\Theta(\omega)\colon \lfloor \omega k/d \rfloor = \lceil \omega' k/d \rceil\}$, for all $\omega' \in [d]^m$.

Spectral method extended from [3]

Step 1 (Unfolding): Unfold $\mathcal{Y}$ into $\text{Mat}(\mathcal{Y}) \in \mathbb{R}^{d^{\lfloor m/2\rfloor}\times d^{\lceil m/2\rceil}}$.
Step 2 (SVD): Obtain SVD of $\text{Mat}(\mathcal{Y}) = \sum_{i\in[d^{\lfloor m/2\rfloor}]} \lambda_i \boldsymbol{u}_i\boldsymbol{v}_i^T$.
Step 3 (Thresholding): Obtain $\text{Mat}(\hat{\Theta}) = \sum_{i\in[d^{\lfloor m/2\rfloor}]} \lambda_i \boldsymbol{u}_i\boldsymbol{v}_i^T \mathbb{1}\{\lambda_i \geq d^{\frac{\lceil m/2\rceil}{2}}\}$ and fold back to tensor $\hat{\Theta}$.

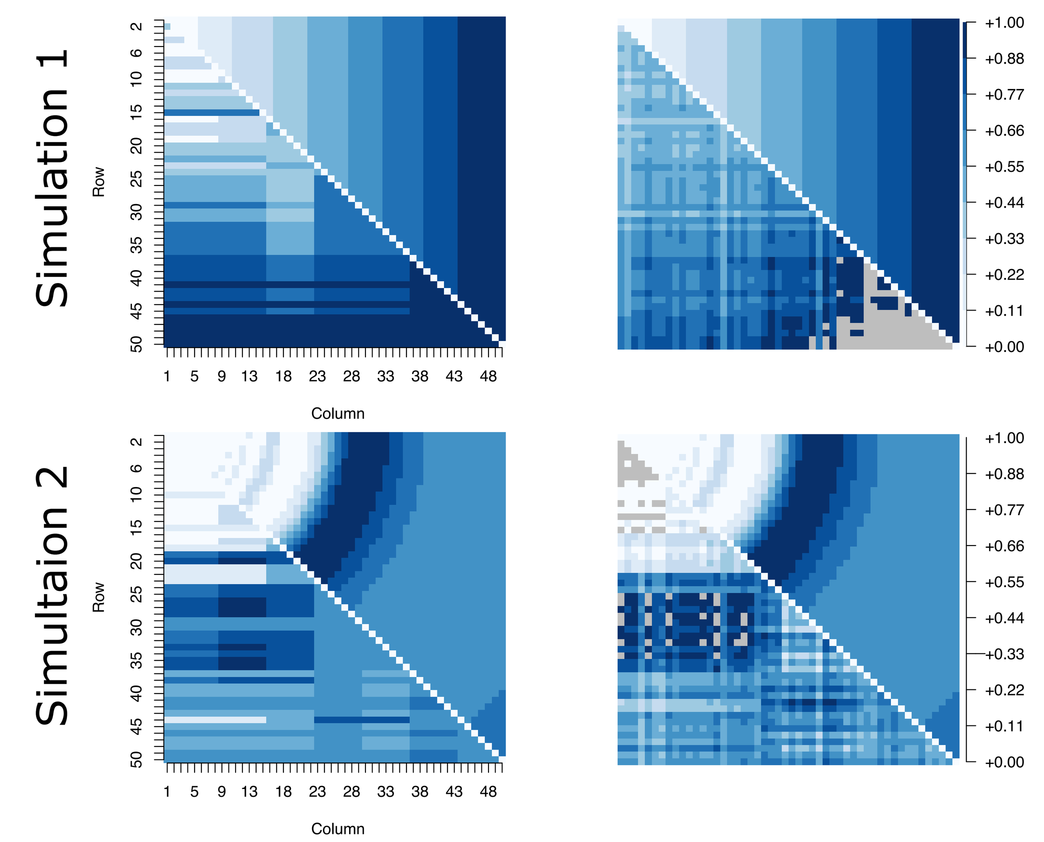


Figure 1. Right triangular matrices show the true signal and left ones show the estimated ones. Simulation 1 follows monotonic degree assumption while Simulation 2 does not.

	SBA	SAS	Spectral
Convergence rate (power of d^{-1})	$\frac{2m}{m+2}$	$\frac{2m}{m+2}$ *	$\frac{4\lfloor m/2\rfloor}{2\lfloor m/2\rfloor+4}$
Polynomial complexity	No	Yes	Yes

* Restricted model

Table 1. Comparison of SBA, SAS, and Spectral method for $\alpha = 1$ and $m > 2$.

Remark: as m increases, convergence rates of both algorithms get closer to that of SBA.

References

[1] Stanley Chan and Edoardo Airoldi. A consistent histogram estimator for exchangeable graph models. In *International Conference on Machine Learning*, pages 208–216. PMLR, 2014.

[2] Rungang Han, Yuetian Luo, Miaoyan Wang, and Anru R Zhang. Exact clustering in tensor block model: Statistical optimality and computational limit. *arXiv preprint arXiv:2012.09996*, 2020.

[3] Jiaming Xu. Rates of convergence of spectral methods for graphon estimation. In *International Conference on Machine Learning*, pages 5433–5442. PMLR, 2018.

IFDS Summer School 2021

The research was supported in part by NSF DMS-1915978, NSF DMS-2023239, and Wisconsin Alumni Research Foundation