# Smooth tensor estimation with unknown permutations

**Chanwoo Lee**
University of Wisconsin-Madison
chanwoo.lee@wisc.edu

**Miaoyan Wang**
University of Wisconsin-Madison
miaoyan.wang@wisc.edu

## 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 $m(m-1)/2$ is sufficient for accurate recovery of order-$m$ tensors, whereas higher degree 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 analysis.

## 1 Introduction

Higher-order tensor datasets are rising 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×song×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 molecular system [**?** ] and computer vision [**?** ]. In both examples, higher-order tensors represent multi-way interactions in an efficient way.

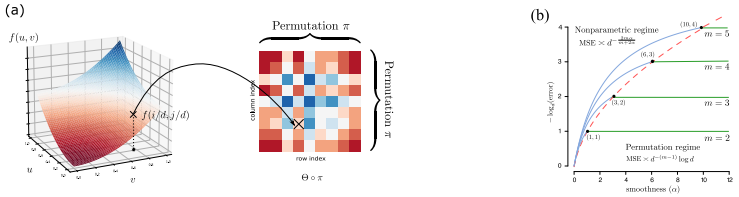



Figure 1: (a): Illustration of order-$m$ $d$-dimensional permuted smooth tensor models with $m=2$. (b): Phase transition of mean squared error (MSE) (on $-\log d$ scale) as a function of smoothness $\alpha$ and tensor order $m$. Bold dots correspond to the critical smoothness level above which higher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

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

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

where $\pi\colon [d] \to [d]$ is an unknown latent permutation, $\Theta \in \mathbb{R}^{d\times\cdots\times d}$ is an unknown symmetric signal tensor, and $f$ is an unknown multivariate function with smoothness index $\alpha > 0$ (see Figure 1(a)), and $\mathcal{E}$ is a symmetric noise tensor consisting of zero-mean, independent sub-Gaussian entries with variance bounded by $\sigma^2$. For simplicity of presentation, we focus on symmetric tensors in the main paper; 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.

**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 [**? ?** ], stochastic transitivity models [**?** ], and crowd labeling [**?** ]. Most of these methods are developed for matrices. The tensor counterparts are far less well understood.

We summarize our major contributions. 1). We develop a general permuted $\alpha$-smooth tensor model for an arbitrary smoothness index $\alpha > 0$. In contrast to earlier work [**? ?** ] that focuses only on $\alpha = 1$, we establish the statistically optimal error rate and its dependence on tensor order, dimension, and smoothness index. 2). We discover an intriguing phase transition phenomenon with respect to the smoothness threshold needed for optimal tensor recovery in model (1). The critical threshold $\alpha^*$ (defined in Theorem 1) characterizes two distinct error dependence behaviors on the smooth index $\alpha$. We proved that the error decreases with $\alpha$ in the range $\alpha < \alpha^*$, whereas the error is a constant of $\alpha$ in the range $\alpha > \alpha^*$ (see Figure 1(b)). These results are distinct from the matrix counterparts [**? ? ?** ], thereby highlighting the fundamental challenges with tensors. 3). We provide an efficient polynomial-time Borda count algorithm that provably achieves optimal rate under monotonicity assumptions. Numerical analysis demonstrates the competitive performance of our algorithm.

**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 $a$. We use $\mathcal{O}(\cdot)$ to denote the big-O notation. 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

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 $\alpha$-Hölder smooth family.

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

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

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

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

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

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

In this setting, we can interpret the permutation $\pi$ 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

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

In this setting, we can interpret the permutation $\pi$ as the affinity measures of authors, and the function $f$ represents the $m$-way interaction among authors.

## 3 Block-wise tensor approximation

Our general strategy for estimating the signal tensor is based on the block-wise tensor approximation. We first introduce the tensor block model [**? ?** ]. Then, we extend this model to the block-wise polynomial approximation.

**Tensor block model.** The tensor block model 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 $z$:

$$\Theta_{i_1,\ldots,i_m} = \mathcal{S}_{z(i_1),\ldots,z(i_m)}, \quad \text{for all } (i_1,\ldots,i_m) \in [d]^m. \tag{2}$$

A tensor $\Theta$ satisfying (3) is called a block-$k$ tensor. Tensor block model have shown great success in discoverying hidden group structure for many applications [**? ?** ]. Despite its popularity and great applicability, the tensor block models cannot describe delicate structure of the signal tensor when the tensor dimension $d$ is very large. This parametric model aims 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). Thus, 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 (3) can be viewed as a discrete version of piece-wise *constant* function. This connection motivates us to use block-wise *polynomial* tensors to approximate $\alpha$-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 propose to approximate the signal tensor $\Theta$ in (1) by degree-$\ell$ polynomial tensor within each $\mathcal{E}_k$-block. Specifically, we use $\mathscr{B}(k,\ell)$ to denote the class of block-$k$, degree-$\ell$ polynomial tensors,

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

where $\mathrm{Poly}_{\ell,\Delta}(\cdot)$ denotes a degree-$\ell$ polynomial function in $\mathbb{R}^m$. Notice that degree-0 polynomial block tensor reduces to the tensor block model (3). We genealized the tensor block model to degree-$\ell$ polynomial block tensor, in a way analogous to the generalization from $k$-bin histogram to $k$-piece-wise polynomial regression.

## 4 Fundamental limits via least-squares estimation

We propose two estimation methods based on the block-wise polynomial approximation. We first introduce a minimax optimal but computationally infeasbile least-squares estimator as statistical benchmark. In Section 5, we will present a polynomial-time algorithm with provably same optimal rates under monotonicity assumptions.

We propose the least-squares estimator for the signal tensor and the permutation $(\Theta, \pi)$ by minimizing the Frobenius loss under block-$k$, degree-$\ell$ polynomial tensor family $\mathscr{B}(k,\ell)$,

$$(\hat{\Theta}^{\mathrm{LSE}}, \hat{\pi}^{\mathrm{LSE}}) = \underset{\Theta\in\mathscr{B}(k,\ell),\ \pi\in S(d)}{\arg\min} \|\mathcal{Y} - \Theta\circ\pi\|_F. \tag{3}$$

The least-squares estimator $(\hat{\Theta}^{\mathrm{LSE}}, \hat{\pi}^{\mathrm{LSE}})$ depends on two tuning parameters: the number of blocks $k$ and the polynomial degree $\ell$. The optimal choice $(k^*, \ell^*)$ is provided in our next theorem. The result establishes the upper bound for the mean squared error of the least square estimator (4).

**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 $\hat{\Theta}^{\mathrm{LSE}} \circ \hat{\pi}^{\mathrm{LSE}}$ in* (4) *satisfies with high probability*

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

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

We discuss the asymptotic error rates as $d \to \infty$ while treating the tensor order $m$ and smoothness $\alpha$ fixed. The least square 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 $\pi$. 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. Based on Theorem 1, the best rate is obtained with the choice of $(\ell^*, k^*) = (0, \lceil d^{\frac{1}{\alpha \wedge 1+1}} \rceil)$ in the matrix case ($m = 2$). This block-wise constant approximation and convergence rate reduce to the results in [**?** **?** ]. We improve the conjectured optimal convergence rate in [**?** ] for higher order tensor case ($m \geq 3$). This improvement stems from polynomial tensor approximation. The work in [**?** ] considers only the block-wise constant approximation. This restriction results in sub-optimality because the optimal degree $\ell^*$ is shown to be greater than 0 for higher-order tensors. This result shows the clear difference from matrices and highlights the challenges with tensors.

We show that the upper bound of Theorem 1 is minimax optimal. The result is based on information-theoretical analysis that combines the minimax rate for nonparametric and permutation estimation.

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

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

## 5 An adaptive and computationally feasible procedure

At this point, we should point out that computing the least square optimizer in (4) with polynomial-time algorithm is unknown. We suspect that the algorithm for (4) may be computationally intractable. In this section, we propose an efficient polynomial-time *Borda count* algorithm with provably same optimal rate under the $\beta$-monotonicity condition. We first introduce $\beta$-monotonicity condition.

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

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

This $\beta$-monotonicity condition can be viewed as an extension of the strict monotonic degree condition in binary-valued networks [**?** ] to general setting. Our $\beta$-monotonicity condition is also closely related to isotonic functions [**?** **?** ] 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 $i \in [d]$. This $\beta$-monotonicity condition allows to estimate the permutation $\pi$ in polynomial-time.

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.

**Algorithm 1** Borda Count algorithm

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

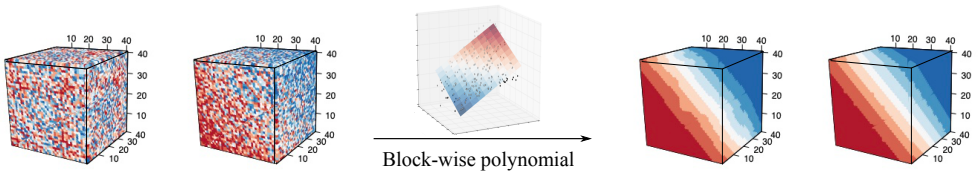
1: **Sorting stage:** Compute a permutation $\hat{\pi}^{\mathrm{BC}}$ such that $\tau \circ (\hat{\pi}^{\mathrm{BC}})^{-1}$ is monotonically increasing, where $\tau(i) = \frac{1}{d^{m-1}} \sum_{(i_2,\ldots,i_m) \in [d]^m} \mathcal{Y}_{i,i_2,\ldots,i_m}$.

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

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

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

Notice that the least square estimation in (4) requires combinatoric search for the permutation resulting in exponential time complexity. However, optimization problem in Algorithm 1 only requires to estimate the degree-$\ell$ polynomial block tensor. Therefore, this step reduces to a degree-$\ell$ polynomial regression problem within each block $\mathcal{E}_k$. The full estimation procedure is illustrated in Figure 2.




(a) Observation (b) Sorted observation (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-$\ell$ polynomial approximation.

We show the consistency of the signal tensor estimation of Borda count estimator.

**Theorem 3** (Estimation error for Borda count)**.** *Suppose that the signal tensor $\Theta$ 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*

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

*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 $\alpha, \beta$, and $m$.*

We find that the Borda count estimator achieves the same minimax-optimal rate as the least-squares estimator for sufficiently smooth tensors under Lipschitz score condition $\beta = 1$. The least-squares estimator requires a combinatoric search with exponential-time complexity. By contrast, the Borda count estimator is polynomial-time solvable. Therefore, Borda count algorithm enjoys both statistical accuracy and computational efficiency.

## 6 Numerical experiments and data application

**Numerical experiments.** We simulate symmetric order-3 $d$-dimensional tensors based on the permuted smooth tensor model (1) with function $f$ in Figure 4a. Notice that considered functions cover a reasonable range of model complexities from low rank to high rank. We generate the entries of the noise tensor i.i.d. from Gaussian distribution $N(0, 0.5^2)$. The permutation $\pi$ is randomly sampled from all permutations. Throughout all experiments, we evaluate the accuracy of the estimation by mean square error (MSE) $= d^{-3}\|\Theta \circ \pi - \hat{\Theta} \circ \hat{\pi}\|_F^2$.

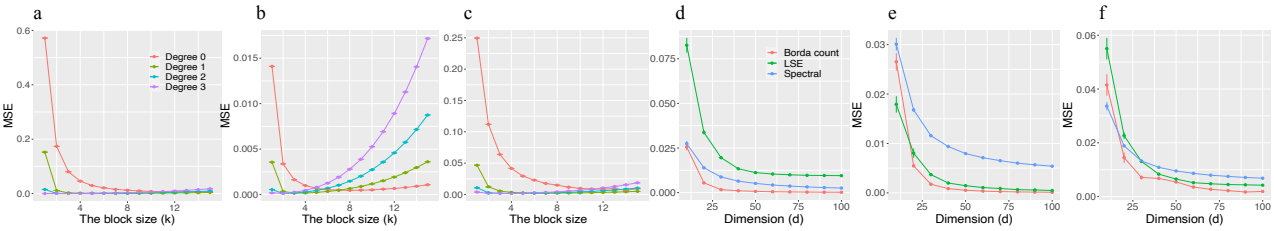



Figure 3: Panels a-c: MSE comparison versus the number of blocks for different polynomial approximation under models 1-3 respectively. Panel d-f: MSE comparison of different methods versus tensor dimension under models 1-3 respectively. MSEs are measured across $n_{\text{sim}} = 20$ replications.

The first experiment examines the impact of the block number $k$ and degree of polynomial $\ell$ for the approximation. We fix the tensor dimension $d = 100$, and vary the number of blocks $k \in \{1, \ldots, 15\}$ and polynomial degree $\ell \in \{0, 1, 2, 3\}$. Figure 3a-c demonstrates the trade-off in accuracy determined by the number of groups for each polynomial degree. In addition, we find that degree-2 polynomial approximation with the optimal $k$ gives the smallest MSE among all considered polynomial approximation. These two observations are well explained by the optimal choice of $(k^*, \ell^*) = (\mathcal{O}(\lceil d^{3/7} \rceil, 2)$ in our theoretical results. The second experiment compares our method (**Borda Count**) with several popular alternative methods: (a) Spectral method (**Spectral**) [**?** ] on unfolded tensor; (b) Least square estimation (**LSE**) [**?** ] with $\ell = 0$; (c) Our **Borda Count** algorithm. 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 **LSE** 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 square optimization (4) is highly non-convex and computationally unstable. The outperformance of **Borda count** demonstrates the efficacy of our method.

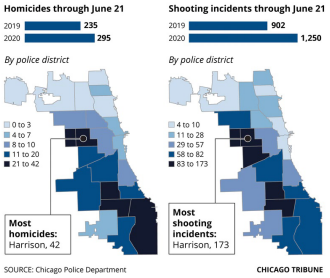
**Applications to Chicago crime data.** Chicago crime tensor dataset is an order-3 tensor with entries representing the log counts of crimes from 24 hours, 77 Chicago community areas, and 32 crime types ranging from January 1st, 2001 to December 11th, 2017. We apply our Borda Count method to Chicago crime dataset. Because the data tensor is asymmetric, we allow different number of blocks across the three modes. 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 4c shows the four areas overlaid on a map of Chicago. Interestingly, we find that the clusters conform the actual locations even though our algorithm did not take any geographic information such as longitude or latitude. In addition, we compare the cluster patterns with benchmark results based on homicides- and shooting incidents-maps in Chicago shown in Figure 4b. We find that our clusters share similar geographical patterns with Figure 4b. The benchmark Figure 4b covers only homicides and shooting incidents in 2020, whereas our result in Figure 4c considers 32 crime types across 2001-2017. The results demonstrate the power of our approach in detecting meaningful pattern from tensor data.

a

| ID | $f(x, y, z)$ | CP rank | Tucker rank |
|---|---|---|---|
| 1 | $xyz$ | 1 | $(1, 1, 1)$ |
| 2 | $\frac{1}{1+\exp(-3x^2+3y^2+3z^2))}$ | 9 | $(4, 4, 4)$ |
| 3 | $\frac{1}{\exp\big(\max(x,y,z)+\sqrt{x}+\sqrt{y}+\sqrt{z}\big)}$ | $\geq 100$ | $(90, 90, 90)$ |

Smooth functions in simulation. We define the numerical CP/Tucker rank as the minimal rank $r$ for which the relative approximation error is below $10^{-4}$. The reported rank in the table is estimated from a $100 \times 100 \times 100$ signal tensor generated by (2).

b




c

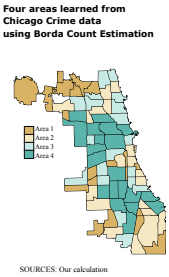



Figure 4: Panel a: smooth functions in simulations. Panels b-c: Chicago crime maps. Panel b shows homicides and shooting incidents in community areas in Chicago. This figure is from *Chicago Tribune* article in 2020 [**?** ]. Panel c shows the four areas estimated by our Borda Count algorithm.

## 7 Conclusion

We have developed permuted smooth tensor model and estimation methods with theoretical guarantees. The efficacy of our procedure is demonstrated through both simulations and analysis of Chicago crime dateset.