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, community detection, and multiway comparison applications. Here, we develop a general family of smooth tensors up to arbitrarily index permutations; the model incorporates the popular block models and graphon models. We show that a constrained least-squares estimate 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 of $(m-2)(m+1)/2$ is sufficient for accurate recovery of order-m tensors, whereas higher smoothness exhibits no further benefits. 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 date applications.

1 Introduction

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 [2]. 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 [4] and recommendation system [6]. In both examples, higher-order tensors represent multi-way interactions in an efficient way.

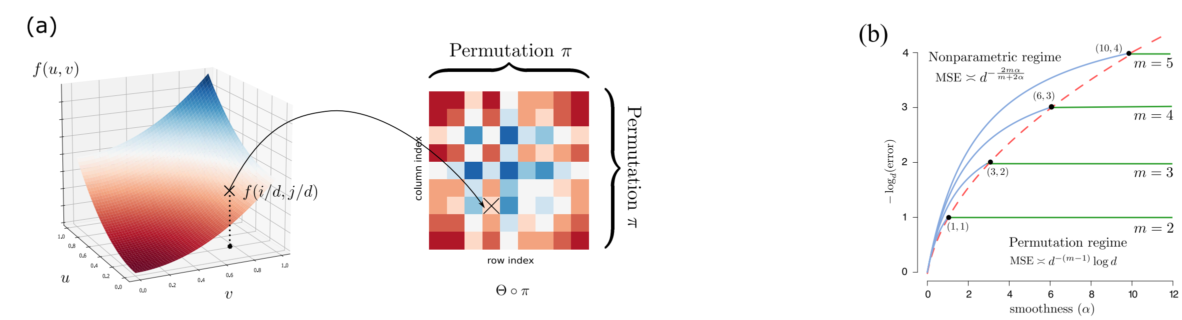


Figure 1: (a): Illustration of order-m d-dimensional permuted smooth tensor models with $m = 2$. (b): Phase transition of mean squared error (MSE) (on -log d scale) as a function of smoothness α and tensor order m. Bold dots correspond to the critical smoothness level above which higher smoothness exhibits no further benefits to tensor estimation. See Theorems 1-3 in Sections 3-4 for details.

Tensor estimation problem cannot be solved without imposing structure. We study a class of structured tensors, *permuted smooth tensors*, of the following form:

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

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 $m = 2$.

For ease of presentation, we focus on symmetric tensors; our models and techniques easily generalize to non-symmetric tensors. Our primary goal is to estimate a permuted smooth signal tensor from a noisy observation (1).

Related work and our contributions. 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 [4, 8], stochastic transitivity models [11], and crowd labeling [9]. Most of these methods are developed for matrices. The tensor counterparts are far less well understood.

The primary goal of our work is to provide statistical and computational estimation accuracy for the permuted smooth tensor model (1). Our major contributions are summarized below.

	Pananjady et al [10]	Balasubramanian [1]	Li et al [9]	**Ours***
Model structure	monotonic	Lipschitz	Lipschitz	α-smoothness
Minimax lower bound	$\surd$	$\times$	$\times$	$\surd$
Error rate for order-m tensor* (e.g., when $m=3$)	d^{-1}	$d^{-2m/(m+2)}$ $(d^{-6/5})$	$d^{-\lfloor m/3 \rfloor}$ (d^{-1})	$d^{-(m-1)}$ (d^{-2})
Polynomial algorithm	$\surd$	$\times$	$\surd$	$\surd$

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; See Theorems 1-3 in Sections 3-4.

- We develop a general permuted α-smooth tensor model, where $\alpha \geq 0$ is some natural measure of functional smoothness (formal definition in Section 2). In contrast to earlier work [1, 9], we establish the statistically optimal rate and fully characterize its dependence on tensor order, dimension, and smoothness index. Table 1 summarizes our improvement over previous works on tensor learning with latent permutations.

- 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 [8] 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. Application to Chicago crime analysis is presented to showcase the usefulness of our method. The software package and all data used have been publicly released at CRAN.

Notation. We use $[d] = \{1, \ldots, d\}$ for d-set with $d \in \mathbb{N}_+$. For a set S, $\mathbb{1}_S$ denotes the indicator function. For positive two sequences $\{a_n\}, \{b_n\}$, we denote $a_n \lesssim b_n$ if $\lim_{n\to\infty} a_n/b_n \leq c$, and $a_n \asymp b_n$ if $c_1 \leq \lim_{n\to\infty} a_n/b_n \leq c_2$ for some constants $c, c_1, c_2 > 0$. Given number $a \in \mathbb{R}$, the floor function $\lfloor a \rfloor$ is the largest integer no greater than a, and the ceiling function $\lceil a \rceil$ is the smallest integer no less than a. An event A is said to occur *with high probability* if $\mathbb{P}(A)$ tends to 1 as the tensor dimension $d \to \infty$.We use $\Theta_{i_1,\ldots,i_m}$ to denote the tensor entry indexed by $(i_1, \ldots, i_m)$, and use $\Theta \circ \pi$ to denote the permuted tensor such that $(\Theta \circ \pi)_{i_1,\ldots,i_m} = \Theta_{\pi(i_1),\ldots,\pi(i_m)}$ for all $(i_1, \ldots, i_m) \in [d]^m$. We use $S(d) = \{\pi \colon [d] \to [d]\}$ to denote all possible permutations on $[d]$.

2 Smooth tensor model with unknown permutation and block-wise approximation

Suppose we observe an order-m d-dimensional symmetric data tensor from the permuted tensors in (1). We assume the generating function f is in the α-Hölder smooth family.

Definition 1 (α-Hölder smooth)**.** A function $f \colon [0,1]^m \to \mathbb{R}$ is α-Hölder smooth, denoted as $f \in \mathcal{H}(\alpha)$, if there exists a polynomial $\text{Poly}_{\lfloor \alpha \rfloor}(\boldsymbol{x} - \boldsymbol{x}_0)$ of degree $\lfloor \alpha \rfloor$, such that

$$|f(\boldsymbol{x}) - \text{Poly}_{\lfloor \alpha \rfloor}(\boldsymbol{x} - \boldsymbol{x}_0)| \leq C \|\boldsymbol{x} - \boldsymbol{x}_0\|_\infty^\alpha, \text{ for all } \boldsymbol{x}, \boldsymbol{x}_0 \in [0,1]^m \text{ and a constant } C > 0. \tag{2}$$

In addition to the function class $\mathcal{H}(\alpha)$, we define the smooth tensor class based on discretization (1),

$$\mathcal{P}(\alpha) = \left\{ \Theta \in \mathbb{R}^{d \times \cdots \times d} \colon \Theta(\omega) = f\left(\frac{\omega}{d}\right) \text{ for all } \omega = (i_1, \ldots, i_m) \in [d]^m \text{ and } f \in \mathcal{H}(\alpha) \right\}. \tag{3}$$

Combining (1) and (2) yields our proposed *permuted smooth tensor model*. The unknown parameters are the smooth tensor $\Theta \in \mathcal{P}(\alpha)$ and latent permutation $\pi \in S(d)$. The model is visualized in Figure 1(a) for the case $m = 2$ (matrices).

We give two concrete examples to show the applicability of our permuted smooth tensor model.

Example 1 (Four-player game tensor)**.** Consider a four-player board game. Suppose there are in total d players, among which all combinations of four have played against each other. The game results are summarized as an order-4 (asymmetric) tensor, with entries encoding the winner of the games. Our model is then given by

$$\mathbb{E}(\mathcal{Y}_{i_1,\ldots,i_4}) = \mathbb{P}(\text{user } i_1 \text{ wins over } (i_2,i_3,i_4)) = f\left(\frac{\pi(i_1)}{d},\cdots,\frac{\pi(i_4)}{d}\right).$$

We can interpret the permutation π as the unknown ranking among d players, and the function f the unknown four-players interaction. Operationally, players with similar ranking would have similar performance encoded by the smoothness of f.

Example 2 (Co-authorship networks)**.** Consider co-authorship networks. Suppose there are in total d authors. We say there exists a hyperedge between nodes $(i_1,\ldots,i_m)$ if the authors $i_1,\ldots,i_m$ have co-authored at least one paper. The resulting hypergraph is represented as an order-m (symmetric) adjacency tensor. Our model is then expressed as

$$\mathbb{E}(\mathcal{Y}_{i_1,\ldots,i_m}) = \mathbb{P}(\text{authors } i_1,\ldots,i_m \text{ co-authored}) = f\left(\frac{\pi(i_1)}{d},\cdots,\frac{\pi(i_m)}{d}\right).$$

In this setting, we can interpret the permutation π as the affinity measures of authors, and the function f represents the m-way interaction among authors. Our nonparametric model learns the unknown function f from data.

Our general strategy for estimating the signal tensor in model (3) is based on the block-wise tensor approximation. We first introduce the tensor block model [6]. Then, we extend the idea to block-wise polynomial approximation.

Tensor block model. The tensor block model [6] describes a checkerbroad pattern in the signal tensor. Specifically, suppose that there are k clusters in the tensor dimension d, and the clusters are represented by a clustering function $z\colon [d]\to[k]$. Then, the tensor block model assumes that signal tensor $\Theta\in\mathbb{R}^{d\times\cdots\times d}$ takes values from a mean tensor $\mathcal{S}\in\mathbb{R}^{k\times\cdots\times k}$ according to the clustering function z:

$$\Theta_{i_1,\ldots,i_m} = \mathcal{S}_{z(i_1),\ldots,z(i_m)}, \qquad \text{for all } (i_1,\ldots,i_m)\in[d]^m. \tag{4}$$

A tensor Θ satisfying (4) is called a block-k tensor. Classical tensor block models aim to explain data with a finite number of blocks; this approach is useful when the sample outsizes the parameters. Our nonparametric models (1), by contrast, use infinite number of parameters to allow growing model complexity as sample increases. Therefore, we shift the goal of tensor block model from discovering hidden group structure to approximating the generative process of the function f in (1). In our setting, the number of blocks k should be interpreted as a resolution parameter (i.e., a bandwidth) of the approximation similar to the notion of number of bins in histogram and polynomial regression.

Block-wise polynomial approximation. The tensor block model (4) can be viewed as a discrete version of piece-wise *constant* function. This connection motivates us to use block-wise *polynomial* tensors to approximate α-Hölder functions. For a given block number k, we use $z\colon[d]\to[k]$ to denote the canonical clustering function that partitions $[d]$ into k clusters, $z(i)=\lceil ki/d\rceil$, for all $i\in[d]$. The collection of inverse images $\{z^{-1}(j)\colon j\in[k]\}$ consists of disjoint and equal-sized subsets in $[d]$, and we have $\cup_{j\in[k]}z^{-1}(j)=[d]$ by the construction. We denote $\mathcal{E}_k$ as the m-way partition as a collection of k^m disjoint, equal-sized blocks in $[d]^m$, such that

$$\mathcal{E}_k = \{z^{-1}(j_1)\times\cdots\times z^{-1}(j_m)\colon (j_1,\ldots,j_m)\in[k]^m\}.$$

We refer to $\Delta\in\mathcal{E}_k$ as the *canonical blocks*. We propose to approximate the signal Θ in (1) by degree-ℓ polynomial tensor within each block $\Delta\in\mathcal{E}_k$. Specifically, we use $\mathscr{B}(k,\ell)$ to denote the class of block-k, degree-ℓ polynomial tensors,

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

where $\text{Poly}_{\ell,\Delta}(\cdot)$ denotes a degree-ℓ polynomial function in $\mathbb{R}^m$. Notice that degree-0 polynomial block tensor reduces to the tensor block model (4). We generalize the tensor block model to degree-ℓ polynomial block tensor, in a way analogous to the generalization from k-bin histogram to k-piece-wise polynomial regression.

3 Fundamental limits via least-squares estimation

We develop two estimation methods based on the block-wise polynomial approximation. We first introduce a minimax optimal but computationally inefficient least-squares estimator as a statistical benchmark. In Section 4, we will present a polynomial-time solvable estimator with provably same optimal rate under monotonicity assumptions.

We propose the least-squares estimation for model (1) by minimizing the Frobenius loss under block-k, degree-ℓ polynomial tensor family $\mathscr{B}(k,\ell)$,

$$(\hat\Theta^{\text{LSE}},\hat\pi^{\text{LSE}}) = \mathop{\arg\min}_{\Theta\in\mathscr{B}(k,\ell),\ \pi\in S(d)}\|\mathcal{Y}-\Theta\circ\pi\|_F. \tag{5}$$

The least-squares estimator $(\hat{\Theta}^{\text{LSE}}, \hat{\pi}^{\text{LSE}})$ depends on two tuning parameters: the number of blocks k and the polynomial degree ℓ. The optimal choice (k^*, ℓ^*) is provided in our next theorem.

Theorem 1 (Least-squares estimation error)**.** *Consider the order-m ($m \geq 2$) permuted smooth tensor model* (1) *with $\Theta \in \mathcal{P}(\alpha)$. Then, the estimator $\hat{\Theta}^{\text{LSE}} \circ \hat{\pi}^{\text{LSE}}$ in* (5) *satisfies with high probability*

$$\frac{1}{d^m}\|\hat{\Theta}^{\text{LSE}} \circ \hat{\pi}^{\text{LSE}} - \Theta \circ \pi\|_F^2 \lesssim \begin{cases} d^{-\frac{2m\alpha}{m+2\alpha}} & \text{when } \alpha < m(m-1)/2, \\ \frac{\log d}{d^{m-1}} & \text{when } \alpha \geq m(m-1)/2, \end{cases} \tag{6}$$

under the optimal choice of $\ell^* = \min(\lfloor \alpha \rfloor, (m-2)(m+1)/2)$ and $k^* = \lceil d^{\frac{m}{m+2\min(\alpha,\ell^*+1)}} \rceil$.

Theorem 1 establishes the upper bound for the mean squared error of the least-squares estimator (5). We discuss the asymptotic error rates as $d \to \infty$ while treating the tensor order m and smoothness α fixed. The least-squares estimation error has two sources of error: the nonparametric error $d^{-\frac{2m\alpha}{m+2\alpha}}$ and the clustering error $\log d / d^{m-1}$. When the function f is smooth enough, estimating the function f becomes relatively easier compared to estimating the permutation π. This intuition coincides with the fact that the clustering error dominates the nonparametric error when $\alpha \geq m(m-1)/2$.

We now compare our results with existing work in the literature. In the matrix case ($m = 2$), our block-wise constant approximation and convergence rate reduce to the results in [8]. For higher order tensor case ($m \geq 3$), earlier work [1] conjectures that constant block approximation ($\ell^* = 0$) remains minimax optimal for tensors. Our Theorem 1 disproves this conjecture, and we reveal a much faster rate $d^{-(m-1)}$ compared to the conjectured lower bound $d^{-2m/(m+2)}$ [1]. In fact, permuted α-smooth tensors of order-3 require quadratic approximation ($\ell^* = 2$) with $k^* \asymp d^{1/3}$ blocks, for all $\alpha \geq 2$. The results show the clear difference from matrices and highlight the challenges with tensors.

The next theorem shows that the critical polynomial degree up to $(m-2)(m+1)/2$ is not only sufficient but also necessary for accurate estimation of order-m permuted smooth tensors.

Theorem 2 (Minimax lower bound)**.** *For any given $\alpha \in (0, \infty)$, the estimation problem based on model* (1) *obeys the minimax lower bound*

$$\inf_{(\hat{\Theta}, \hat{\pi})} \sup_{\Theta \in \mathcal{P}(\alpha), \pi \in S(d)} \mathbb{P}\left(\frac{1}{d^m}\|\Theta \circ \pi - \hat{\Theta} \circ \hat{\pi}\|_F^2 \gtrsim d^{-\frac{2m\alpha}{m+2\alpha}} + d^{-(m-1)} \log d\right) \geq 0.8.$$

The above result demonstrates that the upper bound (6) is minimax optimal. Theorem 2 is obtained via information-theoretical analysis and thus applies to all estimators including, but not limited to, the least-squares estimator (5) and Borda count estimator introduced in next section.

4 An adaptive and computationally feasible procedure

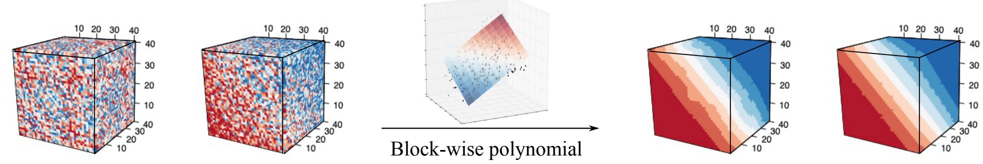
At this point, we should note that the least-squares estimation in (5) is generally computationally hard. In this section, we propose an efficient polynomial-time *Borda count* algorithm with provably same optimal rate under the β-monotonicity condition. We first introduce β-monotonicity condition.

Definition 2 (β-monotonicity)**.** A function $f \colon [0,1]^m \to \mathbb{R}$ is called β-monotonic, denoted as $f \in \mathcal{M}(\beta)$, if

$$\left(\frac{i-j}{d}\right)^{1/\beta} \leq g(i) - g(j) \text{ for all } i > j \in [d], \quad \text{where } g(i) := \frac{1}{d^{m-1}} \sum_{(i_2,\ldots,i_m) \in [d]^m} f\left(\frac{i}{d}, \frac{i_2}{d}, \ldots, \frac{i_m}{d}\right).$$

Our β-monotonicity condition extends the strictly monotonic degree condition in the graphon literature [3]; the latter is a special case of our definition with $\beta = 1, m = 2$. Our β-monotonicity condition is also related to isotonic functions [5, 10] which assume the coordinate-wise monotonicity, i.e., $f(x_1, \ldots, x_d) \leq f(x_1', \ldots, x_d')$ when $x_i \leq x_i'$ for $i \in [d]$.

Now we introduce a Borda count estimator that consists of two stages: sorting and block-wise polynomial approximation. The simplified version of the algorithm is described in Algorithm 1 and Figure 2.



(a) Observation (b) Sorted observation Block-wise polynomial approximation (c) Estimated signal (d) True signal

Figure 2: Procedure of Borda count estimation. We first sort the tensor entries using the proposed procedure. Then, we estimate the signal tensor using block-k degree-ℓ polynomial approximation.

Algorithm 1 Borda Count algorithm

Input: Noisy observed data tensor $\mathcal{Y} \in \mathbb{R}^{d \times \cdots \times d}$

1: **Sorting stage:** Compute a permutation $\hat{\pi}^{\text{BC}}$ such that $\tau \circ (\hat{\pi}^{\text{BC}})^{-1}$ is monotonically increasing, where $\tau(i) = \frac{1}{d^{m-1}} \sum_{(i_2,\ldots,i_m)\in[d]^m} \mathcal{Y}_{i,i_2,\ldots,i_m}$.

2: Obtain a rearranged observation $\tilde{\mathcal{Y}}_{i_1,\ldots,i_m} = \mathcal{Y}_{(\hat{\pi}^{\text{BC}})^{-1}(i_1),\ldots,(\hat{\pi}^{\text{BC}})^{-1}(i_m)}$

3: **Block-wise polynomial approximation stage:** Given degree ℓ and block k, solve the following optimization problem, $\hat{\Theta}^{\text{BC}} = \arg\min_{\mathcal{B}\in\mathscr{B}(k,\ell)} \|\tilde{\mathcal{Y}} - \Theta\|_F$.

Output: Estimated signal tensor and permutation $(\hat{\Theta}^{\text{BC}}, \hat{\pi}^{\text{BC}})$.

Theorem 3 (Estimation error for Borda count; simplified version)**.** *Suppose that the signal tensor Θ is generated as in* (1) *with $f \in \mathcal{H}(\alpha) \cap \mathcal{M}(\beta)$. Then estimators $(\hat{\Theta}^{BC}, \hat{\pi}^{BC})$ from Algorithm 1 satisfies*

$$\frac{1}{d^m}\|\hat{\Theta}^{\text{BC}} \circ \hat{\pi}^{\text{BC}} - \Theta \circ \pi\|_F^2 \lesssim \begin{cases} d^{-\frac{2m\alpha}{m+2\alpha}} & \text{when } \alpha < c(\alpha,\beta,m), \\ \left(\frac{\log d}{d^{m-1}}\right)^{\beta\min(\alpha,1)} & \text{when } \alpha \geq c(\alpha,\beta,m), \end{cases}$$

with high probability under the optimal choice of $\ell^* = \min(\lfloor\alpha\rfloor, \lfloor c(\alpha,\beta,m)\rfloor)$ and $k^* = \lceil d^{\frac{m}{m+2\min(\alpha,\ell^*+1)}}\rceil$. Here $c(\alpha,\beta,m) > 0$ is a constant only depending on α, β, and m.

Theorem 3 shows the estimation consistency of Borda count estimator. We find that the Borda count estimator achieves the same minimax-optimal rate as the least-squares estimator under 1-monotonicity condition. The least-squares estimator requires a combinatoric search with exponential-time complexity. By contrast, Algorithm 1 requires only the estimation of degree-ℓ polynomials within k *canonical blocks*. Therefore, the Borda count estimator is polynomial-time efficient.

5 Numerical experiments and data application

Numerical comparisons. We simulate symmetric order-3 d-dimensional tensors based on the permuted smooth tensor model (1) with diverse functions f. The detailed simulation procedure is described in Appendix. We assess the performance for the four popular tensor methods: (a) Spectral method (**Spectral**) [12] on unfolded tensor; (b) Least-squares estimation (**LSE**) with $\ell = 0$ implied by [4]; (c) Lease square estimation (**BAL**) implied by [1]; (d) Our **Borda Count** algorithm. The performance accuracy is assessed via mean square error (MSE) $= d^{-3}\|\Theta \circ \pi - \hat{\Theta} \circ \hat{\pi}\|_F^2$.

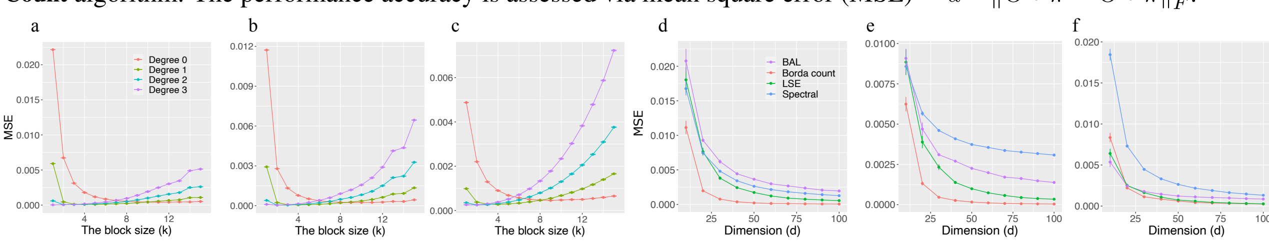


Figure 3: (a-c) MSE comparison versus the number of blocks under simulation models 1,3 and 5 respectively. (d-f) MSE comparison versus tensor dimension under models 1,3 and 5 respectively. MSEs are measured across $n_{\text{sim}} = 20$ replications.

Figure 3a-c examine the impact of the block number k and degree of polynomial ℓ for the approximation. We fix the tensor dimension $d = 100$, and vary the number of blocks $k \in \{1, \ldots, 15\}$ and polynomial degree $\ell \in \{0, 1, 2, 3\}$. The result demonstrates the trade-off in accuracy determined by the number of blocks for each polynomial degree. We find that degree-2 polynomials give the smallest MSE among all considered approximation for order-3 tensors. These observations are consistent with our theoretical results in Sections 3-4. Figure 3d-f shows that our algorithm **Borda Count** achieves the best performance in all scenarios as the tensor dimension increases. The poor performance of **Spectral** can be explained by the loss of multilinear structure in the tensor unfolding procedure. The sub-optimality of the least square estimations is possibly due to its limits in both statistics and computations. Statistically, our theorems have shown that constant block approximation has sub-optimal rates. Computationally, the least-squares estimation (5) is highly non-convex and computationally unstable. The outperformance of **Borda count** demonstrates the efficacy of our method.

Applications to Chicago crime data. Chicago crime tensor dataset is an order-3 tensor with entries representing the log counts of crimes from 24 hours, 77 Chicago community areas, and 32 crime types ranging from January 1st, 2001 to December 11th, 2017. We apply our Borda Count method to Chicago crime dataset. Cross validation result suggests the $(k_1, k_2, k_3) = (6, 4, 10)$, representing the block number for crime hours, community areas, and crime types, respectively. We investigate the four clustered community areas obtained from our Borda Count algorithm. Figure 4a-b shows the four

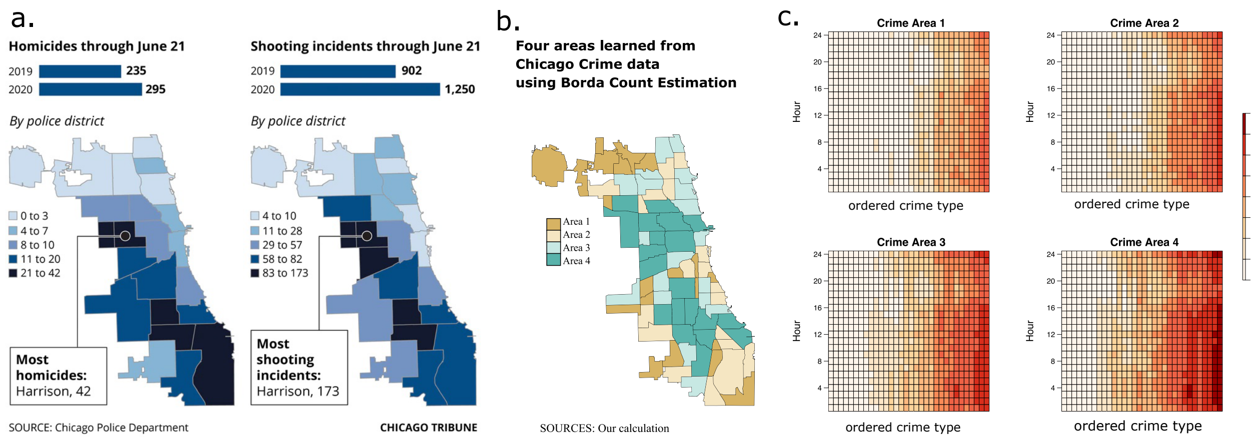


Figure 4: Chicago crime maps based on (a) *Chicago Tribune* article in 2020 [7] and (b) our estimation using Borda Count algorithm. (c) Averaged log counts of crimes according to crime types, hours, and the four areas estimated by our Borda count algorithm. For space consideration, the annotated crime types are described in the Appendix.

areas overlaid on a map of Chicago. We find that our clusters conform the actual locations even though our algorithm did not take any geographic information such as longitude or latitude. In addition, our clusters (Figure 4b) share similar geographical patterns with benchmark result (Figure 4a) based on *Chicago Tribune* article in 2020 [7]. Figure 4c reveals that the major difference among four areas is the crime rates: Area 4 has the highest crime rates, and the crime rates monotonically decrease from Area 4 to Area 1. The variation in crime rates across hour and type, nevertheless, exhibits similarity among the four areas. For example, the number of crimes increases hourly from 8 p.m., peaks at night hours, and then drops to the lowest at 6 p.m. The interpretable similarities and differences among the four community areas demonstrate the applicability of our method in real data.

6 Conclusion

We have presented a suite of statistical theory, estimation methods, and data applications for permuted smooth tensor models. 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.

References

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

[2] Linas Baltrunas, Marius Kaminskas, Bernd Ludwig, Omar Moling, Francesco Ricci, Aykan Aydin, Karl-Heinz Lüke, and Roland Schwaiger. InCarMusic: Context-aware music recommendations in a car. In *International Conference on Electronic Commerce and Web Technologies*, pages 89–100. Springer, 2011.

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

[4] Chao Gao and Zongming Ma. Minimax rates in network analysis: Graphon estimation, community detection and hypothesis testing. *Statistical Science*, 36(1):16–33, 2021.

[5] Qiyang Han, Tengyao Wang, Sabyasachi Chatterjee, and Richard J Samworth. Isotonic regression in general dimensions. *The Annals of Statistics*, 47(5):2440–2471, 2019.

[6] 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.

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

[8] 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.

[9] 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.

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

[11] 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.

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