SUPPLEMENTARY MATERIALS

Appendix: The appendix includes the proofs and extra simulation results.

Software and data: R-package contains code to perform the methods described in the article. The package also contains all datasets used as examples in the article.

A Extra numerical results

A.1 Results for Models 2 and 4 in Table 1

We first present simulation results for Models 2 and 4 omitted in Section 6. Figure S2 compares the estimation performance among the **Borda count, LSE**, and **Spectral methods**. We find that our Borda count algorithm outperforms others in both models. The first two columns in Figure S3 show the impact of the number of blocks k and degree of polynomial ℓ for the approximation with fixed dimension $d = 100$. Similar to results for Models 1, 3 and 5 in the main paper, we find the optimal k balances the trade-off between approximation error and signal tensor estimation error within each block. The last two columns compare our **Borda count** with other alternative methods. We find that our method still outperforms **LSE** and **Spectral** in all scenarios under Models 2 and 4.

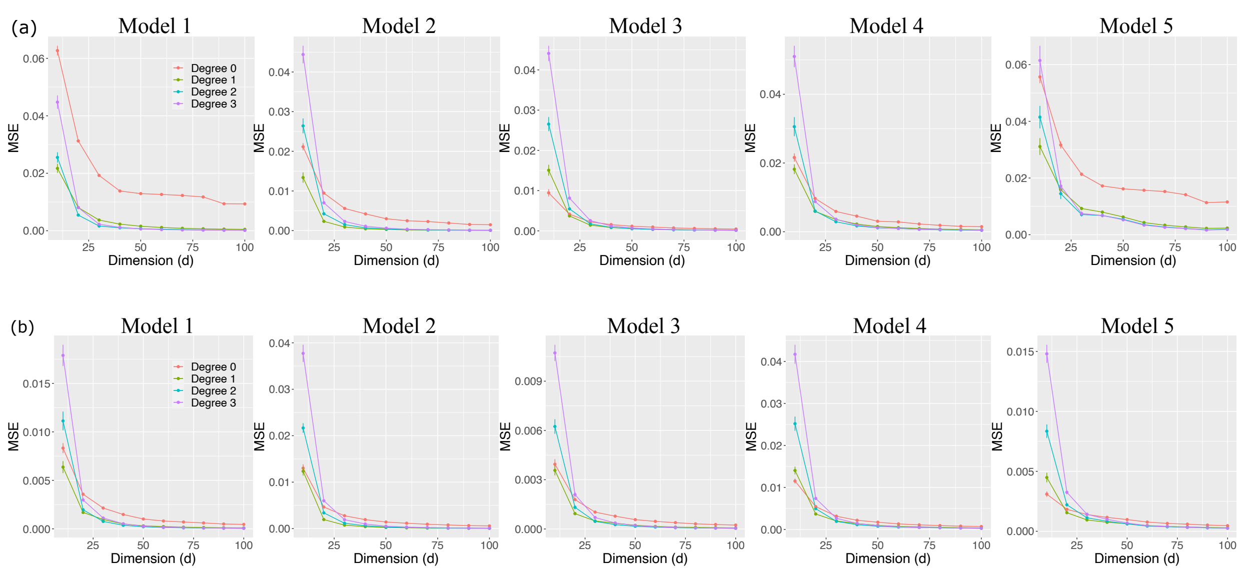

Figure S1: MSE versus the tensor dimension based on different polynomial approximations. Columns 1-5 consider the Models 1-5 in Table 2 respectively. Panel (a) is for continuous tensors, whereas (b) is for the binary tensors.

Suggested by one reviewer, we consider another competing algorithm (denoted as cluster + polyℓ algorithm) for comparison. The algorithm performs clustering first and then

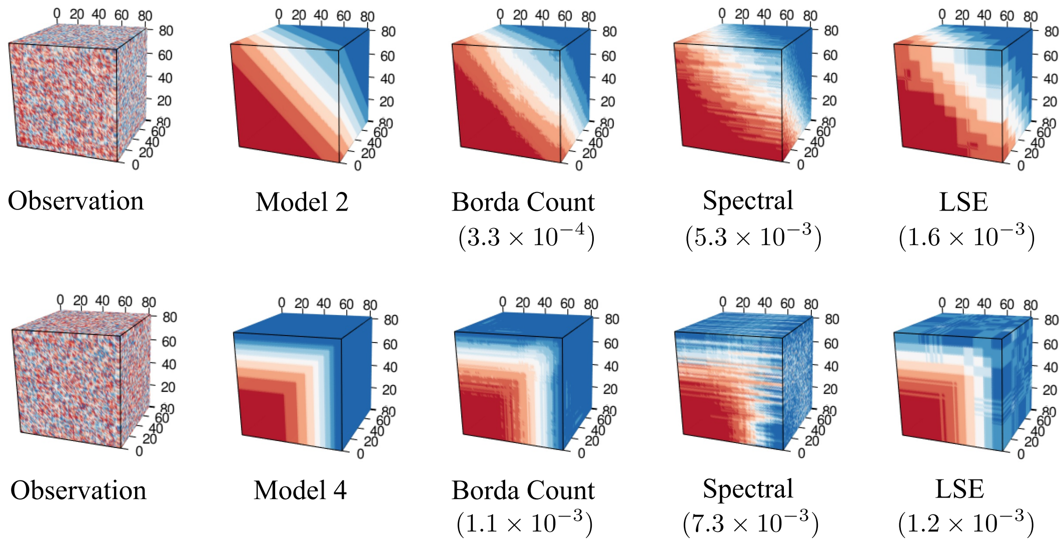

Figure S2: Performance comparison between different methods. The observed data tensors, true signal tensors, and estimated signal tensors are plotted for Models 2 and 4 in Table 2 with fixed dimension $d = 80$. Numbers in parenthesis indicate the mean squared error.

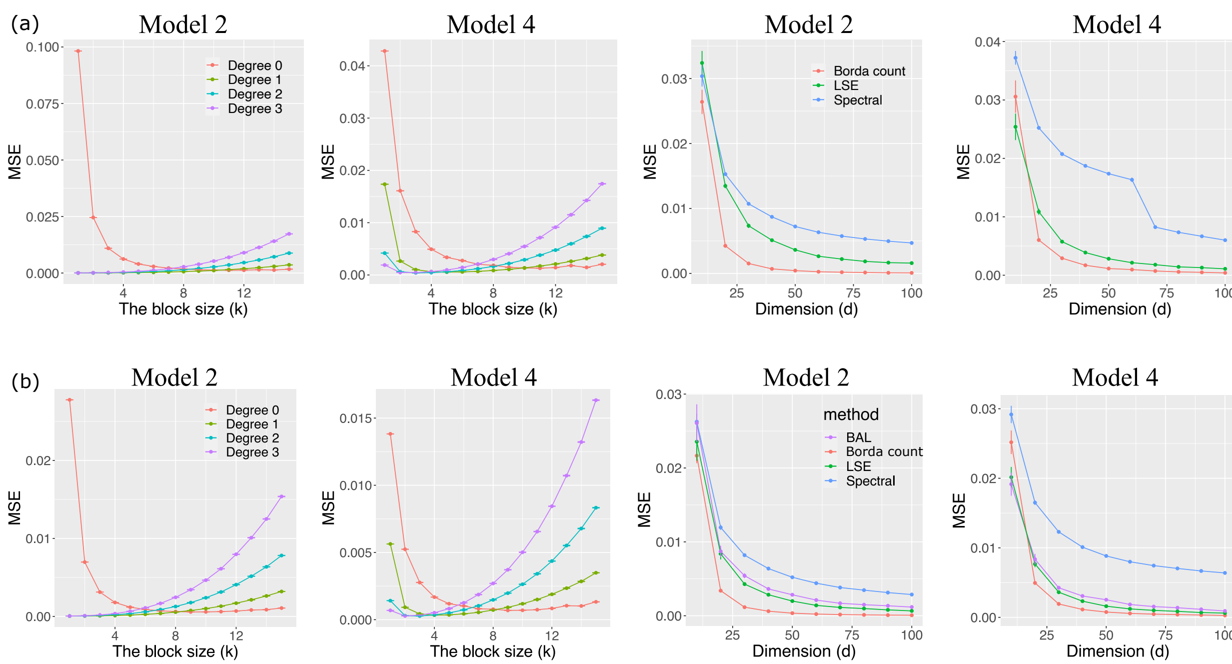

Figure S3: Simulation results for Models 2 and 4 in Table 2. Columns 1-2 plots MSE versus the number of blocks for different polynomial approximation, while Columns 3-4 shows the MSE versus the tensor dimension according to estimation methods. Panel (a) is for continuous tensors, whereas (b) is for the binary tensors.

polynomial approximation within clusters. Since we do not know the permutation, we randomly assign the order of nodes within a block. We perform the simulation using the same setting in Section 6. Figure S4 shows the comparison among all methods, including the new cluster + polyℓ algorithm for $\ell \in \{1, 2, 3\}$. Notice that cluster + poly0 algorithm is equivalent to LSE algorithm. The result shows that the cluster + polyℓ algorithms are unstable except Model 3.

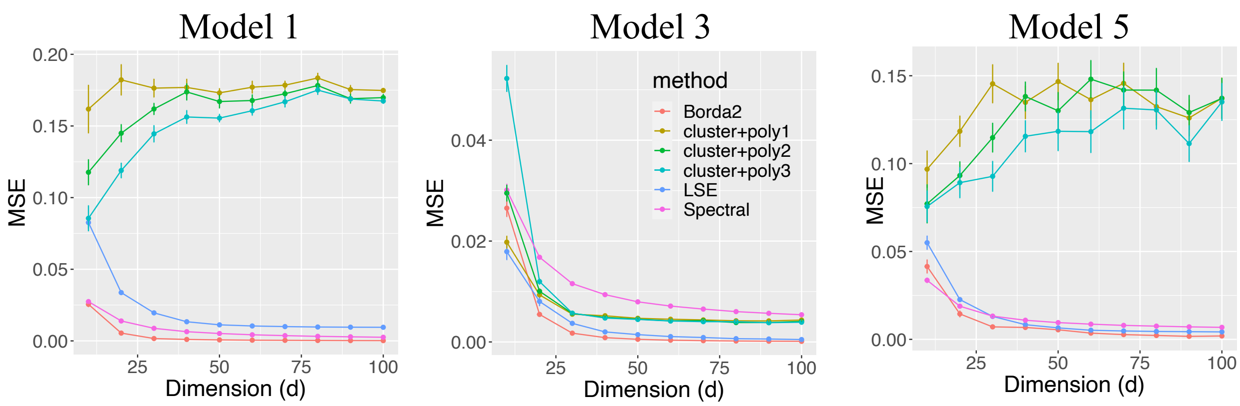

Figure S4: MSE versus the tensor dimension based on different estimation methods. Columns 1-3 consider the Models 1, 3, and 5 with continuous case in Table 2 respectively. cluster+polyℓ means polynomial ℓ-approximation

A.2 Investigation of non-symmetric tensors

Our models and techniques easily extend to non-symmetric tensors. We describe the simulation set-up for non-symmetric tensors and results. We simulate order-3 tensors based on the non-symmetric functions in Table S1.

Model ID	$f(x, y, z)$
1	$xy + z$
2	$x^2 + y + yz^2$
3	$x(1 + \exp(-3(x^2 + y^2 + z^2)))^{-1}$
4	$\log(1 + \max(x, y, z) + x^2 + yz)$
5	$\exp\left(-x - \sqrt{y} - z^3\right)$

Table S1: List of non-symmetric smooth functions in simulation.

We fix the tensor dimension $30 \times 40 \times 50$ and assume that the noise tensors are from Gaussian distribution. Similar to other simulations, we evaluate the accuracy of the estimation by MSE and report the summary statistics across $n_{\text{sim}} = 20$ replicates. The

Method	Model 1	Model 2	Model 3	Model 4	Model 5
Borda count	(2,1,2)	(1,2,2)	(1,3,3)	(2,1,2)	(1,4,4)
LSE	(6,2,3)	(8,5,8)	(6,9,6)	(9,5,6)	(7,9,3)
Spectral	(1,24)	(3,48)	(1,48)	(1,28)	(1,22)

Table S2: Hyperparameters for the methods under Models 1-5 in Table S1. For **Borda count** and **LSE** methods, the values in the table indicate the number of blocks. For **Spectral** method, the first value indicates the tensor unfolding mode, while the second one represents the singular value threshold.

hyperparameters are chosen via cross-validation that give the best accuracy for each method. Table S2 summarizes the choice of hyperparameters.

Table S3 compares the MSEs from repeated simulations based on different methods under Models 1-5 (see Table S1). We find that Borda count estimation outperforms all alternative methods for non-symmetric tensors. The results demonstrate the applicability of our method to general tensors

Method	Model 1	Model 2	Model 3	Model 4	Model 5
Borda count	**0.57 (0.01)**	**0.51 (0.02)**	**0.87 (0.02)**	**1.02 (0.02)**	**2.56 (0.21)**
LSE	23.58 (0.03)	7.70 (0.04)	9.45 (0.05)	3.29 (0.05)	9.93 (0.03)
Spectral	10.76 (0.06)	10.64 (0.05)	6.27 (0.05)	10.90 (0.06)	5.24 (0.04)

Table S3: MSEs from 20 repeated simulations based on different methods. All numbers are displayed on the scales 10^{-3}. Standard errors are reported in parenthesis.

A.3 Extra results on Chicago crime data analysis

We investigate the ten groups of crime types clustered by our method. Table S4 shows that the clustering captures the similar type of crimes. For example, group 2 consists of misdemeanors such as public indecency, non-criminal, and concealed carry license violation, while group 6 represents sex-related offenses such as prostitution, sex offense, and crime sexual assault.

B Details on Example 4

We show that Borda count algorithm works perfectly in the Example 4. For simplicity, assume that d is an even number. The signal tensor is represented by

$$\Theta(i,j,k) = \left(\frac{i}{d} - 0.5\right)^2 + \left(\frac{j}{d}\right)\left(\frac{k}{d}\right), \quad \text{for all } (i,j,k) \in [d]^3.$$

GROUP	I	II	III
CRIME TYPE	RITUALISM, HUMAN TRAFFICKING, OTHER NARCOTIC VIOLATION	PUBLIC INDECENCY, NON-CRIMINAL, CONCEALED CARRY LICENSE VIOLATION	OBSCENITY, STALKING, INTIMIDATION
GROUP	IV	V	VI
CRIME TYPE	KIDNAPPING, GAMBLING, HOMICIDE	LIQUOR LAW VIOLATION, ARSON, INTERFERENCE WITH PUBLIC OFFICER	PROSTITUTION, SEX OFFENSE, CRIM SEXUAL ASSAULT
GROUP	VII	VIII	VIIII
CRIME TYPE	OTHER OFFENSE, CRIMINAL DAMAGE, BATTERY, THEFT, BURGLARY	CRIMINAL TRESPASS, ROBBERY, DECEPTIVE PRACTICE	NARCOTICS, ASSAULT, MOTOR VEHICLE THEFT
GROUP	X		
CRIME TYPE	PUBLIC PEACE VIOLATION, WEAPONS VIOLATION, OFFENSE INVOLVING CHILDREN		

Table S4: Groups of crime types learned based on the Borda count estimation.

We construct permutations π_1, π_2, π_3 and a multivariate function $\bar{f}\colon [0,1]^3 \to \mathbb{R}$ such that

$$\Theta(i,j,k) = \bar{f}\left(\frac{\pi_1(i)}{d}, \frac{\pi_2(j)}{d}, \frac{\pi_3(k)}{d}\right) \pm \frac{1}{d^2}, \quad \text{for all } (i,j,k) \in [d]^3. \tag{23}$$

Define the permutations by

$$\pi_1(i) = \begin{cases} 2i - d, & \text{if } i > \frac{d}{2}, \\ d + 1 - 2i, & \text{if } i \le \frac{d}{2}, \end{cases} \quad \text{and} \quad \pi_2(i) = \pi_3(i) = i. \tag{24}$$

Define the function $\bar{f}\colon [0,1]^3 \to \mathbb{R}$ by

$$\begin{aligned} \bar{f}\colon [0,1]^3 &\to [0,1] \\ (x,y,z) &\mapsto \frac{1}{4}x^2 + yz. \end{aligned} \tag{25}$$

One can verify that $\bar{f}$ is monotonic-and-smooth such that $\bar{f} \in \mathcal{F}(2, 1/4) \cap \mathcal{B}(1/2)$. Furthermore, the construction (24)-(25) satisfies (23), where the perturbation term $1/d^2$ can be absorbed into the approximation error by Proposition 1:

$$\frac{1}{d^m} \inf_{\mathcal{B} \in \mathscr{B}(k,\ell)} \|\Theta - \mathcal{B}\|_F \le \frac{L}{k^{\min(2,\ell+1)}} + \frac{1}{d^2} \le \frac{L+1}{k^{\min(2,\ell+1)}}.$$

In conclusion, the signal tensor constructed in the Example 4 can be regarded as a tensor generated from $\mathcal{F}(2, 5/4) \cap \mathcal{B}(1/2)$. Therefore, our Borda count algorithm is applicable to this case with the claimed accuracy in our paper. This example highlights that our marginal monotonicity assumption is weaker than the usual sense thanks to the unknown permutations.

C Proofs of main theorems

C.1 Proof of Proposition 1

Proof of Proposition 1. We denote $\mathcal{E}_k$ as the m-way partition by

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

where $z\colon [d] \to [k]$ is the canonical clustering function such that $z(i) = \lceil ki/d \rceil$, for all $i \in [d]$, and we use the shorthand $\times_{a=1}^{m}$ to denote the Cartesian product of m sets. For a given $\boldsymbol{j} := (j_1, \ldots, j_m)$ and the related partition $\times_{a=1}^{m} z^{-1}(j_a) \in \mathcal{E}_k$, fix any index $\omega_{\boldsymbol{j}} \in \times_{a=1}^{m} z^{-1}(j_a)$. Then, we have

$$\|\omega - \omega_{\boldsymbol{j}}\|_\infty \leq \frac{d}{k}, \quad \text{for all } \omega \in \bigtimes_{a=1}^{m} z^{-1}(j_a).$$

We define the block-wise degree-ℓ polynomial tensor $\mathcal{B}$ based on the partition $\mathcal{E}_k$ as

$$\mathcal{B}(\omega) = \sum_{\boldsymbol{j}\in[k]^m} \mathrm{Poly}_{\min(\lfloor\alpha\rfloor,\ell),\omega_{\boldsymbol{j}}}\left(\frac{\omega}{d}\right) \mathbb{1}\{\omega \in \bigtimes_{a=1}^{m} z^{-1}(j_a)\},$$

where $\mathrm{Poly}_{\min(\lfloor\alpha\rfloor,\ell),\omega_{\boldsymbol{j}}}$ denotes a degree-ℓ polynomial function defined by

$$\mathrm{Poly}_{\min(\lfloor\alpha\rfloor,\ell),\omega_{\boldsymbol{j}}}(\boldsymbol{x}) = \sum_{\kappa:|\kappa|\leq\min(\lfloor\alpha\rfloor,\ell+1)} \frac{1}{\kappa!}\left(\boldsymbol{x} - \frac{\omega_{\boldsymbol{j}}}{d}\right)^{\kappa} \nabla_\kappa f\left(\frac{\omega_{\boldsymbol{j}}}{d}\right).$$

It follows from Taylor's theorem that

$$\left|f\left(\frac{\omega}{d}\right) - \mathrm{Poly}_{\min(\lfloor\alpha\rfloor,\ell),\omega_{\boldsymbol{j}}}\left(\frac{\omega}{d}\right)\right| \leq L\left\|\frac{\omega-\omega_{\boldsymbol{j}}}{d}\right\|_\infty^{\min(\alpha,\ell+1)} \leq L\left(\frac{1}{k}\right)^{\min(\alpha,\ell+1)}, \tag{26}$$

for all $\omega \in \times_{a=1}^{m} z^{-1}(j_a)$.

Based on the construction of block-wise degree-ℓ polynomial tensor $\mathcal{B}$, we have

$$\begin{aligned}
&\frac{1}{d^m}\|\Theta - \mathcal{B}\|_F^2 \\
&= \frac{1}{d^m}\sum_{\omega\in[d]^m} |\Theta(\omega) - \mathcal{B}(\omega)|^2 \\
&= \frac{1}{d^m}\sum_{\boldsymbol{j}\in[k]^m}\sum_{\omega\in\times_{a=1}^{m} z^{-1}(j_a)} \left|f\left(\frac{\omega}{d}\right) - \mathrm{Poly}_{\min(\lfloor\alpha\rfloor,\ell),\omega_{\boldsymbol{j}}}\left(\frac{\omega}{d}\right)\right|^2 \\
&\leq \frac{L^2}{k^{2\min(\alpha,\ell+1)}},
\end{aligned}$$

where the last inequality uses (26). □

C.2 Proof of Theorem 1

Proof of Theorem 1. By Proposition 1, there exists a block-wise polynomial tensor $\mathcal{B} \in \mathscr{B}(k,\ell)$ such that

$$\|\mathcal{B} - \Theta\|_F^2 \lesssim \frac{L^2 d^m}{k^{2\min(\alpha,\ell+1)}}. \tag{27}$$

By the triangle inequality,

$$\|\hat{\Theta}^{\mathrm{LSE}} \circ \hat{\pi}^{\mathrm{LSE}} - \Theta\circ\pi\|_F^2 \leq 2\|\hat{\Theta}^{\mathrm{LSE}} \circ \hat{\pi}^{\mathrm{LSE}} - \mathcal{B}\circ\pi\|_F^2 + 2\underbrace{\|\mathcal{B}\circ\pi - \Theta\circ\pi\|_F^2}_{\text{Proposition 1}}. \tag{28}$$

Therefore, it suffices to bound $\|\hat{\Theta}^{\text{LSE}} \circ \hat{\pi}^{\text{LSE}} - \mathcal{B} \circ \pi\|_F^2$. By the global optimality of least-square estimator, we have

$$
\begin{aligned}
\|\hat{\Theta}^{\text{LSE}} \circ \hat{\pi}^{\text{LSE}} - \mathcal{B} \circ \pi\|_F &\leq \left\langle \frac{\hat{\Theta}^{\text{LSE}} \circ \hat{\pi}^{\text{LSE}} - \mathcal{B} \circ \pi}{\|\hat{\Theta}^{\text{LSE}} \circ \hat{\pi}^{\text{LSE}} - \mathcal{B} \circ \pi\|_F},\ \mathcal{E} + (\Theta \circ \pi - \mathcal{B} \circ \pi) \right\rangle \\
&\leq \sup_{\pi, \pi' \colon [d] \to [d]} \sup_{\mathcal{B}, \mathcal{B}' \in \mathscr{B}(k,\ell)} \left\langle \frac{\mathcal{B}' \circ \pi' - \mathcal{B} \circ \pi}{\|\mathcal{B}' \circ \pi' - \mathcal{B} \circ \pi\|_F}, \mathcal{E} \right\rangle + \underbrace{\|\mathcal{B} \circ \pi - \Theta \circ \pi\|_F}_{\text{Proposition 1}}.
\end{aligned}
$$

Now we bound inner product term. For fixed π, π', let $\boldsymbol{P}$ and $\boldsymbol{P}'$ be permutation matrices corresponding to permutations π and π' respectively. We express vectorized block-wise degree-ℓ polynomial tensors, $\text{vec}(\mathcal{B})$ and $\text{vec}(\mathcal{B}')$, by discrete polynomial functions. Specifically, denote $\text{vec}(\mathcal{B}) = \boldsymbol{X}\boldsymbol{\beta}$ and $\text{vec}(\mathcal{B}') = \boldsymbol{X}\boldsymbol{\beta}'$, where $\boldsymbol{X} \in \mathbb{R}^{d^m \times k^m(\ell+m)^\ell}$ is a design matrix consisting of m-multivaraite degree-ℓ polynomial basis over grid design $\{1/d, \ldots, d/d\}$ (or $\{x_i\}_{i=1}^d$ for relaxed smooth model (5)), $\boldsymbol{\beta}$ and $\boldsymbol{\beta}' \in \mathbb{R}^{k^m(\ell+m)^\ell}$ are corresponding coefficient vectors. Notice that the number of coefficients for m-multivariate polynomial of degree-ℓ is $\binom{\ell+m}{\ell} \approx (\ell+m)^\ell$; we choose to use $(\ell+m)^\ell$ for each block for notational simplicity. Therefore, we rewrite the inner product by

$$
\begin{aligned}
\left\langle \frac{\mathcal{B}' \circ \pi' - \mathcal{B} \circ \pi}{\|\mathcal{B}' \circ \pi' - \mathcal{B} \circ \pi\|_F}, \mathcal{E} \right\rangle &= \left\langle \frac{(\boldsymbol{P}')^{\otimes m} \text{vec}(\mathcal{B}') - (\boldsymbol{P})^{\otimes m} \text{vec}(\mathcal{B})}{\|(\boldsymbol{P}')^{\otimes m} \text{vec}(\mathcal{B}') - (\boldsymbol{P})^{\otimes m} \text{vec}(\mathcal{B})\|_F}, \mathcal{E} \right\rangle \\
&= \left\langle \frac{(\boldsymbol{P}')^{\otimes m} \boldsymbol{X}\boldsymbol{\beta}' - (\boldsymbol{P})^{\otimes m} \boldsymbol{X}\boldsymbol{\beta}}{\|(\boldsymbol{P}')^{\otimes m} \boldsymbol{X}\boldsymbol{\beta}' - (\boldsymbol{P})^{\otimes m} \boldsymbol{X}\boldsymbol{\beta}\|_F}, \mathcal{E} \right\rangle \\
&= \left\langle \frac{\boldsymbol{A}\boldsymbol{c}}{\|\boldsymbol{A}\boldsymbol{c}\|_2}, \mathcal{E} \right\rangle
\end{aligned}
$$

where we define $\boldsymbol{A} := \begin{pmatrix} \boldsymbol{P}' & -\boldsymbol{P} \end{pmatrix} \begin{pmatrix} \boldsymbol{X} & 0 \\ 0 & \boldsymbol{X} \end{pmatrix} \in \mathbb{R}^{d^m \times 2k^m(\ell+m)^\ell}$ and $\boldsymbol{c} := \begin{pmatrix} \boldsymbol{\beta}' \\ \boldsymbol{\beta} \end{pmatrix} \in \mathbb{R}^{2k^m(\ell+m)^\ell}$.

By Lemma 8, we have

$$
\sup_{\boldsymbol{c} \in \mathbb{R}^{2k^m(\ell+m)^\ell}} \left\langle \frac{\boldsymbol{A}\boldsymbol{c}}{\|\boldsymbol{A}\boldsymbol{c}\|_2}, \mathcal{E} \right\rangle \leq \sup_{\boldsymbol{c} \in \mathbb{R}^{2k^m(\ell+m)^\ell}} \left\langle \frac{\boldsymbol{c}}{\|\boldsymbol{c}\|_2},\ \boldsymbol{e} \right\rangle, \tag{29}
$$

where $\boldsymbol{e} \in \mathbb{R}^{2k^m(\ell+m)^\ell}$ is a sub-Gaussian random vector with variance proxy bounded by σ^2. By the union bound of sub-Gaussian maxima over countable set $\{\pi, \pi' \colon [d] \to [d]\}$, we obtain

$$
\begin{aligned}
&\mathbb{P}\left(\sup_{\pi, \pi' \colon [d] \to [d]} \sup_{\mathcal{B}, \mathcal{B}' \in \mathscr{B}(k,\ell)} \left\langle \frac{\mathcal{B}' \circ \pi' - \mathcal{B} \circ \pi}{\|\mathcal{B}' \circ \pi' - \mathcal{B} \circ \pi\|_F}, \mathcal{E} \right\rangle \geq t \right) \\
&\quad \leq \sum_{\pi, \pi' \in [d]^d} \mathbb{P}\left(\sup_{\mathcal{B}, \mathcal{B}' \in \mathscr{B}(k,\ell)} \left\langle \frac{\mathcal{B}' \circ \pi' - \mathcal{B} \circ \pi}{\|\mathcal{B}' \circ \pi' - \mathcal{B} \circ \pi\|_F}, \mathcal{E} \right\rangle \geq t \right)
\end{aligned}
$$

$$
\begin{aligned}
&\le d^d \mathbb{P}\left(\sup_{\boldsymbol{c}\in\mathbb{R}^{2k^m(\ell+m)^\ell}} \left\langle \frac{\boldsymbol{c}}{\|\boldsymbol{c}\|_2},\ \boldsymbol{e} \right\rangle \ge t \right) \\
&\le \exp\left(-\frac{t^2}{8\sigma^2} + k^m(\ell+m)^\ell \log 6 + d\log d \right), \qquad (30)
\end{aligned}
$$

where the second inequality is from (29) and the last inequality is from Lemma 9. Setting $t = C\sigma\sqrt{k^m(\ell+m)^\ell + d\log d}$ in (30) for a sufficiently large $C > 0$ gives

$$
\sup_{\pi,\pi'\colon [d]\to[d]} \sup_{\mathcal{B},\mathcal{B}'\in\mathscr{B}(k,\ell)} \left\langle \frac{\mathcal{B}'\circ\pi' - \mathcal{B}\circ\pi}{\|\mathcal{B}'\circ\pi' - \mathcal{B}\circ\pi\|_F}, \mathcal{E} \right\rangle \lesssim \sigma\sqrt{k^m(\ell+m)^\ell + d\log d},
$$

with high probability.

Combining the inequalities (27), (28) and (30) yields the desired conclusion

$$
\|\hat{\Theta}^{\mathrm{LSE}} \circ \hat{\pi}^{\mathrm{LSE}} - \Theta\circ\pi\|_F^2 \lesssim \sigma^2\left(k^m(\ell+m)^\ell + d\log d\right) + \frac{L^2 d^m}{k^{2\min(\alpha,\ell+1)}}. \qquad (31)
$$

Finally, optimizing (31) with respect to (k,l) gives that

$$
(31) \lesssim \begin{cases} L^2\left(\frac{\sigma}{L}\right)^{\frac{4\alpha}{m+2\alpha}} d^{-\frac{2m\alpha}{m+2\alpha}}, & \text{when } \alpha < m(m-1)/2, \\ \sigma^2 d^{-(m-1)}\log d, & \text{when } \alpha \ge m(m-1)/2, \end{cases}
$$

under the choice

$$
\ell^* = \min(\lfloor\alpha\rfloor, (m-2)(m+1)/2), \quad k^* = \left\lceil \left(d^m L^2/\sigma^2\right)^{\frac{1}{m+2\min(\alpha,\ell^*+1)}} \right\rceil.
$$

□

C.3 Proof of Theorem 2

Proof of Theorem 2. By the definition of the tensor space, we seek the minimax rate ε^2 in the following expression

$$
\inf_{(\hat{\Theta},\hat{\pi})} \sup_{\Theta\in\mathcal{P}(\alpha,L)} \sup_{\pi\in\Pi(d,d)} \mathbb{P}\left(\frac{1}{d^m}\|\Theta\circ\pi - \hat{\Theta}\circ\hat{\pi}\|_F^2 \ge \varepsilon^2 \right).
$$

On one hand, if we fix a permutation $\pi\in\Pi(d,d)$, the problem can be viewed as a classical m-dimensional α-smooth nonparametric regression with d^m sample points. The minimax lower bound is known to be $\varepsilon^2 = L^2\left(\frac{\sigma}{L}\right)^{\frac{4\alpha}{m+2\alpha}} d^{-\frac{2m\alpha}{m+2\alpha}}$. On the other hand, if we fix $\Theta\in\mathcal{P}(\alpha,L)$, the problem become a new type of convergence rate due to the unknown permutation. We refer to the resulting error as the permutation rate, and we will prove that $\varepsilon^2 = \sigma^2 d^{-(m-1)}\log d$. Since our target is the sum of the two rates, it suffice to prove the two different rates separately. In the following arguments, we will proceed by this strategy.

Nonparametric rate. The nonparametric rate for α-smooth function is readily available in the literature. We cite the results here for self-completeness.

Lemma 1 (Minimax rate for α-smooth function estimation; see Section 2 in Stone [29])**.** *Consider a sample of N data points, $(\boldsymbol{x}_1, Y_1), \ldots, (\boldsymbol{x}_N, Y_N)$, where $\boldsymbol{x}_n = (\frac{i_1}{d}, \ldots, \frac{i_m}{d}) \in [0,1]^m$ is the m-dimensional predictor and $Y_n \in \mathbb{R}$ is the scalar response for $n \in [N]$. Consider the observation model*

$$Y_n = f(\boldsymbol{x}_n) + \varepsilon_n, \quad \text{with } \varepsilon_n \sim \text{i.i.d. } \mathcal{N}(0,1), \quad \text{for all } n \in [N].$$

Assume f is in the α-Hölder smooth function class, denoted by $\mathcal{F}(\alpha, L)$. Then,

$$\inf_{\hat{f}} \sup_{f \in \mathcal{F}(\alpha,L)} \mathbb{P}\left(\|f - \hat{f}\|_2 \geq \sigma^{\frac{4\alpha}{m+2\alpha}} L^{\frac{2m}{m+2\alpha}} N^{-\frac{2\alpha}{m+2\alpha}} \right) \geq 0.9.$$

Our desired nonparametric rate readily follows from Lemma 1 by taking sample size $N = d^m$ and function norm $\|f - \hat{f}\|_2 = \frac{1}{d^m}\|\Theta - \hat{\Theta}\|_F^2$. In summary, for a given permutation $\pi \in \Pi(d, d)$, we have

$$\inf_{\hat{\Theta}} \sup_{\Theta \in \mathcal{P}(\alpha,L)} \mathbb{P}\left(\frac{1}{d^m}\|\hat{\Theta} \circ \pi - \Theta \circ \pi\|_F^2 \geq L^2 \left(\frac{\sigma}{L}\right)^{\frac{4\alpha}{m+2\alpha}} d^{-\frac{2m\alpha}{m+2\alpha}} \right) \geq 0.9. \tag{32}$$

Permutation rate. Since nonparametric rate dominates permutation rate when $\alpha \leq 1$, it is sufficient to prove the permutation rate lower bound for $\alpha \geq 1$. We first show the minimax permutation rate for k-block degree-0 tensor family $\mathcal{B}(k, 0)$, and then construct a smooth $f \in \mathcal{F}(\alpha, L)$ to mimic the constant block tensors.

Let $\Pi(d, k)$ denote the collection of all possible onto mappings from $[d]$ to $[k]$. Lemma 2 shows the permutation rate over k-block degree-0 tensor family $\mathcal{B}(k, 0)$ is $\sigma^2 d^{-(m-1)} \log k$.

Lemma 2 (Permutation error for tensor block model)**.** *Consider the problem of estimating d-dimensional, block-k signal tensors from the Gaussian tensor block model. For every given integer $k \in [d]$, there exists a core tensor $\mathcal{S} \in \mathbb{R}^{k \times \cdots \times k}$ satisfying*

$$\inf_{\hat{\Theta}} \sup_{z \in \Pi(d,k)} \mathbb{P}\left(\frac{1}{d^m}\|\hat{\Theta} - \mathcal{S} \circ z\|_F^2 \gtrsim \frac{\sigma^2 \log k}{d^{m-1}} \right) \geq 0.9. \tag{33}$$

The proof of Lemma 2 is constructive and deferred to Section G. The core tensor $\mathcal{S}$ in Lemma 2 has a special pattern of zero's that will be used in the proof of Theorem 2; see the Section G for details.

To prove the permutation rate, we fix a core tensor $\mathcal{S} \in \mathbb{R}^{k \times \cdots \times k}$ satisfying (33), and use it to construct the smooth tensor. we construct a function $f \in \mathcal{F}(\alpha, L)$ that mimics the core tensor $\mathcal{S}$ in block tensor family $\mathcal{B}(k, 0)$. For notational simplicity, we do not distinguish the fractional number and its rounding integer. For example, we simply write $d/2$ (instead of

$\lceil d/2 \rceil$ or $\lfloor d/2 \rfloor$) to represent its rounding integer. Define $k = d^\delta$ for some $\delta \in (0,1)$, which will be specified later. Consider a smooth function $K(x)$ that is infinitely differentiable,

$$K(x) = C_k \exp\left(-\frac{1}{1-64x^2}\right) \mathbb{1}\left\{|x| < \frac{1}{8}\right\},$$

where $C_k > 0$ satisfies $\int K(x)dx = 1$. Then, we define a smooth convolution function as

$$\psi(x) = \int_{-3/8}^{3/8} K(x-y)dy.$$

The smooth convolution function has support $[-1/2, 1/2]$ and takes value 1 on the interval $[-1/4, 1/4]$. For a given core tensor $\mathcal{S}$ in Lemma 2, we define an α-smooth function

$$f(x_1, \ldots, x_m) = \sum_{(a_1,\ldots,a_m)\in[k]^m} \left(\mathcal{S}(a_1,\ldots,a_m) - \frac{1}{2}\right) \prod_{i=1}^m \psi\left(kx_i - a_i + \frac{1}{2}\right) + \frac{1}{2}. \quad (34)$$

One can verify that $f \in \mathcal{F}(\alpha, L)$ as long as we choose sufficiently small δ depending on α and L. Notice that for any $(a_1, \ldots, a_m) \in [k]^m$,

$$f(x_1,\ldots,x_m) = \mathcal{S}(a_1,\ldots,a_m), \quad \text{if } (x_1,\ldots,x_m) \in \bigtimes_{i=1}^m \left[\frac{a_i - 3/4}{k}, \frac{a_i - 1/4}{k}\right]. \quad (35)$$

From this observation, we define a sub-domain $I \subset [d]$ such that

$$I = \left(\bigcup_{a=1}^k \left[\frac{d(a-3/4)}{k}, \frac{d(a-1/4)}{k}\right]\right) \bigcap [d]. \quad (36)$$

We have that $|I| = d/2$ by definition. Let $\Theta(\mathcal{S}) \in \mathbb{R}^{d\times\cdots\times d}$ denote the tensor induced by f in (34). We use subscript I to denote objects when restricted in the indices set I. For example, $\Theta_I(\mathcal{S}) \in \mathbb{R}^{d/2\times\cdots\times d/2}$ denotes the sub-tensor with indices in I, and $\|\cdot\|_{F,I}$ denotes the sum of squares over indices in I. Based on (35) and (36), $\Theta_I(\mathcal{S})$ has block structure with the core tensor $\mathcal{S}$. We use $\Pi(d/2, d/2)$ to denote the set of all permutations on I while fixing indices on $[d]\backslash I$; that is, $\Pi(d/2, d/2) = \{\pi\colon I \to I\} \cong \{\pi \in \Pi(d,d)\colon \pi(i) = i \text{ for } i \in [d]\backslash I\}$). Then, we have

$$\begin{aligned}
&\inf_{(\hat\Theta,\hat\pi)} \sup_{\pi\in\Pi(d,d)} \mathbb{P}\left(\frac{1}{d^m}\|\hat\Theta\circ\hat\pi - \Theta(\mathcal{S})\circ\pi\|_F^2 \geq \varepsilon^2\right)\\
\overset{(*)}{=}\ &\inf_{\hat\Theta} \sup_{\pi\in\Pi(d,d)} \mathbb{P}\left(\frac{1}{d^m}\|\hat\Theta - \Theta(\mathcal{S})\circ\pi\|_F^2 \geq \varepsilon^2\right)\\
\geq\ &\inf_{\hat\Theta} \sup_{\pi\in\Pi(d,d)} \mathbb{P}\left(\frac{1}{d^m}\|\hat\Theta - \Theta(\mathcal{S})\circ\pi\|_{F,I}^2 \geq \varepsilon^2\right)\\
\overset{(**)}{\geq}\ &\inf_{\hat\Theta} \sup_{\pi\in\Pi(d/2,d/2)} \mathbb{P}\left(\frac{1}{(d/2)^m}\|\hat\Theta_I - \Theta_I(\mathcal{S})\circ\pi\|_F^2 \geq 2^m\varepsilon^2\right)
\end{aligned}$$

$$= \inf_{\hat{\Theta}} \sup_{z\in\Pi(d/2,k)} \mathbb{P}\left(\frac{1}{(d/2)^m}\|\hat{\Theta}_I - \mathcal{S}\circ z\|_F^2 \geq 2^m\varepsilon^2\right), \tag{37}$$

where $(*)$ absorbs the estimate $\hat{\pi}$ into the estimate $\hat{\Theta}$, and $(**)$ uses the constructed function (34) and the permutation collections $\Pi(d/2, d/2)$. Based on the construction of $\mathcal{S}$ in Lemma 2, the cross terms in $(\mathcal{S}\circ z)(I, I^c, \ldots)$ are zero. Therefore, we reduce the problem of estimating $\pi\colon [d] \to [d]$ to estimating $z : I \to [k]$ in the sub-tensor. Applying Lemma 2 to (37) by using $d/2$ in the place of d and $k = d^\delta$ for a constant $\delta > 0$ yields the desired conclusion.

Combining two rates. Now, we combine (32) and (37) to get the desired lower bound. For any Θ generated as in (4) with $f \in \mathcal{F}(\alpha, L)$, by union bound, we have

$$\begin{aligned}
&\mathbb{P}\left\{\frac{1}{d^m}\|\hat{\Theta}-\Theta\|_F^2 \gtrsim L^2\left(\frac{\sigma}{L}\right)^{\frac{4\alpha}{m+2\alpha}} d^{-\frac{2m\alpha}{m+2\alpha}} + \frac{\sigma^2\log d}{d^{m-1}}\right\}\\
\geq\ &\mathbb{P}\left\{\frac{1}{d^m}\|\hat{\Theta}-\Theta\|_F^2 \gtrsim L^2\left(\frac{\sigma}{L}\right)^{\frac{4\alpha}{m+2\alpha}} d^{-\frac{2m\alpha}{m+2\alpha}}\right\} + \mathbb{P}\left\{\frac{1}{d^m}\|\hat{\Theta}-\Theta\|_F^2 \gtrsim \frac{\sigma^2\log d}{d^{m-1}}\right\} - 1.
\end{aligned}$$

Taking sup on both sides with the property

$$\sup_{\substack{\Theta\in\mathcal{P}(\alpha,L)\\ \pi\in\Pi(d,d)}} (f(\pi)+g(\Theta)) = \sup_{\pi\in\Pi(d,d)} f(\pi) + \sup_{\Theta\in\mathcal{P}(\alpha,L)} g(\Theta)$$

yields the desired rate (12). □

C.4 Proof of Theorem 3

The proof of Theorem 3 leverages results of hypergraphic planted clique and constant higher-order clustering problems. We first briefly explain the constant higher-order clustering problems. We then prove the main result.

C.4.1 Constant higher-order clustering and computational lower bound

Let $\boldsymbol{k} = (k_1, \ldots, k_m)$ and $\boldsymbol{d} = (d_1, \ldots, d_m)$. We introduce the constant high-order clustering (CHC) problem [24]. Consider a data tensor $\mathcal{Y} \in \mathbb{R}^{d_1\times\cdots\times d_m}$ generated from the signal plus noise model

$$\mathcal{Y} = \Theta + \mathcal{E}, \tag{38}$$

where the entries in $\mathcal{E}$ are i.i.d. drawn from Gaussian distribution, and the signal tensor Θ contains a constant planted structure:

$$\Theta \in \Theta_{\text{CHC}}(\boldsymbol{k}, \boldsymbol{d}, \lambda) := \{\lambda' \mathbb{1}_{I_1}\otimes\cdots\otimes\mathbb{1}_{I_m} : |I_i| = k_i,\ \text{for all } i\in[m],\ \lambda'\geq\lambda\}.$$

Here $I_i \subset [d_i]$ denotes a subset of indices, $|\cdot|$ denotes the cardinality of the set, $\mathbb{1}_{I_i}$ is the d_i-dimensional indicator vector such that $(\mathbb{1}_{I_i})_j = 1$ if $j \in I_i$ and 0 otherwise. The CHC detection problem is to test the following hypothesis based on the observed tensor $\mathcal{Y}$,

$$H_0\colon \Theta = 0 \quad \text{v.s.} \quad H_1\colon \Theta \in \Theta_{\mathrm{CHC}}(\boldsymbol{k}, \boldsymbol{d}, \lambda). \tag{39}$$

The following proposition provides the asymptotic regime for impossible polynomial-time detection of CHC under Conjecture 1. This proposition plays important role to prove our Theorem 3.

Proposition 2 (Theorem 15 in [24]). *Consider CHC detection problem in* (39) *in the Gaussian noise model* (38) *under the asymptotic regime $d \to \infty$ satisfying*

$$d = d_1 = \cdots = d_m, \quad k = k_1 = \cdots = k_m = d^{\delta}, \quad \lambda = d^{-\gamma},$$

with $0 \le \delta \le 1$ and $\gamma > (m\delta - m/2) \vee 0$. Then, under Conjecture 1, for any polynomial-time test sequence $\{\phi\}_d \colon \mathcal{Y} \mapsto \{0,1\}$, we have

$$\liminf_{d\to\infty} \left\{ \mathbb{P}_{\mathcal{H}_0}(\phi(\mathcal{Y}) = 1) + \sup_{\Theta \in \Theta_{\mathrm{CHC}}(\boldsymbol{k},\boldsymbol{d},\lambda)} \mathbb{P}_{\Theta}(\phi(\mathcal{Y}) = 0) \right\} \ge \frac{1}{2}.$$

C.4.2 Proof of Theorem 3

Proof of Theorem 3. Assume that the true signal $\Theta \in \mathbb{R}^{d\times\cdots\times d}$ has constant planted structure with a given $\boldsymbol{k} = (k, \ldots, k)$ such that

$$\Theta \in \Theta_{\mathrm{CHC}}(\boldsymbol{k}, \boldsymbol{d}, \lambda) = \{\lambda \mathbb{1}_I \otimes \cdots \otimes \mathbb{1}_I \colon |I| = k\}.$$

We have $\Theta_{\mathrm{CHC}} \subset \mathcal{P}_{\mathrm{rex}}$, because we can set the infinitely smooth function $f : [0,1]^m \to [0,1]$ by

$$f(x_1, \ldots, x_m) = \lambda \prod_{i\in[m]} x_i. \tag{40}$$

Then, there is one-to-one correspondence between tensors in $\Theta_{\mathrm{CHC}}(\boldsymbol{k}, \boldsymbol{d}, \lambda)$ and tensors generated by the above f under the choice $x_i = \mathbb{1}_{i\in I}$ for all $i \in [m]$.

We consider the regime where polynomial-time solvable test is impossible based on Proposition 2. We set $\delta = 1/2$, $k = c_1 d^{\delta}$, and $\lambda = c_2 d^{-\gamma}$ for any fixed $\gamma \in (0, \frac{2\alpha - m}{m})$, so that any polynomial-time test sequence ϕ satisfies

$$\liminf_{d\to\infty} \left\{ \mathbb{P}_{\mathcal{H}_0}(\phi(\mathcal{Y}) = 1) + \sup_{\Theta \in \Theta_{\mathrm{CHC}}(\boldsymbol{k},\boldsymbol{d},\lambda)} \mathbb{P}_{\Theta}(\phi(\mathcal{Y}) = 0) \right\} \ge \frac{1}{2}.$$

The choice of λ is possible given that $\alpha > m/2$. Notice that the choice $\lambda \lesssim O(1)$ ensures the function (40) satisfies the definition (1) for all $\alpha > 0$.

We prove by contradiction. Assume that there exists a hypothetical estimator $\hat{\Theta}$ from a polynomial-time algorithm that attains the rate $\mathrm{Rate}(d)$. Specifically, there exists a constant $b > 0$ such that

$$\limsup_{d\to\infty} \frac{1}{\mathrm{Rate}(d)} \sup_{\Theta\in\Theta_{\mathrm{CHC}}(\boldsymbol{k},\boldsymbol{d},\lambda)} \frac{1}{d^m}\mathbb{E}\|\hat{\Theta}-\Theta\|_F^2 \le b. \tag{41}$$

By Markov's inequality, the inequality (41) implies that, when d is sufficiently large, for all $\Theta \in \Theta_{\mathrm{CHC}}(\boldsymbol{k},\boldsymbol{d},\lambda)$ and all $u > 0$, we have

$$\|\hat{\Theta}-\Theta\|_F \le u\sqrt{\mathrm{Rate}(d)d^m}, \tag{42}$$

with probability at least $1 - b/u$. Consider the hypothesis test in (39). We employ the following test

$$\phi(\mathcal{Y}) = \mathbb{1}(\|\hat{\Theta}\|_F \ge u\sqrt{\mathrm{Rate}(d)d^m}).$$

The Type I error of the test ϕ is controlled by

$$\mathbb{P}_0(\|\hat{\Theta}\|_F \ge u\sqrt{\mathrm{Rate}(d)d^m}) = \mathbb{P}_0(\|\hat{\Theta}-\Theta\|_F \ge u\sqrt{\mathrm{Rate}(d)d^m}) \le b/u.$$

For Type II error, we obtain,

$$\begin{aligned}
\sup_{\Theta\in\Theta_{\mathrm{CHC}}(k,d,\lambda)} \mathbb{P}_\Theta(\phi(\mathcal{Y})=0) &= \sup_{\Theta\in\Theta_{\mathrm{CHC}}(k,d,\lambda)} \mathbb{P}_\Theta(\|\hat{\Theta}\|_F < u\sqrt{\mathrm{Rate}(d)d^m}) \\
&\le \sup_{\Theta\in\Theta_{\mathrm{CHC}}(k,d,\lambda)} \mathbb{P}_\Theta(\|\hat{\Theta}-\Theta\|_F^2 > \|\Theta\|_F^2 - u^2\mathrm{Rate}(d)d^m) \\
&\overset{(*)}{\le} \sup_{\Theta\in\Theta_{\mathrm{CHC}}(\boldsymbol{k},\boldsymbol{d},\lambda)} \mathbb{P}_\Theta(\|\hat{\Theta}-\Theta\|_F^2 > u^2\mathrm{Rate}(d)d^m) \\
&\overset{(**)}{\le} b/u.
\end{aligned}$$

The inequality $(*)$ holds because

$$\|\Theta\|_F^2 \ge \lambda^2 k^m = c_1^m c_2^2 d^{\frac{m}{2}-\gamma} \ge 2u^2\mathrm{Rate}(d)d^m \asymp d^{\frac{m}{2}-\frac{2\alpha-m}{m}}$$

where the last inequality is true under the regime $c_1^m c_2 > 2u^2$. We can always choose constants c_1 and c_2 given the value u. The inequality $(**)$ holds because of the statement (42). Putting Type I and II errors together, we obtain

$$\mathbb{P}_{\mathcal{H}_0}(\phi(\mathcal{Y})=1) + \sup_{\Theta\in\Theta_{\mathrm{CHC}}(\boldsymbol{k},\boldsymbol{d},\lambda)} \mathbb{P}_\Theta(\phi(\mathcal{Y})=0) \le 2b/u < 1/2,$$

for $u > 4b$. This fact contradicts the Proposition 2. Therefore, there is no polynomial-time $\hat{\Theta}$ satisfying (41). □

C.5 Proof of Theorem 4

We first present a lemma to show the estimation error of $\hat{\pi}$. The exponent β measures the difficulty for permutation estimation. We find that a larger β guarantees a faster consistency rate of $\hat{\pi}^{\mathrm{BC}}$, which is represented below 2.

Lemma 3 (Permutation error)**.** *Consider the sub-Gaussian tensor model* (3) *with* $f \in \mathcal{M}(\beta)$. *With high probability, we have*

$$\mathrm{Loss}(\pi, \hat{\pi}^{\mathrm{BC}}) := \frac{1}{d} \max_{i\in[d]} |\pi(i) - \hat{\pi}^{\mathrm{BC}}(i)| \lesssim \left(d^{-(m-1)/2}\sqrt{\log d}\right)^{\beta}.$$

The proof of Lemma 3 is provided in Section G.

Proof of Theorem 4. By Proposition 1, there exists a block-wise polynomial tensor $\mathcal{B} \in \mathscr{B}(k,\ell)$ satisfying (27). By the triangle inequality, we decompose estimation error into three terms,

$$\begin{aligned} &\|\hat{\Theta}^{\mathrm{BC}} \circ \hat{\pi}^{\mathrm{BC}} - \Theta \circ \pi\|_F \\ &\leq \|\hat{\Theta}^{\mathrm{BC}} \circ \hat{\pi}^{\mathrm{BC}} - \mathcal{B} \circ \hat{\pi}^{\mathrm{BC}}\|_F + \|\mathcal{B} \circ \hat{\pi}^{\mathrm{BC}} - \Theta \circ \hat{\pi}^{\mathrm{BC}}\|_F + \|\Theta \circ \hat{\pi}^{\mathrm{BC}} - \Theta \circ \pi\|_F \\ &= \underbrace{\|\hat{\Theta}^{\mathrm{BC}} - \mathcal{B}\|_F}_{\text{Nonparametric error}} + \underbrace{\|\Theta \circ \hat{\pi}^{\mathrm{BC}} - \Theta \circ \pi\|_F}_{\text{Permutation error}} + \underbrace{\|\mathcal{B} - \Theta\|_F}_{\text{Proposition 1}}. \end{aligned} \tag{43}$$

Therefore, it suffices to bound two terms $\|\Theta \circ \hat{\pi}^{\mathrm{BC}} - \Theta \circ \pi\|_F$ and $\|\hat{\Theta}^{\mathrm{BC}} - \mathcal{B}\|_F$ separately.
Permutation error. For any $(i_1, \ldots, i_m) \in [d]^m$, we have

$$\begin{aligned} &\left|\Theta\left(\hat{\pi}^{\mathrm{BC}}(i_1), \ldots, \hat{\pi}^{\mathrm{BC}}(i_m)\right) - \Theta\left(\pi(i_1), \ldots, \pi(i_m)\right)\right| \\ &\leq \left\|\left(\frac{\hat{\pi}^{\mathrm{BC}}(i_1)}{d}, \ldots, \frac{\hat{\pi}^{\mathrm{BC}}(i_m)}{d}\right) - \left(\frac{\pi(i_1)}{d}, \ldots, \frac{\pi(i_m)}{d}\right)\right\|_\infty^{\min(\alpha,1)} \\ &\leq \left[\frac{1}{d}\max_{i\in[d]}\left|\hat{\pi}^{\mathrm{BC}}(i) - \pi(i)\right|\right]^{\min(\alpha,1)} \\ &\lesssim \left(\sigma d^{-(m-1)/2}\sqrt{\log d}\right)^{\beta \min(\alpha,1)}, \end{aligned}$$

where the first inequality is from the α-Hölder smoothness of Θ, and the last inequality is from Lemma 3. Therefore, we obtain the upper bound of the permutation error

$$\frac{1}{d^m}\|\Theta \circ \hat{\pi}^{\mathrm{BC}} - \Theta \circ \pi\|_F^2 \lesssim \left(\sigma^2 \frac{\log d}{d^{m-1}}\right)^{\beta \min(\alpha,1)}. \tag{44}$$

Nonparametric error. Recall that Borda count estimation is defined by $\hat{\Theta}^{\mathrm{BC}} := \arg\min_{\Theta \in \mathscr{B}(k,\ell)} \|\tilde{\mathcal{Y}} - \Theta\|_F^2$, where $\tilde{\mathcal{Y}} = \mathcal{Y} \circ (\hat{\pi}^{\mathrm{BC}})^{-1}$. By the optimality of least-square estimator, we have

$$\|\hat{\Theta}^{\mathrm{BC}} - \mathcal{B}\|_F \leq \left\langle \frac{\hat{\Theta}^{\mathrm{BC}} - \mathcal{B}}{\|\hat{\Theta}^{\mathrm{BC}} - \mathcal{B}\|_F},\ \mathcal{Y} \circ \pi \circ (\hat{\pi}^{\mathrm{BC}})^{-1} - \mathcal{B} \right\rangle$$

$$
\begin{aligned}
&\equiv \left\langle \frac{\hat{\Theta}^{\mathrm{BC}} - \mathcal{B}}{\|\hat{\Theta}^{\mathrm{BC}} - \mathcal{B}\|_F}, \mathcal{E} + \left(\Theta \circ \pi \circ (\hat{\pi}^{\mathrm{BC}})^{-1} - \mathcal{B}\right) \right\rangle \\
&\leq \sup_{\mathcal{B},\mathcal{B}' \in \mathscr{B}(k,\ell)} \left\langle \frac{\mathcal{B}' - \mathcal{B}}{\|\mathcal{B} - \mathcal{B}\|_F}, \mathcal{E} \right\rangle + \|\Theta \circ \pi - \mathcal{B} \circ \hat{\pi}^{\mathrm{BC}}\|_F \\
&\leq \sup_{\mathcal{B},\mathcal{B}' \in \mathscr{B}(k,\ell)} \left\langle \frac{\mathcal{B}' - \mathcal{B}}{\|\mathcal{B} - \mathcal{B}\|_F}, \mathcal{E} \right\rangle + \underbrace{\|\Theta \circ \pi - \Theta \circ \hat{\pi}^{\mathrm{BC}}\|_F}_{\text{Permutation error (44)}} + \underbrace{\|\Theta - \mathcal{B}\|_F}_{\text{Proposition 1}}
\end{aligned}
$$

Now we bound the inner product term. By the same argument in the proof of Theorem 1, the space embedding $\mathscr{B}(k,\ell) \subset \mathbb{R}^{(\ell+m)^\ell k^m}$ implies the space embedding $\{(\mathcal{B}' - \mathcal{B}) \colon \mathcal{B}, \mathcal{B}' \in \mathscr{B}(k,\ell)\} \subset \mathbb{R}^{2(\ell+m)^\ell k^m}$. Therefore, we have

$$
\sup_{\mathcal{B},\mathcal{B}' \in \mathscr{B}(k,\ell)} \left\langle \frac{\mathcal{B}' - \mathcal{B}}{\|\mathcal{B}' - \mathcal{B}\|_F}, \mathcal{E} \right\rangle \leq \sup_{\boldsymbol{c} \in \mathbb{R}^{2k^m(\ell+m)^\ell}} \left\langle \frac{\boldsymbol{c}}{\|\boldsymbol{c}\|_2},\ e \right\rangle, \tag{45}
$$

where $e \in \mathbb{R}^{2k^m(\ell+m)^\ell}$ is a sub-Gaussian random vector with variance proxy σ^2. Combining (45) and Lemma 9 yields

$$
\begin{aligned}
\mathbb{P}\left(\sup_{\mathcal{B},\mathcal{B}' \in \mathscr{B}(k,\ell)} \left\langle \frac{\mathcal{B}' - \mathcal{B}}{\|\mathcal{B}' - \mathcal{B}\|_F}, \mathcal{E} \right\rangle \geq t \right) &\leq \mathbb{P}\left(\sup_{\boldsymbol{c} \in \mathbb{R}^{2k^m(\ell+m)^\ell}} \left\langle \frac{\boldsymbol{c}}{\|\boldsymbol{c}\|_2},\ e \right\rangle \geq t \right) \\
&\leq \exp\left(-\frac{t^2}{8\sigma^2} + k^m(\ell+m)^\ell \log 6 \right),
\end{aligned}
$$

Setting $t = C\sigma\sqrt{k^m(\ell+m)^\ell}$ for sufficiently large $C > 0$ gives

$$
\sup_{\mathcal{B},\mathcal{B}' \in \mathscr{B}(k,\ell)} \left\langle \frac{\mathcal{B}' - \mathcal{B}}{\|\mathcal{B}' - \mathcal{B}\|_F}, \mathcal{E} \right\rangle \lesssim \sigma\sqrt{k^m(\ell+m)^\ell}, \tag{46}
$$

with high probability.

Finally, combining all sources of error from Proposition 1 and inequalities (44), (46), (43) yields

$$
\frac{1}{d^m}\|\hat{\Theta}^{\mathrm{BC}} \circ \hat{\pi}^{\mathrm{BC}} - \Theta \circ \pi\|_F \lesssim \left(\sigma^2 \frac{\log d}{d^{m-1}}\right)^{\beta \min(\alpha,1)} + \sigma^2 \frac{k^m(\ell+m)^\ell}{d^m} + \frac{L^2}{k^{2\min(\alpha,\ell+1)}}. \tag{47}
$$

Finally, optimizing (47) with respect to (k,l) gives that

$$
(47) \ \lesssim \begin{cases} L^2 \left(\frac{\sigma}{L}\right)^{\frac{4\alpha}{m+2\alpha}} d^{-\frac{2m\alpha}{m+2\alpha}}, & \text{when } \alpha < c(\alpha,\beta,m), \\ \left(\frac{\sigma^2 \log d}{d^{m-1}}\right)^{\beta \min(\alpha,1)}, & \text{when } \alpha \geq c(\alpha,\beta,m), \end{cases}
$$

under the choice

$$
\ell^* = \min\left(\lfloor \alpha \rfloor, \lfloor c(\alpha,\beta,m) \rfloor\right), \quad k^* = c_1 d^{m/(m+2\min(\alpha,\ell^*+1))},
$$

where $c(\alpha,\beta,m) := \frac{m(m-1)\beta\min(\alpha,1)}{\max(0,\ 2(m-(m-1)\beta\min(\alpha,1)))}$.

$\square$

C.6 Proof of Theorem 5

We prove a stronger result than Theorem 5 in the main paper. Specifically, we show that the conclusion (21) holds for either 1-monotonic in Definition 2 or isotonic functions defined in (22). Let $\mathcal{M}$ denote either 1-monotonic or isotonic function class. Define the monotonic Lipschitz tensor class

$$\mathcal{P}_{M,L} = \left\{\Theta \in \mathbb{R}^{d\times\cdots\times d}\colon \Theta(\omega) = f\left(\frac{\omega}{d}\right) \text{ for all } \omega = (i_1,\ldots,i_m) \in [d]^m \text{ and } f \in \mathcal{H}(1,L)\cap\mathcal{M}\right\}.$$

We will show the estimation problem under the Gaussian model (1) obeys the minimax lower bound

$$\inf_{(\hat{\Theta},\hat{\pi})}\ \sup_{\Theta\in\mathcal{P}_{M,L},\pi\in\Pi(d,d)} \mathbb{P}\left(\text{MSE}(\hat{\Theta}\circ\hat{\pi},\ \Theta\circ\pi) \gtrsim d^{-\frac{2m}{m+2}} + d^{-(m-1)}\log d\right) \geq 0.8.$$

Proof of Theorem 5. We prove the nonparametric rate and permutation rate separately.

Nonparametric rate. We first show the nonparametric rate for a given permutation $\pi \in \Pi(d,d)$

$$\inf_{\hat{\Theta}}\ \sup_{\Theta\in\mathcal{P}_{M,L}} \mathbb{P}\left(\text{MSE}(\hat{\Theta}\circ\hat{\pi},\ \Theta\circ\pi) \geq cd^{-\frac{2m}{m+2}}\right) \geq 0.9. \tag{48}$$

We introduce the following function

$$K(x_1,\ldots,x_m) = \prod_{i=1}^m (1-2|x_i|)\mathbb{1}\{|x_i|\leq 1/2\}.$$

Let us take $k = \lceil c_1 d^{\frac{m}{2+m}}\rceil$ for some constant $c_1>0$ to be determined later. For any $\boldsymbol{a} = (a_1,\ldots,a_m) \in [k]^m$, define the function

$$\phi_{\boldsymbol{a}}(x_1,\ldots,x_m) = Mk^{-1}K\left(kx_1 - a_1 + \frac{1}{2},\ldots,kx_m - a_m + \frac{1}{2}\right), \tag{49}$$

where $M>0$ is a constant only depending on L. By the construction of the function, we have the following lemma

Lemma 4. *For some $M>0$, depending on L, the function* (49) *satisfies*

1. $\phi_{\boldsymbol{a}}(x_1,\ldots,x_m) \in \mathcal{F}(1,L)$
2. $\sum_{(i_1,\ldots,i_m)\in[d]^m} \phi_{\boldsymbol{a}}(i_1/d,\ldots,i_m/d) \geq \frac{M^2}{3^m} d^m k^{-2-m}$

The proof of Lemma 4 is deferred to Section G. Let $\Gamma = \{0,1\}^N$ be the set of all binary sequences of length $N = \binom{k+m-1}{m}$. For any $\gamma = \{\gamma_{\boldsymbol{a}}\}_{1\leq a_1\leq\cdots\leq a_m\leq k} \in \Gamma$, we define the symmetric function h^γ by

$$h^\gamma(x_1,\ldots,x_m) = \frac{1}{2} + \sum_{1\leq a_1\leq\cdots\leq a_m\leq k} \gamma_{\boldsymbol{a}}\phi_{\boldsymbol{a}}(x_1,\ldots,x_m), \quad \text{for } x_1\geq\cdots\geq x_m.$$

We further define monotonically increasing function given ω by

$$f^{\gamma} = g + \delta h^{\gamma},$$

where $g(x_1, \ldots, x_m) = \frac{1}{2} + \frac{1}{m}\sum_{i=1}^{m} x_i$. We choose small constant $\delta > 0$ so that f^{γ} is a monotonic function for all $\gamma \in \Gamma$ with respect to both isotonic and 1-monotonic definition.

Notice that f^{γ} is continuous and has repeating structure across each block $\bigtimes_{i=1}^{m}\left[\frac{a_i-1}{k}, \frac{a_i}{k}\right]$. It suffices tho check the monottonicity for any $(x_1, x_2, \ldots, x_m) \in \bigtimes_{i=1}^{m}\left[\frac{a_i-1}{k}, \frac{a_i}{k}\right]$ where $a_i \in [k]$ for all $i \in [m]$. For fixed $(x_2, \ldots, x_m)$, we have

$$f^{\gamma}(x_1, x_2, \ldots, x_m) = \frac{1+\delta}{2} + \frac{1}{m}\sum_{i=1}^{m} x_i + \frac{\delta\gamma_{\boldsymbol{a}} M}{k}\left(1 - 2\left|kx_1 - a_1 + \frac{1}{2}\right|\right)\prod_{i=2}^{m}\left(1 - 2\left|kx_i - a_i + \frac{1}{2}\right|\right)$$

Since $\gamma_{\boldsymbol{a}} = 0$ gives trivial result, we assume that $\gamma_{\boldsymbol{a}} = 1$. When $x_1 \in \left[\frac{(a_1-1)}{k}, \frac{(a_1-1/2)}{k}\right]$, we have linear function whose coefficient greater than $1/m$ with respect to x_1. When $x_1 \in \left[\frac{(a_1-1/2)}{k}, \frac{a_1}{k}\right]$, we have

$$f^{\gamma}(x_1, \ldots, x_m) = \frac{1+\delta}{2} + \frac{1}{m}x_1 - \delta M C_1 x_1 + C_2,$$

where C_1 and C_2 are constant only depending on m, a_1 and $(x_1, \ldots, x_m)$. Therefore, choosing sufficiently small constant $\delta > 0$ always makes f^{γ} monotonic with respect to both isotonic and 1-montonic definition. Since both g and h^{γ} are Lipschitz, we show that $f^{\gamma} \in \mathcal{F}(1, L) \cap \mathcal{M}$ for all $\gamma \in \Gamma$.

Now we consider the subspace $\mathcal{F} = \{f^{\gamma} \colon \gamma \in \Gamma\} \subset \mathcal{F}(1, L) \cap \mathcal{M}$. To apply Lemma 11, we first upper bound $\sup_{f, f' \in \mathcal{F}} D(\mathbb{P}_f, \mathbb{P}_{f'})$. For any $f \in \mathcal{F}$, denote $f(\omega) = f(i_1/d, \ldots, i_m/d)$ where $\omega = (i_1/d, \ldots, i_m/d)$. Then for any $f, f' \in \mathcal{F}$, we have

$$\begin{aligned}
D(\mathbb{P}_f | \mathbb{P}_{f'}) &\leq \frac{1}{2\sigma^2}\sum_{\omega \in [d]^m}(f(\omega) - f'(\omega))^2 \\
&\leq \frac{1}{2\sigma^2}\sum_{\omega \in [d]^m}\delta^2(h(\omega) - h'(\omega))^2 \\
&\leq \frac{1}{2\sigma^2}\delta^2 M^2 d^m k^{-2}.
\end{aligned}$$

where the first inequality holds by Lemma 12.

Next we lower bound the packing number of $\mathcal{F}$. For any $f^{\gamma}, f^{\gamma'} \in \mathcal{F}$, we have

$$\begin{aligned}
\rho^2(f^{\gamma}, f^{\gamma'}) &\geq \frac{1}{d^m}\sum_{\omega \in [d]^m}\sum_{1 \leq a_1 \leq \cdots \leq a_m \leq k}\delta^2(\gamma_{\boldsymbol{a}} - \gamma'_{\boldsymbol{a}})^2\phi_{\boldsymbol{a}}^2(\omega) \\
&\geq \frac{\delta^2 M^2}{3^m}k^{-2-m}\rho_H(\gamma, \gamma'),
\end{aligned}$$

where the last inequality uses Lemma 4 and ρ_H is defined in Lemma 10. By Lemma 10, we select a subset $S \subset \Gamma$ such that $|S| \geq \exp(k^m/16)$ and $\rho_H(\gamma, \gamma') \geq k^m/8$ for any $\gamma \neq \gamma' \in S$. Therefore setting $\epsilon^2 = ck^{-2\alpha}$ for some constant $c > 0$, we have

$$\log \mathcal{M}(\epsilon, \mathcal{F}, \rho) \geq k^m/16.$$

By the choice of $k = \lceil c_1 d^{\frac{m}{2+m}} \rceil$ and Lemma 11, we have

$$\inf_{\hat{\Theta}} \sup_{f \in \mathcal{F}} \mathbb{P}\left(\frac{1}{d^m} \sum_{\omega \in [d]^m} \left(\hat{\Theta}(\omega) - f(\omega/d) \right)^2 \geq c d^{-\frac{2m}{m+2}} \right) \geq 0.9,$$

which completes the proof of (48).

Permutation rate. Similar to the proof of Theorem 2, we first show the minimax permutation rate for k-block isotonic tensor family, and then construct a monotonic (with respect to both isotonic and 1-monotonic definition) and a function $f \in \mathcal{F}(1, L) \cap \mathcal{M}(1)$ to mimic the constant isotonic block tensors.

The following lemma serves the same role as Lemma 2, except that now we has an isotonic tensor $\mathcal{S} \in \mathbb{R}^{k \times \cdots \times k}$ satisfying (33). The proof of Lemma 5 is constructive and deferred to the end of this subsection.

Lemma 5 (Permutation error for isotonic tensor block model)**.** *Consider the problem of estimating d-dimensional, block-k signal tensors from Gaussian tensor block models. For every given integer $k \in [d]$, there exists an isotonic core tensor $\mathcal{S} \in \mathbb{R}^{k \times \cdots \times k}$ satisfying*

$$\inf_{\hat{\Theta}} \sup_{z \in \Pi(d,k)} \mathbb{P}\left\{ \frac{1}{d^m} \sum_{(i_1, \ldots, i_m) \in [d]^m} \left[\hat{\Theta}(i_1, \ldots, i_m) - \mathcal{S}\left(z(i_1), \ldots, z(i_m)\right) \right]^2 \gtrsim \frac{\sigma^2 \log k}{d^{m-1}} \right\} \geq 0.9.$$

The remaining proof is the same as Theorem 2, and we omit the proof here for brevity. $\square$

D Sharp extension of Theorem 4 to isotonbic functions

We provide a sharp upper bound under isotonic assumption similar to Theorem 4. Define isotonic function class $\mathcal{M}$

$$\mathcal{M} = \{ f \colon [0,1]^m \to \mathbb{R} \mid f(x_1, \ldots, x_m) \leq f(x_1', \ldots, x_m') \text{ when } x_i \leq x_i' \text{ for } i \in [m] \}.$$

Unlike β-monotonic functions, the permutation π per se may not be accurately estimated for isotonic functions. However, we find that the composition $\Theta \circ \pi$ can still be accurately estimated due to the joint monotonicity.

Lemma 6 (Permutation error of Borda count algorithm under isotonic assumption). *Consider the sub-Gaussian tensor model* (3) *with* $f \in \mathcal{M}$. *Let* $\hat{\pi}$ *be the permutation such that the permuted empirical score function* $\tau \circ \hat{\pi}^{-1}$ *is monotonically increasing. Then with high probability*

$$\frac{1}{d^m}\|\Theta \circ \hat{\pi} - \Theta \circ \pi\|_F^2 \lesssim \frac{\log d}{d^{m-1}}.$$

The proof of Lemma 6 is provided in Section G. Now, we obtain the same statistical accuracy of the Borda count estimator for the isotonic functions as in Theorem 4.

Theorem 6 (Estimation error for Borda count algorithm under isotonic assumption). *Consider the sub-Gaussian tensor model with* $f \in \mathcal{H}(\alpha, L) \cap \mathcal{M}$. *Let* $(\hat{\Theta}^{\mathrm{BC}}, \hat{\pi}^{\mathrm{BC}})$ *be the Borda count estimator in* (16)-(17) *with* $\ell^* = \min(\lfloor \alpha \rfloor, (m-2)(m+1)/2)$ *and* $k^* \asymp d^{m/(m+2\min(\alpha, \ell^*+1))}$. *Then, we have in high probability,*

$$(9) \ \lesssim \begin{cases} d^{-\frac{2m\alpha}{m+2\alpha}}, & \text{when } \alpha < m(m-1)/2, \\ d^{-(m-1)}\log d, & \text{when } \alpha \geq m(m-1)/2. \end{cases}$$

Proof of Theorem 6. The proof of Theorem 6 is the exactly same as in Theorem 4, except that we now use Lemma 6 in place of Lemma 3. We omit the proof for brevity. □

E Challenges for extending Theorem 5 to arbitrary α

Our Theorem 5 is based on Lipschitz mononoe ($\alpha = \beta = 1$). Here we provide a high level explanation for the technical challenges of constructing minimax lower bound for the arbitrary $\alpha > 0$ under monotonicity assumption.

Recall the proof of minimax lower bound for arbitrary α without monotonicity in Theorem 2. The main idea is to fix the permutation first and construct a subset of $\mathcal{F}(\alpha, L)$ to get nonparametric rate. Then we fix the smooth function and let the permutation vary to get the permutation rate. Finally we combine two rates to obtain the minimax lower bound. To obtain the nonparametric and permutation rates, we use Fano's method in Lemma 11 via a finite subset of the function class $\mathcal{F}(\alpha, L)$ for each case. Technical challenges come from constructing a subset of the α-smooth function class with extra monotonicity structure. We cannot directly use function classes constructed for Theorem 2.

Challenge in nonparametric rates. In the general proof of the nonparametric rate, we partition the domain into k^m uniform blocks and consider a set of α-smooth functions $\{\phi_{\boldsymbol{a}}\}_{\boldsymbol{a} \in [k]^m}$ whose support is one of the partition in $[0,1]^m$. The function class consists of functions defined as

$$h^{\gamma}(x_1, \ldots, x_m) = \frac{1}{2} + \sum_{1 \leq a_1 \leq \cdots \leq a_m \leq k} \gamma_{\boldsymbol{a}} \phi_{\boldsymbol{a}}(x_1, \ldots, x_m), \quad \text{for } x_1 \geq \cdots \geq x_m,$$

where $\gamma_{\boldsymbol{a}} \in \{0,1\}^{\binom{k+m-1}{m}}$.

However, the extra monotonicity constraints make this function class non-applicable because h^γ is not monotonic in general. One possible remedy is to construct the function class of the form

$$f^\gamma = g + \delta h^\gamma,$$

where $g(x_1,\ldots,x_m) = 1/2 + \frac{1}{m}\sum_{i=1}^m x_i$ and δ is small constant not depending on tensor dimension d. This new construction allows to have moderate sized function class while keeping both α-smoothness and monotonicity. When $\alpha \geq 1$, we are able to set small constant δ to make the function f^γ α-smooth and monotonic. When $0 < \alpha < 1$, however, we find that δ cannot be a constant but depends on k (or equivalently, on d) if f^γ is α-smooth. This dependency of δ severely worsens the minimax lower bound resulting in the loose bound. How to construct such function classes is unknown.

Challenge in permutation rate. In permutation rate, we construct the smooth function that mimics the constant block tensor over $d/2$ points. This construction reduces our problem to that of estimating a block constant tensor. Then we obtain the permutation rate applying minimax lower bound of estimating constant block tensor (Lemma 2). Since we now consider the monotonic α-smooth function, we construct the smooth function that mimics the isotonic constant block tensor and provide minimax lower bound Lemma 5 corresponding to Lemma 2. The issue we have is how to construct the monotonic α-smooth function that mimics the isotonic constant tensor block over $d/2$ points. Without monotonic constraint, we use the infinitely differentiable function called smooth convolution function whose support is a partition of $[0,1]^m$. This allows the constructed function to have constant block structure over at least $d/2$ points, smoothly connecting the other points.

However, the monotonicity constraint makes the application of the smooth convolution function impossible because the smooth convolution functions are not monotonic. We overcome this challenge using piece-wise linear functions when $\alpha = 1$. We construct Lipschitz monotonic function where $d/2$ points mimic isotonic constant block tensor and the other half points are connected by piecewise linear function. How to construct the monotonic smooth function for general $\alpha \neq 1$ is unknown.

F Extension to relaxed smooth models

We now extend the grid design (4) to random design (5). Specifically, suppose that the signal tensor is generated from the following random design model

$$\Theta(i_1,\ldots,i_m) = f(x_{i_1},\ldots,x_{i_m}), \tag{50}$$

where the sequence $\{x_i\}_{i=1}^d$ are i.i.d. random variables sampled from a probability distribution supported on $[0,1]$.

We briefly discuss the extension of Theorem 1 to random design. The proof of Theorem 1 is based on Proposition 1 and the union bound over $(\mathcal{B},\pi)$. For random designs, we change the union bound over the set $\{\pi\colon [d]\to[d]\}$ to union bound over the set $\{(x_1,\ldots,x_d): x_i\in[0,1] \text{ for all } i\in[d]\}$. These two sets have the same order of complexity up to log terms. Therefore, it suffices to present the extension of Proposition 1 to random designs.

Lemma 7 (Block-wise polynomial tensor approximation under random design)**.** *Consider the random design model* (50)*. For every block number $k\leq d$ and degree $\ell\in\mathbb{N}$, we have the approximation error*

$$\inf_{\mathcal{B}\in\mathscr{B}(k,\ell)}\frac{1}{d^m}\mathbb{E}\|\Theta-\mathcal{B}\|_F^2\lesssim\frac{L^2}{k^{2\min(\alpha,\ell+1)}}.$$

Proof of Lemma 7. The proof is similar to that of Proposition 1 except that we now construct the block-wise degree-ℓ polynomial tensor $\mathcal{B}$ conditional on the random variables $\{x_i\}_{i\in[d]}$. In Proposition 1, we use the canonical clustering function such that $z(i)=\lceil ki/d\rceil$ for all $i\in[d]$. Here, we use the clustering function $z^*\colon[d]\to[k]$ conditional on $\{x_i\}_{i\in[d]}$ such that

$$z^*(i)=\lceil k*x_i\rceil.$$

The remaining proof is exactly the same as in Proposition 1 and thus omitted. $\square$

G Technical lemmas

Proof of Lemma 2. We provide the proof for $m=3$ only. The extension to higher orders ($m\geq 4$) uses exactly the same techniques and thus is omitted. For notational simplicity, we do not distinguish the fractional number and its rounding integer. For example, we simply write $k/3$, instead of $\lceil k/3\rceil$, to represent its rounding integer.

Let us pick $\omega_1,\ldots,\omega_{k/3}\in\{0,1\}^{k^2/9}$ such that $\rho_H(\omega_p,\omega_q)\geq k^2/36$ for all $p\neq q\in[k/3]$. This selection is possible by lemma 10. Fixing such $\omega_1,\ldots,\omega_{k/3}$, we define a symmetric core tensor $\mathcal{S}\in\mathbb{R}^{k\times k\times k}$ for $p<q<r$,

$$\mathcal{S}(p,q,r)=\begin{cases} s_{p,q,r} & \text{if } p\in\{1,\ldots,k/3\}, q\in\{k/3+1,\ldots,2k/3\},\\ & \quad r\in\{2k/3+1,\ldots,k\},\\ 0 & \text{Otherwise},\end{cases}$$

where $\{s_{p,q,r}\colon p\in\{1,\ldots,k/3\}, q\in\{k/3+1,\ldots,2k/3\}, r\in\{2k/3+1,\ldots,k\}\}$ satisfies

$$\boldsymbol{s}(r):=\text{vec}\left(\mathcal{S}\left(1:\frac{k}{3},\ \left(\frac{k}{3}+1\right):\frac{2k}{3},\ r\right)\right)$$

$$= \sqrt{\frac{c\sigma^2 \log k}{d^2}}\omega_{r-2k/3} \quad \text{for any } r \in \{2k/3+1, \dots, k\}. \tag{51}$$

The choice of constant $c > 0$ is deferred to a later part of the proof. Notice that for any $r_1, r_2 \in \{2k/3+1, \dots, k\}$, we have

$$\|\boldsymbol{s}(r_1) - \boldsymbol{s}(r_2)\|_F^2 \geq \frac{c\sigma^2 k^2 \log k}{36d^2}. \tag{52}$$

Define a subset of permutation set $\Pi(d,k)$ by

$$\mathcal{Z} = \left\{ z \in \Pi(d,k) \colon |z^{-1}(a)| = \frac{d}{k} \text{ for all } a \in [k], z^{-1}(a) = \left\{ \frac{(a-1)d}{k} + 1, \dots, \frac{ad}{k} \right\} \text{ for } a \in [2k/3] \right\}.$$

Each $z \in \mathcal{Z}$ induces a block structure in $\mathcal{B}(k,0)$, with known positions for the $2k/3$ indices and unknown positions for the remaining $k/3$ indices. We consider the collection of block tensors induced by $\mathcal{Z}$ and $\mathcal{S}$; i.e.,

$$\mathcal{B}(\mathcal{Z}) = \{\Theta^z \in \mathbb{R}^{d\times d\times d} \colon \Theta^z(i,j,k) = \mathcal{S}(z(i), z(j), z(k)) \text{ for } z \in \mathcal{Z}\}.$$

To apply Lemma 11, we seek for the upper bound for $\sup_{\Theta,\Theta' \in \mathcal{B}(\mathcal{Z})} D(\mathbb{P}_\Theta | \mathbb{P}_{\Theta'})$ and the lower bound for $\log \mathcal{M}(\epsilon, \mathcal{B}(\mathcal{Z}), \rho)$, where ρ is defined by $\rho(\Theta, \Theta') = \frac{1}{d^3}\|\Theta - \Theta'\|_F^2$. For the Gaussian model, we have

$$D(\mathbb{P}_\Theta | \mathbb{P}_{\Theta'}) \leq \frac{1}{2\sigma^2}\|\Theta - \Theta'\|_F \leq \frac{1}{2\sigma^2} d^3 \frac{c\sigma^2 \log k}{d^2} = \frac{cd \log k}{2}, \tag{53}$$

where the first inequality holds for any $\Theta, \Theta' \in \mathcal{B}(\mathcal{Z})$ by Lemma 12.

Next we provide a lower bound of the packing number $\log \mathcal{M}(\epsilon, \mathcal{B}(\mathcal{Z}), \rho)$ with $\epsilon^2 \asymp \frac{\sigma^2 \log k}{d^2}$. From the construction of $\mathcal{S}$ in (51), we have one-to-one correspondence between $\mathcal{Z}$ and $\mathcal{B}(\mathcal{Z})$. Thus $\mathcal{M}(\epsilon, \mathcal{B}(\mathcal{Z}), \rho) = \mathcal{M}(\epsilon, \mathcal{Z}, \rho')$ for the metric ρ' on $\mathcal{Z}$ defined by $\rho'(z_1, z_2) = \rho(\Theta^{z_1}, \Theta^{z_2})$. Let P be the packing set in $\mathcal{Z}$ with the same cardinality of $\mathcal{M}(\epsilon, \mathcal{Z}, \rho')$. Given any $z \in \mathcal{Z}$, define its ϵ-neighbor by $\mathcal{N}(z, \epsilon) = \{z' \in \mathcal{Z} \colon \rho'(z, z') \leq \epsilon\}$. Then, we have $\cup_{z\in P}\mathcal{N}(z,\epsilon) = \mathcal{Z}$, because the cardinality of P is same as packing number $\mathcal{M}(\epsilon, \mathcal{Z}, \rho')$. Therefore, we have

$$|\mathcal{Z}| \leq \sum_{z \in P} |\mathcal{N}(z,\epsilon)| \leq |P| \max_{z\in P} |\mathcal{N}(z,\epsilon)|. \tag{54}$$

It remains to find the upper bound of $\max_{z\in P}|\mathcal{N}(z,\epsilon)|$. For any $z_1, z_2 \in \mathcal{Z}$, $z_1(i) = z_2(i)$ for $i \in [2d/3]$ and $|z_1^{-1}(p)| = d/k$ for all $p \in [k]$. Therefore,

$$\begin{aligned}
\rho'^2(z_1, z_2) &\geq \frac{1}{d^3} \sum_{1\leq i_1 \leq d/3 < i_2 \leq 2d/3 \leq i_3 \leq d} \left(\mathcal{S}\left(z_1(i_1), z_1(i_2), z_1(i_3)\right) - \mathcal{S}\left(z_2(i_1), z_2(i_2), z_2(i_3)\right)\right)^2 \\
&= \frac{1}{d^3} \sum_{2n/3 < i_3 \leq n} \sum_{1 \leq p \leq k/3 < q \leq 2k/3} \sum_{i_1 \in z_1^{-1}(p), i_2 \in z_1^{-1}(q)} \left(\mathcal{S}\left(p, q, z_1(i_3)\right) - \mathcal{S}\left(p, q, z_2(i_3)\right)\right)^2
\end{aligned}$$

$$
\begin{aligned}
&= \frac{1}{d^3} \sum_{2n/3 < i_3 \le n} \sum_{1 \le p \le k/3 < q \le 2k/3} \left(\frac{d}{k}\right)^2 \left(\mathcal{S}\left(p, q, z_1(i_3)\right) - \mathcal{S}\left(p, q, z_2(i_3)\right)\right)^2 \\
&= \frac{1}{d^3} \sum_{2n/3 < i_3 \le n} \left(\frac{d}{k}\right)^2 \|\boldsymbol{s}(z_1(i_3) - \boldsymbol{s}(z_2(i_3))\|_F^2 \\
&\ge \frac{c\sigma^2 \log k}{36 d^3} |\{j \colon z_1(j) \neq z_2(j)\}|,
\end{aligned}
$$

where the last inequality is from (52). Hence with the choice of $\epsilon^2 = \frac{c\sigma^2 \log k}{288 d^2}$, we have $|\{j \colon z(j) \neq z'(j)\}| \le d/8$ for any $z' \in \mathcal{N}(z, \epsilon)$. This implies

$$
|\mathcal{N}(z, \epsilon)| \le \binom{d}{d/8} k^{d/8} \le (8e)^{d/8} k^{d/8} \le \exp\left(\frac{1}{5} d \log k\right), \tag{55}
$$

for sufficiently large k. Next we find the lower bound of $|\mathcal{Z}|$ based on Stirling's formula,

$$
|\mathcal{Z}| = \frac{(d/3)!}{[(d/k)!]^{k/3}} = \exp\left(\frac{1}{3} d \log k + o(d \log k)\right) \ge \exp\left(\frac{1}{4} d \log k\right). \tag{56}
$$

Plugging (55) and (56) into (54) yields

$$
\mathcal{M}(\epsilon, \mathcal{B}(\mathcal{Z}), \rho) = |P| \ge \frac{\max_{z \in P} \mathcal{N}(z, \epsilon)}{|\mathcal{Z}|} \ge \exp\left(\frac{1}{20} d \log k\right). \tag{57}
$$

Finally, applying Lemma 11 based on (53) and (57) gives

$$
\inf_{\hat{\Theta}} \sup_{\Theta \in \mathcal{B}(\mathcal{Z})} \mathbb{P}\left(\frac{1}{d^3} \|\hat{\Theta} - \Theta\|_F^2 \ge \frac{C\sigma^2 \log k}{d^2}\right) = \inf_{\hat{\Theta}} \sup_{z \in \mathcal{Z}} \mathbb{P}\left(\frac{1}{d^3} \|\hat{\Theta} - \mathcal{S} \circ z\|_F^2 \ge \frac{C\sigma^2 \log k}{d^2}\right) \ge 0.9,
$$

with some constant $C > 0$ for sufficiently small $c > 0$ in (51). □

Proof of Lemma 3. Without loss of generality, assume that π is the identity permutation. Notice that $g(i) - \tau(i)$ is the sample average of d^{m-1} independent mean-zero Gaussian random variables with the variance proxy σ^2. Based on the maxima of sub-Gaussian random variables, we have

$$
\max_{i \in [d]} |g(i) - \tau(i)| < 2\sigma d^{-(m-1)/2} \sqrt{\log d}, \tag{58}
$$

with probability $1 - \frac{2}{d^2}$.

By the β-monotonicity of the function g, we have

$$
g(1) \le g(2) \le \cdots \le g(d-1), \tag{59}
$$

The estimated permutation $\hat{\pi}$ is defined for which

$$
\tau \circ \hat{\pi}^{-1}(1) \le \tau \circ \hat{\pi}^{-1}(2) \le \cdots \le \tau \circ \hat{\pi}^{-1}(d-1) \le \tau \circ \hat{\pi}^{-1}(d). \tag{60}
$$

For any given index i, we examine the error $|i - \hat{\pi}(i)|$. By (59) and (60), we have

$$i = \underbrace{|\{j\colon g(j) \le g(i)\}|}_{=:\mathrm{I}}, \quad \text{and} \quad \hat{\pi}(i) = \underbrace{|\{j\colon \tau(j) \le \tau(i)\}|}_{=:\mathrm{II}},$$

where $|\cdot|$ denotes the cardinality of the set. We claim that the sets I and II differ only in at most $d^{(m-1)\beta/2}$ elements. To prove this, we partition the indices in $[d]$ in two cases.

1. Long-distance indices in $\{j\colon |j-i|/d \ge C\left(\sigma d^{-(m-1)/2}\sqrt{\log d}\right)^{\beta}\}$ for some sufficient large constant $C > 0$. In this case, the ordering of (i, j) remains the same in (59) and (60), i.e.,

$$g(i) < g(j) \Longleftrightarrow \tau(i) < \tau(j). \tag{61}$$

   We only prove the right side direction in (61) here. The other direction can be similarly proved. Suppose that $g(i) < g(j)$. Then we have

$$\begin{aligned} \tau(j) - \tau(i) &\ge -|g(j) - \tau(j)| - |g(i) - \tau(i)| + g(j) - g(i) \\ &> -4\sigma d^{-(m-1)/2}\sqrt{\log d} + g(j) - g(i) \\ &\ge 0, \end{aligned}$$

   where the second inequality is from (58) with probability at least $(1 - 2/d^2)^d$ and the last inequality uses β-monotonicity of $g(\cdot)$, and the assumption $|j - i|/d \ge C\left(\sigma d^{-(m-1)/2}\sqrt{\log} d\right)^{\beta}$. Therefore we show that $g(i) < g(j)$ implies $\tau(i) < \tau(j)$. In this case, we conclude that none of long-distance indices belongs to IΔII.

2. Short-distance indices in $\{j\colon |j-i|/d < \left(\sigma d^{-(m-1)/2}\sqrt{\log d}\right)^{\beta}\}$. In this case, (59) and (60) may yield different ordering of (i, j).

Combining the above two cases gives that

$$\left\{j\colon \frac{1}{d}|j-i| \le \left(4\sigma d^{-(m-1)/2}\sqrt{\log d}\right)^{\beta}\right\} \supset \mathrm{I}\Delta\mathrm{II}.$$

Finally, we have

$$\mathrm{Loss}(\pi, \hat{\pi}) := \frac{1}{d}\max_{i\in[d]}|\pi(i) - \hat{\pi}(i)| \le \frac{1}{d}|\mathrm{I}\Delta\mathrm{II}| \le 2\left(4\sigma d^{-(m-1)/2}\sqrt{\log d}\right)^{\beta},$$

with high probability. □

Proof of Lemma 4. Here we provide the proof of Lemma 4 for the general $\alpha \in (0, 1]$.

1. $\phi_{\boldsymbol{a}} \in \mathcal{F}(\alpha, L)$: We first claim that

$$|K(x_1, \ldots, x_m) - K(y_1, \ldots, y_m)| \le 2\sum_{i=1}^{m}|x_i - y_i|. \tag{62}$$

We prove this by induction. When $m = 2$, we have

$$|K(x_1, x_2) - K(y_1, y_2)| \le 2(|x_1 - y_1| + |x_2 - y_2|).$$

Suppose $m = \ell$, we have

$$|K(x_1, \ldots, x_\ell) - K(y_1, \ldots, y_\ell)| \le 2 \sum_{i=1}^{\ell} |x_i - y_i|. \tag{63}$$

Then we have

$$\begin{aligned}
&|K(x_1, \ldots, x_\ell, x_{\ell+1}) - K(y_1, \ldots, y_\ell, y_{\ell+1})| \\
&= |K(x_1, \ldots, x_\ell)(1 - 2|x_{\ell+1}|)\mathbb{1}\{|x_{\ell+1}| \le 1/2\} - K(y_1, \ldots, y_\ell)(1 - 2|y_{\ell+1}|)\mathbb{1}\{|y_{\ell+1}| \le 1/2\}| \\
&\le 2K(x_1, \ldots, x_\ell)||x_{\ell+1} - y_{\ell+1}| + (1 - 2|y_{\ell+1}|)|K(x_1, \ldots, x_\ell) - K(y_1, \ldots, y_\ell)| \\
&\le 2|x_{\ell+1} - y_{\ell+1}| + 2 \sum_{i=1}^{\ell} |x_i - y_i| \\
&\le 2 \sum_{i=1}^{\ell+1} |x_i - y_i|,
\end{aligned}$$

where the second inequality uses (63). By mathematical induction, (62) holds true for any integer $m > 0$.

For any $\boldsymbol{x} = (x_1, \ldots, x_m), \boldsymbol{y} = (y_1, \ldots, y_m)$ in the support of $\phi_{\boldsymbol{a}}$, we have

$$\begin{aligned}
|\phi_{\boldsymbol{a}}(x_1, \ldots, x_m) - \phi_{\boldsymbol{a}}(y_1, \ldots, y_m)| &\le 2Mk^{1-\alpha} \sum_{i=1}^{m} |x_i - y_i| \\
&\le 2Mmk^{1-\alpha} \|\boldsymbol{x} - \boldsymbol{y}\|_\infty \\
&\le 2Mm \|\boldsymbol{x} - \boldsymbol{y}\|_\infty^\alpha,
\end{aligned}$$

where the last inequality uses the fact that $\|\boldsymbol{x} - \boldsymbol{y}\|_\infty \le 1/k$ and $\alpha \in (0, 1]$. Setting $M = L/2m$, we prove $\phi_{\boldsymbol{a}} \in \mathcal{F}(\alpha, L)$.

2. $\sum_{(i_1, \ldots, i_m) \in [d]^m} \phi_{\boldsymbol{a}}(i_1/d, \ldots, i_m/d) \ge \frac{M^2}{3^m} d^m k^{-2\alpha - m}$: Notice that

$$\begin{aligned}
\sum_{(i_1, \ldots, i_m) \in [d]^m} \phi_{\boldsymbol{a}}^2 \left(\frac{i_1}{d}, \ldots, \frac{i_m}{d} \right) &= M^2 k^{-2\alpha} \sum_{(i_1, \ldots, i_m) \in [d]^m} K^2 \left(\frac{ki_1}{d} - a_1 + \frac{1}{2}, \ldots, \frac{ki_m}{d} - a_m + \frac{1}{2} \right) \\
&= \left(\sum_{\frac{d(a-1)}{k} < i \le \frac{da}{k}} \left(1 - 2 \left| \frac{ki}{d} - a + \frac{1}{2} \right| \right)^2 \right)^m,
\end{aligned}$$

and

$$\sum_{\frac{d(a-1)}{k}<i\le\frac{na}{k}}\left(1-2\left|\frac{ki}{d}-a+\frac{1}{2}\right|\right)^2 = \sum_{\frac{d(a-1)}{k}<i\le\frac{na}{k}}\left(1-\frac{2k}{n}\left|i-\frac{(a-1/2)n}{k}\right|\right)^2$$
$$= 4\sum_{0\le t\frac{d}{2k}}\left(1-\frac{2k}{d}t\right)^2$$
$$\ge 2\int_0^{d/2k}\left(1-\frac{2k}{d}t\right)^2$$
$$= \frac{d}{3k}.$$

Therefore, we complete the proof.

□

Proof of Lemma 5. We follow the proof of Lemma 2. Major difference from the proof of Lemma 2 is that we construct an isotonic block tensor instead of the block tensor in Lemma 2. Therefore, we only provide how to construct isotonic block tensor $\mathcal{S}$ which shares the same property of the block tensor used in Lemma 2.

Let us pick $\omega_1,\ldots,\omega_{k/3}\in\{0,1\}^{k/3\times k/3}$ such that $\rho_H(\omega_p,\omega_q)\ge k^2/36$ for all $p\neq q\in[k/3]$. This selection is possible by lemma 10. For the simplicity, assume that k is a multiple of 3. Fixing such $\omega_1,\ldots,\omega_{k/3}$, we define a symmetric core tensor $\mathcal{S}\in\mathbb{R}^{k\times k\times k}$ for $p<q<r$,

$$\mathcal{S}(p,q,r)=\begin{cases} s_{p,q,r} & \text{if } p\in\{1,\ldots,k/3\}, q\in\{k/3+1,\ldots,2k/3\},\\ & \ r\in\{2k/3+1,\ldots,k\},\\ \lfloor\frac{p-1}{3}\rfloor+\lfloor\frac{q-1}{3}\rfloor+\lfloor\frac{r-1}{3}\rfloor & \text{Otherwise.}\end{cases}$$

Here we define a cubic $\{s_{p,q,r}\}$ as

$$s_{p,q,r}=\lfloor\frac{p-1}{3}\rfloor+\lfloor\frac{q-1}{3}\rfloor+\lfloor\frac{r-1}{3}\rfloor+\delta\sqrt{\frac{c\sigma^2\log k}{d^2}}\left(\frac{p+q+r}{k}-1\right)+\sqrt{\frac{c\sigma^2\log k}{d^2}}[\omega_{r-2k/3}]_{p,q-k/3},$$

where a constant $\delta>0$ is set to make $\mathcal{S}$ isotonic and the choice of constant $c>0$ is deferred to a later part of the proof. For a given $r\in\{2k/3+1,\ldots,k\}$, define

$$\boldsymbol{s}(r):=\operatorname{vec}\left(\mathcal{S}\left(1:\frac{k}{3},\ \left(\frac{k}{3}+1\right):\frac{2k}{3},\ r\right)\right). \tag{64}$$

By the construction, we have, for any $r_1,r_2\in\{2k/3+1,\ldots,k\}$

$$\|\boldsymbol{s}(r_1)-\boldsymbol{s}(r_2)\|_F^2\ge\frac{C_{c,\delta}\sigma^2k^2\log k}{36d^2}, \tag{65}$$

for some constant $C_{c,\delta} > 0$.

Define a subset of permutation set $\Pi(d,k)$ by

$$\mathcal{Z} = \left\{ z \in \Pi(d,k) \colon |z^{-1}(a)| = \frac{d}{k} \text{ for } a \in [k], z^{-1}(a) = \left\{ \frac{(a-1)d}{k} + 1, \ldots, \frac{ad}{k} \right\} \text{ for } a \in [2k/3] \right\}.$$

Each $z \in \mathcal{Z}$ induces a block tensor in $\mathcal{B}(k,0)$. We consider the collection of block tensors induced by $\mathcal{Z}$; i.e.,

$$\mathcal{B}(\mathcal{Z}) = \{\Theta^z \in \mathbb{R}^{d \times d \times d} \colon \Theta^z(i,j,k) = \mathcal{S}(z(i), z(j), z(k)) \text{ for } z \in \mathcal{Z}\}.$$

For the Gaussian model, we have

$$D(\mathbb{P}_\Theta | \mathbb{P}_{\Theta'}) \leq \frac{1}{2\sigma^2} \|\Theta - \Theta'\|_F \leq \frac{1}{2\sigma^2} d^3 \frac{c'\sigma^2 \log k}{d^2} = \frac{c' d \log k}{2}, \tag{66}$$

where the first inequality holds for any $\Theta, \Theta' \in \mathcal{B}(\mathcal{Z})$ by Lemma 12. Based on (65) and (66), we verify that all the bounds for Lemma 11 are the same as those in the proof of Lemma 2. Therefore, using the same argument in the proof of Lemma 2 we obtain

$$\inf_{\hat{\Theta}} \sup_{z \in \Pi(d,k)} \mathbb{P}\left(\frac{1}{d^3} \|\hat{\Theta} - \mathcal{S} \circ z\|_F^2 \geq \frac{C\sigma^2 \log k}{d^2} \right) \geq 0.9,$$

with some constant $C > 0$ for sufficiently small $c > 0$ in (64). □

Proof of Lemma 6. Without loss of generality, assume that π is the identity permutation. Notice that $g(i) - \tau(i)$ is the sample average of d^{m-1} independent mean-zero sub-Gaussian random variables with the variance proxy σ^2. Based on the maxima of sub-Gaussian random variables, we have

$$\max_{i \in [d]} |g(i) - \tau(i)| \lesssim 2\sigma \left(\frac{\log d}{d^{m-1}} \right)^{1/2},$$

with probability $1 - \frac{2}{d^2}$. Denote $\delta = 2\sigma\sqrt{\log d / d^{m-1}}$. The estimated permutation $\hat{\pi}$ is defined for which

$$\tau \circ \hat{\pi}^{-1}(1) \leq \tau \circ \hat{\pi}^{-1}(2) \leq \cdots \leq \tau \circ \hat{\pi}^{-1}(d-1) \leq \tau \circ \hat{\pi}^{-1}(d).$$

By definition of the $\hat{\pi}$, we have for any $i > j$ but $\hat{\pi}(i) = j$,

$$\frac{1}{d^{m-1}} \sum_{(i_2,\ldots,i_m) \in [d]^{m-1}} |\Theta(i, i_2, \ldots i_m) - \Theta(\hat{\pi}(i), i_2, \ldots, i_m)| \lesssim \delta. \tag{67}$$

Similarly for any $i < j$ but $\hat{\pi}(i) = j$, we have

$$\frac{1}{d^{m-1}} \sum_{(i_2,\ldots,i_m) \in [d]^{m-1}} |\Theta(\hat{\pi}(i), i_2, \ldots i_m) - \Theta(i, i_2, \ldots, i_m)| \lesssim \delta. \tag{68}$$

Therefore we obtain

$$\begin{aligned}
&\frac{1}{d^m}\sum_{(i_1,\ldots,i_m)\in[d]^m}|\Theta(\hat\pi(i_1),\hat\pi(i_2),\ldots,\hat\pi(i_m))-\Theta(i_1,i_2\ldots,i_m)|\\
&\leq \frac{1}{d^m}\sum_{(i_1,\ldots,i_m)\in[d]^m}\Big(|\Theta(\hat\pi(i_1),\hat\pi(i_2),\ldots,\hat\pi(i_m))-\Theta(i_1,\hat\pi(i_2),\ldots,\hat\pi(i_m))|\\
&\qquad\qquad+|\Theta(i_1,\hat\pi(i_2),\ldots,\hat\pi(i_m))-\Theta(i_1,i_2,\ldots,\hat\pi(i_m))|\\
&\qquad\qquad+\cdots+|\Theta(i_1,i_2,\ldots,\hat\pi(i_m))-\Theta(i_1,i_2,\ldots,i_m)|\Big)\\
&\leq \frac{m}{d^m}\sum_{i_1\in[d]}\sum_{(i_2,\ldots,i_m)}|\Theta(\hat\pi(i_1),i_2,\ldots,i_m)-\Theta(i_1,i_2,\ldots,i_m)|\\
&\lesssim \delta,
\end{aligned}$$

where the second inequality uses the symmetricity of the tensor and the last inequality uses (67) and (68) in the case of wrong permutations. Since the Frobenius norm is bounded by ℓ_1 norm, we complete the proof. □

Definition 3 (sub-Gaussian random vectors)**.** A random vector $\boldsymbol{e}\in\mathbb{R}^d$ is said to be sub-Gaussian with variance proxy σ^2 if $\mathbb{E}\boldsymbol{e}=0$ and $\langle\frac{\boldsymbol{c}}{\|\boldsymbol{c}\|_2},\boldsymbol{e}\rangle$ is sub-Gaussian with variance proxy σ^2 for any vector $\boldsymbol{c}\in\mathbb{R}^d$. Note that the entries in $\boldsymbol{e}$ need not be independently nor identically distributed.

Lemma 8 (Sub-Gaussian maxima under embedding)**.** *Let $\boldsymbol{A}\in\mathbb{R}^{d_1\times d_2}$ be a deterministic matrix with rank $r\leq\min(d_1,d_2)$. Let $\boldsymbol{y}\in\mathbb{R}^{d_1}$ be a vector consisting of independent sub-Gaussian entries with variance proxy σ^2. Then, there exists a sub-Gaussian random vector $\boldsymbol{x}\in\mathbb{R}^r$ with variance proxy σ^2 such that*

$$\max_{\boldsymbol{p}\in\mathbb{R}^{d_2}}\left\langle\frac{\boldsymbol{A}\boldsymbol{p}}{\|\boldsymbol{A}\boldsymbol{p}\|_2},\boldsymbol{y}\right\rangle=\max_{\boldsymbol{q}\in\mathbb{R}^r}\left\langle\frac{\boldsymbol{q}}{\|\boldsymbol{q}\|_2},\boldsymbol{x}\right\rangle.$$

Proof of Lemma 8. Let $\boldsymbol{u}_i\in\mathbb{R}^{d_1},\boldsymbol{v}_j\in\mathbb{R}^{d_2}$ singular vectors and $\lambda_i\in\mathbb{R}$ be singular values of of $\boldsymbol{A}$ such that $\boldsymbol{A}=\sum_{i=1}^r\lambda_i\boldsymbol{u}_i\boldsymbol{v}_i^T$. Then for any $\boldsymbol{p}\in\mathbb{R}^{d_2}$, we have

$$\boldsymbol{A}\boldsymbol{p}=\sum_{i=1}^r\lambda_i\boldsymbol{u}_i\boldsymbol{v}_i^T\boldsymbol{p}=\sum_{i=1}^r\lambda_i(\boldsymbol{v}_i^T\boldsymbol{p})\boldsymbol{u}_i=\sum_{i=1}^r\alpha_i\boldsymbol{u}_i,$$

where $\boldsymbol{\alpha}(\boldsymbol{p})=(\alpha_1,\ldots,\alpha_r)^T:=\big(\lambda_1(\boldsymbol{v}_1^T\boldsymbol{p}),\ldots,\lambda_r(\boldsymbol{v}_r^T\boldsymbol{p})\big)^T\in\mathbb{R}^r$. Notice that $\boldsymbol{\alpha}(\boldsymbol{p})$ covers $\mathbb{R}^r$ in the sense that $\{\boldsymbol{\alpha}(\boldsymbol{p})\colon\boldsymbol{p}\in\mathbb{R}^{d_2}\}=\mathbb{R}^r$. Therefore, we have

$$\begin{aligned}
\max_{\boldsymbol{p}\in\mathbb{R}^{d_2}}\left\langle\frac{\boldsymbol{A}\boldsymbol{p}}{\|\boldsymbol{A}\boldsymbol{p}\|_2},\boldsymbol{y}\right\rangle&=\max_{\boldsymbol{p}\in\mathbb{R}^{d_2}}\sum_{i=1}^r\frac{\alpha_i}{\|\boldsymbol{\alpha}(\boldsymbol{p})\|_2}\boldsymbol{u}_i^T\boldsymbol{y}\\
&=\max_{\boldsymbol{p}\in\mathbb{R}^{d_2}}\left\langle\frac{\boldsymbol{\alpha}(\boldsymbol{p})}{\|\boldsymbol{\alpha}(\boldsymbol{p})\|_2},\boldsymbol{x}\right\rangle
\end{aligned}$$

$$= \max_{\boldsymbol{q} \in \mathbb{R}^r} \left\langle \frac{\boldsymbol{q}}{\|\boldsymbol{q}\|_2}, \boldsymbol{x} \right\rangle,$$

where we define $\boldsymbol{U} = [\boldsymbol{u}_1, \boldsymbol{u}_2, \ldots, \boldsymbol{u}_r] \in \mathbb{R}^{d_2 \times r}$ and $\boldsymbol{x} = \boldsymbol{U}^T \boldsymbol{y} \in \mathbb{R}^r$. Since $\boldsymbol{U}$ has orthonormal columns, the vector $\boldsymbol{x}$ is a sub-Gaussian random vector with variance proxy σ^2. □

Lemma 9 (Theorem 1.19 in Rigollet and Hütter [27]). *Let $\boldsymbol{e} \in \mathbb{R}^d$ be a sub-Gaussian vector with variance proxy σ^2. Then,*

$$\mathbb{P}\left(\max_{\boldsymbol{c} \in \mathbb{R}^d} \left\langle \frac{\boldsymbol{c}}{\|\boldsymbol{c}\|_2}, \boldsymbol{e} \right\rangle \geq t \right) \leq \exp\left(-\frac{t^2}{8\sigma^2} + d \log 6 \right).$$

Lemma 10 (Varshamov-Gilbert bound). *There exists a sequence of subset $\omega_1, \ldots, \omega_N \in \{0,1\}^d$ such that*

$$\rho_H(\omega_i, \omega_j) := \|\omega_i - \omega_j\|_F^2 \geq \frac{d}{4} \text{ for any } i \neq j \in [N],$$

for some $N \geq \exp(d/8)$.

For any two probability measures $\mathbb{P}$ and $\mathbb{Q}$, define the Kullback-Leibler divergence by

$$D(\mathbb{P}|\mathbb{Q}) = \int \left(\log \frac{d\mathbb{P}}{d\mathbb{Q}} \right) d\mathbb{P}.$$

Lemma 11 (Proposition 4.1 in Gao et al. [16]). *Let (Ξ, ρ) be a metric space and $\{\mathbb{P}_\xi \colon \xi \in \Xi\}$ be a collection of probability measure. For any totally bounded $T \subset \Xi$, define the Kullback-Leibler diameter of T by $d_{KL}(T) = \sup_{\xi, \xi' \in T} D(\mathbb{P}_\xi | \mathbb{P}_{\xi'})$. Then, for any $\epsilon > 0$,*

$$\inf_{\hat{\xi}} \sup_{\xi \in \Xi} \mathbb{P}_\xi \left\{ \rho^2(\hat{\xi}, \xi) \geq \frac{\epsilon^2}{4} \right\} \geq 1 - \frac{d_{KL}(T) + \log 2}{\log \mathcal{M}(\epsilon, T, \rho)},$$

where $\mathcal{M}(\epsilon, T, \rho)$ is the packing number of T by ϵ-radius balls with respect to the metric ρ.

Lemma 12 (Proposition 17 in Gao et al. [15]). *Let $\mathbb{P}_\Theta$ (respectively, $\mathbb{P}_{\Theta'}$) denote the distribution of a Gaussian tensor with mean Θ (respectively, Θ') and entrywise i.i.d. noise from $N(0, \sigma^2)$. Then we have*

$$D(\mathbb{P}_\Theta | \mathbb{P}_{\Theta'}) \leq \frac{1}{2\sigma^2} \sum_{\omega \in [d]^m} (\Theta(\omega) - \Theta'(\omega))^2.$$