Smooth tensor estimation with unknown permutation

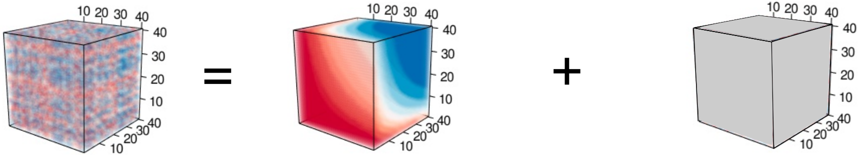
Chanwoo Lee[1] and Miaoyan Wang[2]

Department of Statistics, University of Wisconsin - Madison

NeurIPS workshop on Quantum Tensor Networks in Machine Learning

[1] chanwoo.lee@wisc.edu [2] miaoyan.wang@wisc.edu

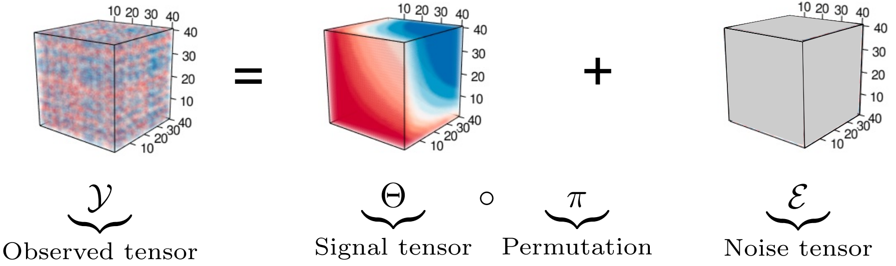
Main problems: the permuted signal plus noise model



$$\underbrace{\mathcal{Y}}_{\text{Observed tensor}} = \underbrace{\Theta}_{\text{Signal tensor}} \circ \underbrace{\pi}_{\text{Permutation}} + \underbrace{\mathcal{E}}_{\text{Noise tensor}}$$

- Question: How to estimate the permuted signal tensor $\Theta \circ \pi$?

Main problems: the permuted signal plus noise model



- Question: How to estimate the permuted signal tensor $\Theta \circ \pi$?
- We assume that there exists a multivariate function $f : [0,1]^m \to \mathbb{R}$ underlying the signal tensor, such that

$$\Theta_{i_1,\ldots,i_m} = f\left(\frac{i_1}{d}, \ldots, \frac{i_m}{d}\right), \text{ for all } i_1, \ldots, i_m \in [d].$$

Our contribution

	Pananjady and Samworth (2020)	Balasubramanian (2021)	Li et al. (2019)	**Ours***
model structure	monotonic	Lipschitz	Lipschitz	α-smoothness
minimax lower bound	$\surd$	$\times$	$\times$	$\surd$
error rate for order-3 tensors	d^{-1}	$d^{-6/5}$	d^{-1}	d^{-2}
polynomial algorithm	$\surd$	$\times$	$\surd$	$\surd$

We list here only the result for infinitely smooth order-3 tensors.

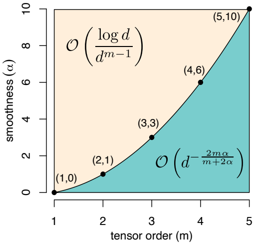
- We develop a general permuted model for an arbitrary smoothness and order of tensors with optimal rate.

Our contribution

	Pananjady and Samworth (2020)	Balasubramanian (2021)	Li et al. (2019)	**Ours***
model structure	monotonic	Lipschitz	Lipschitz	α-smoothness
minimax lower bound	$\surd$	$\times$	$\times$	$\surd$
error rate for order-3 tensors	d^{-1}	$d^{-6/5}$	d^{-1}	d^{-2}
polynomial algorithm	$\surd$	$\times$	$\surd$	$\surd$

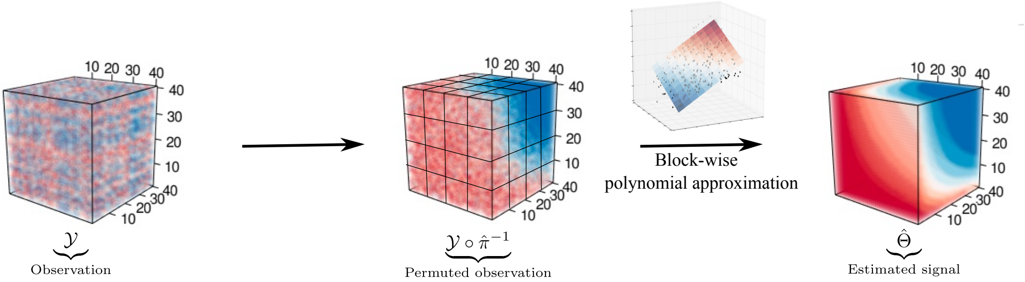
We list here only the result for infinitely smooth order-3 tensors.

- We develop a general permuted model for an arbitrary smoothness and order of tensors with optimal rate.

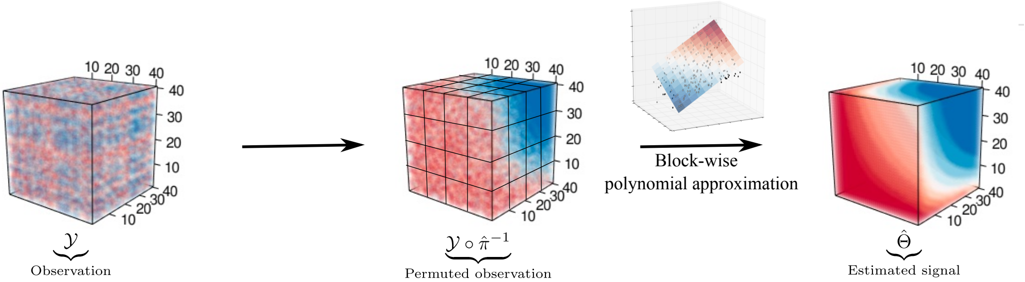


- We discover a phase transition phenomenon with respect to the smoothness threshold needed for optimal tensor recovery.
- We provide an efficient polynomial-time Borda count algorithm that provably achieves optimal rate.

Block-wise polynomial approximation



Block-wise polynomial approximation



- We propose the least square estimation,

$$(\hat{\Theta}^{\text{LSE}}, \hat{\pi}^{\text{LSE}}) = \underset{\Theta \in \mathscr{B}(k,\ell),\ \pi \in [d] \to [d]}{\arg\min} \|\mathcal{Y} - \Theta \circ \pi\|_F \quad \text{where,}$$

$$\mathscr{B}(k,\ell) = \left\{ \mathcal{B} \in (\mathbb{R}^d)^{\otimes m} \colon \mathcal{B}(\omega) = \sum_{\Delta \in \mathcal{E}_k} \text{Poly}_{\ell,\Delta}(\omega) \mathbb{1}\{\omega \in \Delta\} \text{ for all } \omega \in [d]^m \right\}.$$

Least-squares estimation error and its optimality

For two tensor Θ_1, Θ_2, define $\text{MSE}(\Theta_1, \Theta_2) = \frac{1}{d^m}\|\Theta_1 - \Theta_2\|_F^2$.

Least-squares estimation error (L. and Wang 2021)

Suppose that the generating function f is α-Hölder smooth. For optimally chosen polynomial degree ℓ^* and the number of groups k^*,

$$\text{MSE}(\hat{\Theta}^{\text{LSE}} \circ \hat{\pi}^{\text{LSE}}, \Theta \circ \pi) \lesssim \begin{cases} d^{-\frac{2m\alpha}{m+2\alpha}} & \text{when } \alpha < \frac{m(m-1)}{2}, \\ \frac{\log d}{d^{m-1}} & \text{when } \alpha \geq \frac{m(m-1)}{2}. \end{cases}$$

$\ell^* = \min(\lceil \alpha \rceil, m(m-1)/2) - 1$ and $k^* = \lceil d^{m/(m+2\min(\alpha, \ell^*+1))} \rceil$

- The error consists of the nonparametric error and permutation error.
- The dominating error depends on the smoothness and order of tensor.
- We show that the least-square estimation is minimax rate-optimal.

Least-squares estimation error and its optimality

For two tensor Θ_1, Θ_2, define $\text{MSE}(\Theta_1, \Theta_2) = \frac{1}{d^m}\|\Theta_1 - \Theta_2\|_F^2$.

Least-squares estimation error (L. and Wang 2021)

Suppose that the generating function f is α-Hölder smooth. For optimally chosen polynomial degree ℓ^* and the number of groups k^*,

$$\text{MSE}(\hat{\Theta}^{\text{LSE}} \circ \hat{\pi}^{\text{LSE}}, \Theta \circ \pi) \lesssim \begin{cases} d^{-\frac{2m\alpha}{m+2\alpha}} & \text{when } \alpha < \frac{m(m-1)}{2}, \\ \frac{\log d}{d^{m-1}} & \text{when } \alpha \geq \frac{m(m-1)}{2}. \end{cases}$$

$\ell^* = \min(\lceil \alpha \rceil, m(m-1)/2) - 1$ and $k^* = \lceil d^{m/(m+2\min(\alpha, \ell^*+1))} \rceil$

- The error consists of the nonparametric error and permutation error.
- The dominating error depends on the smoothness and order of tensor.
- We show that the least-square estimation is minimax rate-optimal.

However, the algorithm for the least square estimation is computationally intractable.

Polynomial-time algorithm: Borda count estimation

1. **Sorting stage**: Estimate a permutation $\hat{\pi}^{\text{BC}}$ such that the permuted score function $\tau \circ (\hat{\pi}^{\text{BC}})^{-1}$ is monotonically increasing, where

$$\tau(i) = \frac{1}{d^{m-1}} \sum_{(i_2,\ldots,i_m)\in[d]^{m-1}} \mathcal{Y}(i, i_2, \ldots, i_m).$$

2. **Polynomial approximation stage**: Estimate the degree-ℓ polynomial block tensor

$$\hat{\Theta}^{\text{BC}} = \operatorname*{arg\,min}_{\Theta\in\mathscr{B}(k,\ell)} \|\mathcal{Y} \circ (\hat{\pi}^{\text{BC}})^{-1} - \Theta\|_F.$$

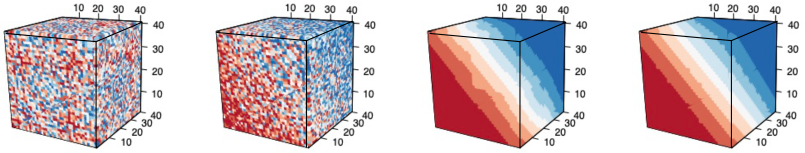
Polynomial-time algorithm: Borda count estimation

1. **Sorting stage**: Estimate a permutation $\hat{\pi}^{\text{BC}}$ such that the permuted score function $\tau \circ (\hat{\pi}^{\text{BC}})^{-1}$ is monotonically increasing, where

$$\tau(i) = \frac{1}{d^{m-1}} \sum_{(i_2,\ldots,i_m)\in[d]^{m-1}} \mathcal{Y}(i, i_2, \ldots, i_m).$$

2. **Polynomial approximation stage**: Estimate the degree-ℓ polynomial block tensor

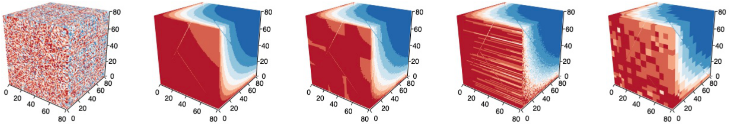
$$\hat{\Theta}^{\text{BC}} = \underset{\Theta\in\mathscr{B}(k,\ell)}{\arg\min} \|\mathcal{Y} \circ (\hat{\pi}^{\text{BC}})^{-1} - \Theta\|_F.$$



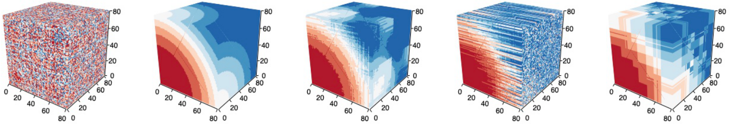
Observation Sorted observation Polynomial approximation True signal

Borda count algorithm provably achieves optimal rate under monotonicity assumptions

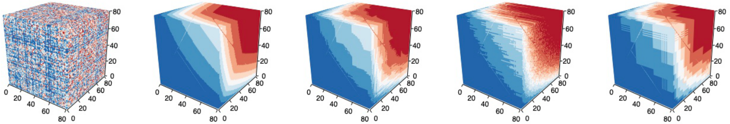
Simulation results



Observation | Model 1 | Borda Count (5.4×10^{-4}) | Spectral (3.1×10^{-3}) | LSE (5.3×10^{-3})



Observation | Model 3 | Borda Count (3.6×10^{-4}) | Spectral (6.5×10^{-3}) | LSE (1.1×10^{-3})



Observation | Model 5 | Borda Count (2.5×10^{-3}) | Spectral (7.5×10^{-3}) | LSE (3.6×10^{-3})

Thank you!

References I

Balasubramanian, K. (2021). Nonparametric modeling of higher-order interactions via hypergraphons. *arXiv preprint arXiv:2105.08678*.

Li, Y., Shah, D., Song, D., and Yu, C. L. (2019). Nearest neighbors for matrix estimation interpreted as blind regression for latent variable model. *IEEE Transactions on Information Theory*, 66(3):1760–1784.

Pananjady, A. and Samworth, R. J. (2020). Isotonic regression with unknown permutations: Statistics, computation, and adaptation. *arXiv preprint arXiv:2009.02609*.