Title: Smooth Tensor Estimation with Unknown Permutations

Abstract: We consider the problem of structured tensor denoising in the presence of unknown permutations. Such data problems arise commonly in recommendation system, neuroimaging, community detection, and multiway comparison applications. Here, we develop a general family of smooth tensor models up to arbitrary index permutations; the model incorporates the popular tensor block models and Lipschitz hypergraphon models as special cases. We show that a constrained least-squares estimator in the block-wise polynomial family achieves the minimax error bound. A phase transition phenomenon is revealed with respect to the smoothness threshold needed for optimal recovery. In particular, we find that a polynomial of degree up to $(m-2)(m+1)/2$ is sufficient for accurate recovery of order-m tensors, whereas higher degree exhibits no further benefits. This phenomenon reveals the intrinsic distinction for smooth tensor estimation problems with and without unknown permutations. Furthermore, we provide an efficient polynomial-time Borda count algorithm that provably achieves optimal rate under monotonicity assumptions. The efficacy of our procedure is demonstrated through both simulations and Chicago crime data analysis.

Authors: Chanwoo Lee, Miaoyan Wang, UW-Madison

Paper status: This paper is currently under review by JASA Theory and Methods. Part of the work was selected as **Oral Presentation** at NeurIPS 2021 Workshop on Quantum Tensor Networks in Machine Learning. We believe our results will be of interest to both machine learning and statistics communities. Since JSM has complementary audience compared to NeurIPS, the first author would like to apply for IMS support to present the work in JSM 2022.

Summary: Higher-order tensor datasets arise ubiquitously in modern data science applications. Tensor structure provides effective representation of data that classical vector- and matrix-based methods fail to capture. One example is music recommendation system that records ratings of songs from users on different contexts. This three-way tensor of user $\times$ song $\times$ context allows us to investigate interaction of users and songs under a context-specific manner. Another example is network analysis that studies the connection pattern among nodes. Pairwise interactions are often insufficient to capture the complex relationships, whereas multi-way interactions improve understanding the networks in social sciences and recommendation system. In both examples, higher-order tensors represent multi-way interactions in an efficient way.

Our paper studies a class of structured tensors, *permuted smooth tensors*, of the following form:

$$\mathcal{Y} = \Theta \circ \pi + \mathcal{E}, \quad \text{where} \quad \Theta_{i_1,\dots,i_m} = f\left(\frac{i_1}{d}, \dots, \frac{i_m}{d}\right). \tag{1}$$

where $\pi\colon [d] \to [d]$ is an *unknown* latent permutation, Θ is an *unknown* order-m d-dimensional signal tensor, and f is an *unknown* multivariate function with certain notion of smoothness, $\Theta \circ \pi$ denotes the permuted tensor after reordering the indices along each of the m modes, and $\mathcal{E}$ is a symmetric noise tensor consisting of zero-mean, independent sub-Gaussian entries with variance bounded by σ^2. Figure 1(a) shows an example of this generative model for the matrix case $m=2$. Our primary goal is to estimate a permuted smooth signal tensor from a noisy observation (1).

The estimation problem of (1) falls into the general category of structured learning with *latent permutation*, which has recently observed a surge of interest. Models involving latent permutations include graphon [2, 4], stochastic transitivity models [7], and crowd labeling [5]. Most of these methods are developed for matrices. The tensor counterparts are far less well understood. In the main paper, we provide statistical and computational estimation accuracy for the permuted smooth tensor model (1). Our major contributions are summarized below.
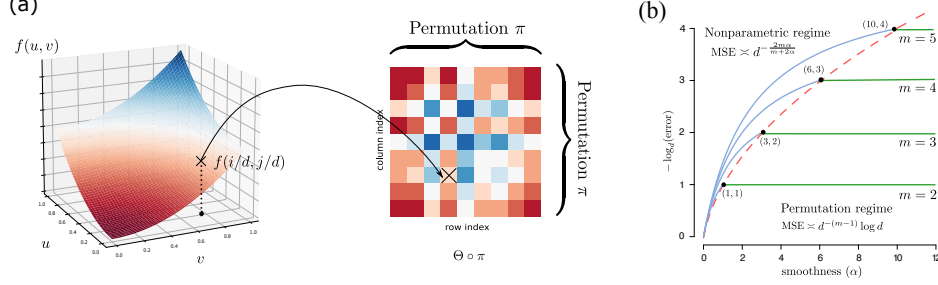
Figure 1: (a): Illustration of order-m d-dimensional permuted smooth tensor models with $m = 2$. (b): Phase transition of mean squared error (MSE) (on -$\log d$ scale) as a function of smoothness α and tensor order m. Bold dots correspond to the critical smoothness level above which higher smoothness exhibits no further benefits to tensor estimation; See Theorems 1-3 in the full paper.

	Pananjady et al [6]	Balasubramanian [1]	Li et al [5]	**Ours***
Model structure	monotonic	Lipschitz	Lipschitz	α-smoothness
Minimax lower bound	$\checkmark$	$\times$	$\times$	$\checkmark$
Error rate for order-m tensor*	d^{-1}	$d^{-2m/(m+2)}$	$d^{-\lfloor m/3 \rfloor}$	$d^{-(m-1)}$
(e.g., when $m = 3$)		$(d^{-6/5})$	(d^{-1})	(d^{-2})
Polynomial algorithm	$\checkmark$	$\times$	$\checkmark$	$\checkmark$

Table 1: Comparison of our results with previous works. *We list here only the result for infinitely smooth order-3 tensors. Our results allow general tensors of arbitrary order m and smoothness α.

- We develop a general permuted α-smooth tensor model, where $\alpha \geq 0$ is some natural measure of functional smoothness (formal definition in the full paper). Table 1 summarizes our improvement over previous works on tensor learning with latent permutations. We establish the upper bound of the least-square estimate with high probability and show that this upper bound matches with minimax lower bound implying its optimality.

- We discover a phase transition phenomenon with respect to the smoothness threshold needed for optimal recovery in model (1). Figure 1(b) plots the dependence of estimation error in terms of smoothness level α for tensors of order m. We find that the estimation accuracy improves with smoothness in the regime $\alpha \leq m(m-1)/2$, but then it becomes a constant of α in the regime $\alpha > m(m-1)/2$. The phenomenon is distinctive from matrix problems [4] and classical *non-permuted* smooth function estimation, thereby highlighting the fundamental challenges in our new setting.

- We propose two estimation algorithms with accuracy guarantees: the least-squares estimation and Borda count estimation. The least-squares estimation is minimax optimal but computationally hard. The Borda count algorithm is polynomial-solvable, and we show it provably achieves the same optimal rate under extra monotonicity assumptions. The algorithm handles a broad range of data types, including continuous and binary observations.

We illustrate the efficacy of our method through both simulations and Chicago crime data application. Chicago crime dataset consists of crime counts reported in the city of Chicago, ranging from January 1st, 2001 to December 11th, 2017. The observed tensor is an order-3 tensor with entries representing the log counts of crimes from 24 hours, 77 community areas, and 32 crime types. We use our Borda count algorithm to identify four community areas based on crime counts. Figure 2(b) shows the four areas overlaid on the Chicago map. We find that the clusters are consistent with actual locations. In addition, we compare the cluster patterns with benchmark maps based on homicides and shooting incidents in Chicago shown in Figure 2(a). We find that our clusters share similar geographical patterns with benchmark maps. The results demonstrate the power of our approach in detecting

meaningful pattern from tensor data.

(a)

Homicides through June 21

2019 | 235
2020 | 295

By police district

0 to 3
4 to 7
8 to 10
11 to 20
21 to 42

Most homicides:
Harrison, 42

SOURCE: Chicago Police Department

Shooting incidents through June 21

2019 | 902
2020 | 1,250

By police district

4 to 10
11 to 28
29 to 57
58 to 82
83 to 173

Most shooting incidents:
Harrison, 173

CHICAGO TRIBUNE

(b)

Four areas learned from Chicago Crime data using Borda Count Estimation

Area 1
Area 2
Area 3
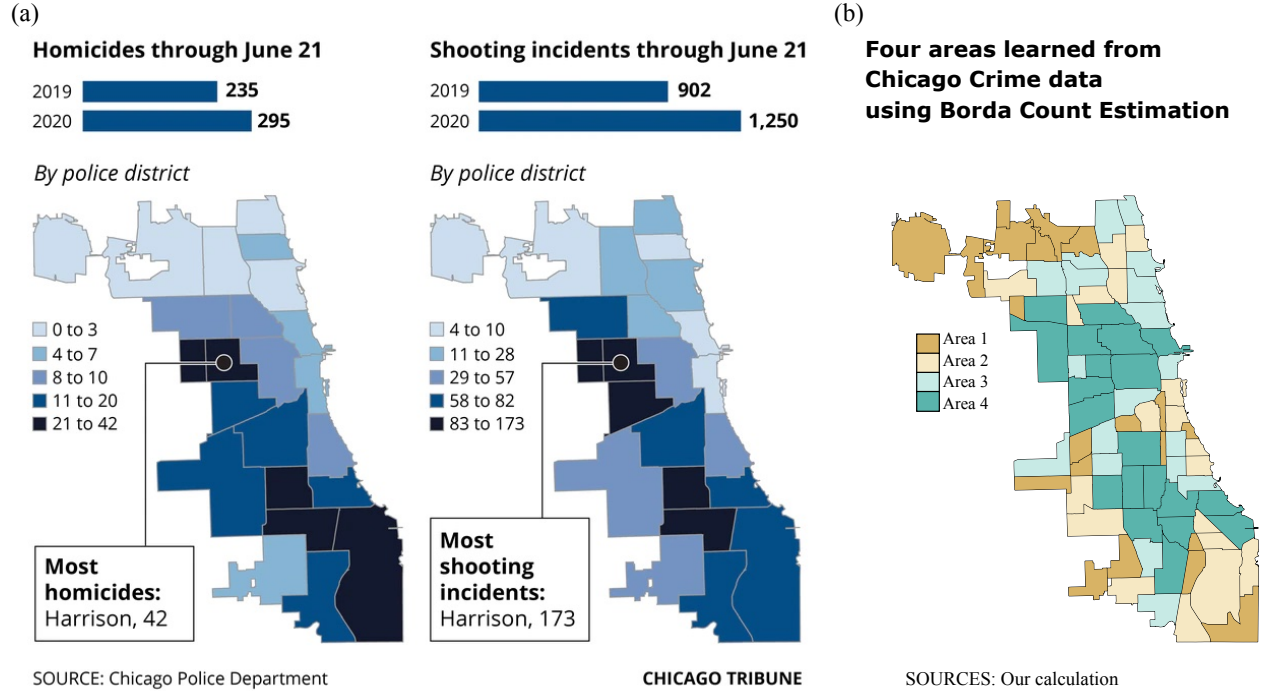Area 4

SOURCES: Our calculation

Figure 2: Chicago crime maps. Figure(a) is the benchmark map based on homicides and shooting incidents in community areas in Chicago [3]. Figure(b) shows the four clustered areas identified by our method.

We believe our results will be of interest to a very broad readership – from those interested in foundations of tensor methods to those in tensor data applications. Our method will help the practitioners efficiently analyze tensor datasets in various areas. Toward this end, the software package and all data used have been publicly released at CRAN.

Software and data: https://cloud.r-project.org/web/packages/SmoothTensor/index.html

References

[1] Krishnakumar Balasubramanian. Nonparametric modeling of higher-order interactions via hypergraphons. *Journal of Machine Learning Research*, 22:1–25, 2021.

[2] Chao Gao, Yu Lu, and Harrison H Zhou. Rate-optimal graphon estimation. *The Annals of Statistics*, 43(6):2624–2652, 2015.

[3] Gorner Jeremy. A trying first half of 2020 included spike in shootings and homicides in chicago. *Chicago Tribune*, 2020.

[4] Olga Klopp, Alexandre B Tsybakov, and Nicolas Verzelen. Oracle inequalities for network models and sparse graphon estimation. *The Annals of Statistics*, 45(1):316–354, 2017.

[5] Yihua Li, Devavrat Shah, Dogyoon Song, and Christina Lee Yu. Nearest neighbors for matrix estimation interpreted as blind regression for latent variable model. *IEEE Transactions on Information Theory*, 66(3):1760–1784, 2019.

[6] Ashwin Pananjady and Richard J Samworth. Isotonic regression with unknown permutations: Statistics, computation, and adaptation. *The Annals of Statistics, in press*, 2021.

[7] N Shah, Sivaraman Balakrishnan, and M Wainwright. Low permutation-rank matrices: Structural properties and noisy completion. *Journal of Machine Learning Research*, 20:1–43, 2019.