Supplements for "Tensor denoising and completion based on ordinal observations"

1 Proofs

1.1 Estimation error for tensor denoising

Proof of Theorem 4.1. We suppress the subscript Ω in the proof, because the tensor denoising assumes complete observation $\Omega = [d_1] \times \cdots \times [d_K]$. It follows from the expression of $\mathcal{L}_{\mathcal{Y}}(\Theta)$ that

$$\frac{\partial \mathcal{L}_{\mathcal{Y}}}{\partial \theta_\omega} = \sum_{\ell \in [L]} \mathbb{1}_{\{y_\omega = \ell\}} \frac{\dot{g}_\ell(\theta_\omega)}{g_\ell(\theta_\omega)},$$
$$\frac{\partial^2 \mathcal{L}_{\mathcal{Y}}}{\partial \theta_\omega^2} = \sum_{\ell \in [L]} \mathbb{1}_{\{y_\omega = \ell\}} \frac{\ddot{g}_\ell(\theta_\omega) g_\ell(\theta_\omega) - \dot{g}_\ell^2(\theta_\omega)}{g_\ell^2(\theta_\omega)} \text{ and } \quad \frac{\partial^2 \mathcal{L}_{\mathcal{Y}}}{\partial \theta_\omega \theta_{\omega'}} = 0 \text{ if } \omega \neq \omega', \tag{1}$$

for all $\omega \in [d_1] \times \cdots \times [d_K]$. Define $d_{\text{total}} = \prod_k d_k$. Let $\nabla_\Theta \mathcal{L}_{\mathcal{Y}} \in \mathbb{R}^{d_1 \times \cdots \times d_K}$ denote the tensor of gradient with respect to $\Theta \in \mathbb{R}^{d_1 \times \cdots \times d_K}$, and $\nabla^2_\Theta \mathcal{L}_{\mathcal{Y}}$ the corresponding Hession matrix of size d_{total}-by-d_{total}. Here, $\text{Vec}(\cdot)$ denotes the operation that turns a tensor into a vector. By (24), $\nabla^2_\Theta \mathcal{L}_{\mathcal{Y}}$ is a diagonal matrix. Recall that

$$U_\alpha = \max_{\ell \in [L], |\theta| \leq \alpha} \frac{|\dot{g}_\ell(\theta)|}{g_\ell(\theta)} > 0 \quad \text{and} \quad L_\alpha = \min_{\ell \in [L], |\theta| \leq \alpha} \frac{\dot{g}_\ell^2(\theta) - \ddot{g}_\ell(\theta) g_\ell(\theta)}{g_\ell^2(\theta)} > 0. \tag{2}$$

Therefore, the entries in $\nabla_\Theta \mathcal{L}_{\mathcal{Y}}$ are upper bounded in magnitude by $U_\alpha > 0$, and all diagonal entries in $\nabla^2_\Theta \mathcal{L}_{\mathcal{Y}}$ are upper bounded by $-L_\alpha < 0$.
By the second-order Taylor's expansion of $\mathcal{L}_{\mathcal{Y}}(\Theta)$ around Θ^{true}, we obtain

$$\mathcal{L}_{\mathcal{Y}}(\Theta) = \mathcal{L}_{\mathcal{Y}}(\Theta^{\text{true}}) + \langle \text{Vec}(\nabla_\Theta \mathcal{L}_{\mathcal{Y}}),\ \text{Vec}(\Theta - \Theta^{\text{true}}) \rangle + \frac{1}{2} \text{Vec}(\Theta - \Theta^{\text{true}})^T \nabla^2_\Theta \mathcal{L}_{\mathcal{Y}}(\check{\Theta}) \text{Vec}(\Theta - \Theta^{\text{true}}), \tag{3}$$

$\check{\Theta} = \gamma \Theta^{\text{true}} + (1 - \gamma)\Theta$ for some $\gamma \in [0, 1]$, and $\nabla^2_\Theta \mathcal{L}_{\mathcal{Y}}(\check{\Theta})$ denotes the $\prod_k d_k$-by-$\prod_k d_k$ Hession matrix evaluated at $\check{\Theta}$.
We first bound the linear term in (3). Note that, by Lemma 4,

$$|\langle \text{Vec}(\nabla_\Theta \mathcal{L}_{\mathcal{Y}}(\Theta^{\text{true}}), \text{Vec}(\Theta - \Theta^{\text{true}}) \rangle| \leq \|\nabla_\Theta \mathcal{L}_{\mathcal{Y}}(\Theta^{\text{true}})\|_\sigma \|\Theta - \Theta^{\text{true}}\|_*, \tag{4}$$

where $\|\cdot\|_\sigma$ denotes the tensor spectral norm and $\|\cdot\|_*$ denotes the tensor nuclear norm. Define

$$s_\omega = \frac{\partial \mathcal{L}_{\mathcal{Y}}}{\partial \theta_\omega}\Big|_{\Theta = \Theta^{\text{true}}} \quad \text{for all } \ \omega \in [d_1] \times \cdots \times [d_K].$$

Based on (24) and the definition of U_α, $\nabla_\Theta \mathcal{L}_{\mathcal{Y}}(\Theta^{\text{true}}) = [\![s_\omega]\!]$ is a random tensor whose entries are independently distributed satisfying

$$\mathbb{E}(s_\omega) = 0, \quad |s_\omega| \leq U_\alpha, \quad \text{for all } \omega \in [d_1] \times \cdots \times [d_K]. \tag{5}$$

By lemma 6, with probability at least $1 - \exp(-C_1 \sum_k d_k)$, we have

$$\|\nabla_\Theta \mathcal{L}_{\mathcal{Y}}(\Theta^{\text{true}})\|_\sigma \leq C_2 U_\alpha \sqrt{\sum_k d_k}, \tag{6}$$

where C_1, C_2 are two positive constants that depend only on K. Furthermore, note that $\text{rank}(\Theta) \leq \boldsymbol{r}$, $\text{rank}(\Theta^{\text{true}}) \leq \boldsymbol{r}$, so $\text{rank}(\Theta - \Theta^{\text{true}}) \leq 2\boldsymbol{r}$. By lemma 3, $\|\Theta - \Theta^{\text{true}}\|_* \leq (2r_{\max})^{\frac{K-1}{2}} \|\Theta - \Theta^{\text{true}}\|_F$. Combining (4), (5) and (6), we have that, with probability at least $1 - \exp(-C_1 \sum_k d_k)$,

$$|\langle \text{Vec}(\nabla_\Theta \mathcal{L}_\mathcal{Y}(\Theta^{\text{true}})), \text{Vec}(\Theta - \Theta^{\text{true}}) \rangle| \leq C_2 U_\alpha \sqrt{r_{\max}^{K-1} \sum_k d_k} \|\Theta - \Theta^{\text{true}}\|_F. \tag{7}$$

We next bound the quadratic term in (3). Note that

$$\begin{aligned} \text{Vec}(\Theta - \Theta^{\text{true}})^T \nabla^2_\Theta \mathcal{L}_\mathcal{Y}(\check{\Theta}) \, \text{Vec}(\Theta - \Theta^{\text{true}}) &= \sum_\omega \left(\frac{\partial^2 \mathcal{L}_\mathcal{Y}}{\partial \theta_\omega^2} \Big|_{\Theta = \check{\Theta}} \right) (\theta_\omega - \theta_{\text{true},\omega})^2 \\ &\leq -L_\alpha \sum_\omega (\Theta_\omega - \Theta_{\text{true},\omega})^2 \\ &= -L_\alpha \|\Theta - \Theta^{\text{true}}\|_F^2, \end{aligned} \tag{8}$$

where the second line comes from the fact that $\|\check{\Theta}\|_\infty \leq \alpha$ and the definition of L_α.
Combining (3), (7) and (8), we have that, for all $\Theta \in \mathcal{P}$, with probability at least $1 - \exp(-C_1 \sum_k d_k)$,

$$\mathcal{L}_\mathcal{Y}(\Theta) \leq \mathcal{L}_\mathcal{Y}(\Theta^{\text{true}}) + C_2 U_\alpha \left(r_{\max}^{K-1} \sum_k d_k \right)^{1/2} \|\Theta - \Theta^{\text{true}}\|_F - \frac{L_\alpha}{2} \|\Theta - \Theta^{\text{true}}\|_F^2.$$

In particular, the above inequality also holds for $\hat{\Theta} \in \mathcal{P}$. Therefore,

$$\mathcal{L}_\mathcal{Y}(\hat{\Theta}) \leq \mathcal{L}_\mathcal{Y}(\Theta^{\text{true}}) + C_2 U_\alpha \left(r_{\max}^{K-1} \sum_k d_k \right)^{1/2} \|\hat{\Theta} - \Theta^{\text{true}}\|_F - \frac{L_\alpha}{2} \|\hat{\Theta} - \Theta^{\text{true}}\|_F^2.$$

Since $\hat{\Theta} = \arg\max_{\Theta \in \mathcal{P}} \mathcal{L}_\mathcal{Y}(\Theta)$, $\mathcal{L}_\mathcal{Y}(\hat{\Theta}) - \mathcal{L}_\mathcal{Y}(\Theta^{\text{true}}) \geq 0$, which gives

$$C_2 U_\alpha \left(r_{\max}^{K-1} \sum_k d_k \right)^{1/2} \|\hat{\Theta} - \Theta^{\text{true}}\|_F - \frac{L_\alpha}{2} \|\hat{\Theta} - \Theta^{\text{true}}\|_F^2 \geq 0.$$

Henceforth,

$$\frac{1}{\sqrt{\prod_k d_k}} \|\hat{\Theta} - \Theta^{\text{true}}\|_F \leq \frac{2 C_2 U_\alpha \sqrt{r_{\max}^{K-1} \sum_k d_k}}{L_\alpha \sqrt{\prod_k d_k}} = \frac{2 C_2 U_\alpha r_{\max}^{(K-1)/2}}{L_\alpha} \sqrt{\frac{\sum_k d_k}{\prod_k d_k}}.$$

This completes the proof. □

Proof of Corollary 1. The result follows immediately from Theorem 4.1 and Lemma 8. □

Proof of Theorem 4.2. Let $d_{\text{total}} = \prod_{k \in [K]} d_k$, and $\gamma \in [0, 1]$ be a constant to be specified later. Our strategy is to construct a finite set of tensors $\mathcal{X} = \{\Theta_i : i = 1, \ldots\} \subset \mathcal{P}$ satisfying the properties of (i)-(iv) in Lemma 9. By Lemma 9, such a subset of tensors exist. For any tensor $\Theta \in \mathcal{X}$, let $\mathbb{P}_\Theta$ denote the distribution of $\mathcal{Y}|\Theta$, where $\mathcal{Y}$ is the ordinal tensor. In particular, $\mathbb{P}_\mathbf{0}$ is the distribution of $\mathcal{Y}$ induced by the zero parameter tensor $\mathbf{0}$, i.e., the distribution of $\mathcal{Y}$ conditional on the parameter tensor $\Theta = \mathbf{0}$. Based on the Remark for Lemma 8, we have

$$KL(\mathbb{P}_\Theta || \mathbb{P}_\mathbf{0}) \leq C \|\Theta\|_F^2, \tag{9}$$

where $C = \frac{(4L-6)\dot{f}^2(0)}{A_\alpha} > 0$ is a constant independent of the tensor dimension and rank. Combining the inequality (9) with property (iii) of $\mathcal{X}$, we have

$$\mathrm{KL}(\mathbb{P}_\Theta || \mathbb{P}_{\mathbf{0}}) \le \gamma^2 R_{\max} d_{\max}. \tag{10}$$

From (10) and the property (i), we deduce that the condition

$$\frac{1}{\mathrm{Card}(\mathcal{X}) - 1} \sum_{\Theta \in \mathcal{X}} \mathrm{KL}(\mathbb{P}_\Theta, \mathbb{P}_{\mathbf{0}}) \le \varepsilon \log_2 \{\mathrm{Card}(\mathcal{X}) - 1\} \tag{11}$$

holds for any $\varepsilon \ge 0$ when $\gamma \in [0,1]$ is chosen to be sufficiently small depending on ε, e.g., $\gamma \le \sqrt{3\varepsilon}$. By applying Lemma 11 to (11), and in view of the property (iv), we obtain that

$$\inf_{\hat{\Theta}} \sup_{\Theta^{\mathrm{true}} \in \mathcal{X}} \mathbb{P}\left(\|\hat{\Theta} - \Theta^{\mathrm{true}}\|_F \ge \frac{\gamma}{8} \min\left\{ \alpha \sqrt{d_{\mathrm{total}}}, C^{-1/2} \sqrt{R_{\max} d_{\max}} \right\} \right) \ge \frac{1}{2} \left(1 - 2\varepsilon - \sqrt{\frac{16\varepsilon}{R_{\max} d_{\max}}} \right). \tag{12}$$

Note that $\mathrm{Loss}(\hat{\Theta}, \Theta^{\mathrm{true}}) = \|\hat{\Theta} - \Theta^{\mathrm{true}}\|_F^2 / d_{\mathrm{total}}$ and $\mathcal{X} \subset \mathcal{P}$. By taking $\varepsilon = 1/8$ and $\gamma = 1/2$, we conclude from (12) that

$$\inf_{\hat{\Theta}} \sup_{\Theta^{\mathrm{true}} \in \mathcal{P}} \mathbb{P}\left(\mathrm{Loss}(\hat{\Theta}, \Theta^{\mathrm{true}}) \ge \frac{1}{256} \min\left\{ \alpha^2, \frac{C^{-1} R_{\max} d_{\max}}{d_{\mathrm{total}}} \right\} \right) \ge \frac{1}{2} \left(\frac{3}{4} - \frac{2}{R_{\max} d_{\max}} \right) \ge \frac{1}{8}.$$

This completes the proof. □

1.2 Sample complexity for tensor completion

Proof of Theorem 4.3. For notational convenience, we use $\|\Theta\|_{F,\Omega} = \sum_{\omega \in \Omega} \Theta_\omega^2$ to denote the sum of squared entries over the observed set Ω, for a tensor $\Theta \in \mathbb{R}^{d_1 \times \cdots \times d_K}$.
Following a similar argument as in the proof of Theorem 4.1, we have

$$\mathcal{L}_{\mathcal{Y},\Omega}(\Theta) = \mathcal{L}_{\mathcal{Y},\Omega}(\Theta^{\mathrm{true}}) + \langle \mathrm{Vec}(\nabla_\Theta \mathcal{L}_{\mathcal{Y},\Omega}),\ \mathrm{Vec}(\Theta - \Theta^{\mathrm{true}}) \rangle + \frac{1}{2} \mathrm{Vec}(\Theta - \Theta^{\mathrm{true}})^T \nabla_\Theta^2 \mathcal{L}_{\mathcal{Y},\Omega}(\check{\Theta}) \mathrm{Vec}(\Theta - \Theta^{\mathrm{true}}), \tag{13}$$

where

1. $\nabla_\Theta \mathcal{L}_{\mathcal{Y},\Omega}$ is a $d_1 \times \cdots \times d_K$ tensor with $|\Omega|$ nonzero entries, and each entry is upper bounded by $U_\alpha > 0$.
2. $\nabla_\Theta^2 \mathcal{L}_{\mathcal{Y},\Omega}$ is a diagonal matrix of size d_{total}-by-d_{total} with $|\Omega|$ nonzero entries, and each entry is upper bounded by $-L_\alpha < 0$.

Similar to (4) and (8), we have

$$|\langle \mathrm{Vec}(\nabla_\Theta \mathcal{L}_{\mathcal{Y},\Omega}),\ \mathrm{Vec}(\Theta - \Theta^{\mathrm{true}}) \rangle| \le C_2 U_\alpha \sqrt{r_{\max}^{K-1} \sum_k d_k} \|\Theta - \Theta^{\mathrm{true}}\|_{F,\Omega}$$

and

$$\mathrm{Vec}(\Theta - \Theta^{\mathrm{true}})^T \nabla_\Theta^2 \mathcal{L}_{\mathcal{Y}}(\check{\Theta}) \mathrm{Vec}(\Theta - \Theta^{\mathrm{true}}) \le -L_\alpha \|\Theta - \Theta^{\mathrm{true}}\|_{F,\Omega}^2. \tag{14}$$

Combining (13)-(14) with the fact that $\mathcal{L}_{\mathcal{Y},\Omega}(\hat{\Theta}) \ge \mathcal{L}_{\mathcal{Y},\Omega}(\Theta^{\mathrm{true}})$, we have

$$\|\hat{\Theta} - \Theta^{\mathrm{true}}\|_{F,\Omega} \le \frac{2 C_2 U_\alpha r_{\max}^{(K-1)/2}}{L_\alpha} \sqrt{\sum_k d_k}. \tag{15}$$

Lastly, we invoke the result regarding the closeness of Θ to its sampled version Θ_Ω, under the entrywise bound condition. Note that $\|\hat{\Theta}-\Theta^{\text{true}}\|_\infty \le 2\alpha$ and $\text{rank}(\hat{\Theta}-\Theta^{\text{true}}) \le 2\boldsymbol{r}$. By Lemma 2, $\|\hat{\Theta}-\Theta^{\text{true}}\|_M \le 2^{(3K-1)/2}\alpha\left(\frac{\prod r_k}{r_{\max}}\right)^{3/2}$. Therefore, the condition in Lemma 12 holds with $\beta = 2^{(3K-1)/2}\alpha\left(\frac{\prod r_k}{r_{\max}}\right)^{3/2}$. Applying Lemma 12 to (15) gives

$$\begin{aligned}\|\hat{\Theta}-\Theta^{\text{true}}\|_{F,\Pi}^2 &\le \frac{1}{m}\|\hat{\Theta}-\Theta^{\text{true}}\|_{F,\Omega}^2 + c\beta\sqrt{\frac{\sum_k d_k}{|\Omega|}} \\ &\le C_2 r_{\max}^{K-1}\frac{\sum_k d_k}{|\Omega|} + C_1\alpha r_{\max}^{3(K-1)/2}\sqrt{\frac{\sum_k d_k}{|\Omega|}},\end{aligned}$$

with probability at least $1-\exp(-\frac{\sum_k d_k}{\sum_k \log d_k})$ over the sampled set Ω. Here $C_1, C_2 > 0$ are two constants independent of the tensor dimension and rank. Therefore,

$$\|\hat{\Theta}-\Theta^{\text{true}}\|_{F,\Pi}^2 \to 0, \quad \text{as} \quad \frac{|\Omega|}{\sum_k d_k} \to \infty,$$

provided that $r_{\max} = O(1)$. □

1.3 Convexity of the log-likelihood function

Theorem 1.1. *Define the function*

$$\mathcal{L}_{\mathcal{Y},\Omega}(\Theta, \boldsymbol{b}) = \sum_{\omega\in\Omega}\sum_{\ell\in[L]}\left\{\mathbb{1}_{\{y_\omega=\ell\}}\log\left[f(\theta_\omega+b_\ell)-f(\theta_\omega+b_{\ell-1})\right]\right\}, \tag{16}$$

where $f(\cdot)$ satisfies Assumption 1. Then, $\mathcal{L}_{\mathcal{Y},\Omega}(\Theta, \boldsymbol{b})$ is concave in $(\Theta, \boldsymbol{b})$.

Proof. Define $d_{\text{total}} = \prod_k d_k$. With a little abuse of notation, we use $(\Theta, \boldsymbol{b})$ to denote the length-$(d_{\text{total}}+L-1)$-vector collecting all parameters together. Let us denote a bivariate function

$$\begin{aligned}\lambda :&\mathbb{R}^2 \mapsto \mathbb{R} \\ &(u,v) \mapsto \lambda(u,v) = \log\left[f(u)-f(v)\right].\end{aligned}$$

It suffices to show that $\lambda(u,v)$ is concave in (u,v) where $u > v$.
Suppose that the claim holds (which we will prove in the next paragraph). Based on (16), u, v are both linear functions of $(\Theta, \boldsymbol{b})$:

$$u = \boldsymbol{a}_1^T(\Theta, \boldsymbol{b}), \quad v = \boldsymbol{a}_2^T(\Theta, \boldsymbol{b}), \quad \text{for some } \boldsymbol{a}_1, \boldsymbol{a}_2 \in \mathbb{R}^{d_{\text{total}}+L-1}.$$

Then, $\lambda(u,v) = \lambda(\boldsymbol{a}_1^T(\Theta, \boldsymbol{b}),\ \boldsymbol{a}_2^T(\Theta, \boldsymbol{b}))$ is concave in $(\Theta, \boldsymbol{b})$ by the definition of concavity. Therefore, we can conclude that $\mathcal{L}_{\mathcal{Y},\Omega}(\Theta, \boldsymbol{b})$ is concave in $(\Theta, \boldsymbol{b})$ because $\mathcal{L}_{\mathcal{Y},\Omega}(\Theta, \boldsymbol{b})$ is the sum of $\lambda(u,v)$.
Now, we prove the concavity of $\lambda(u,v)$. Note that

$$\lambda(u,v) = \log\left[f(u)-f(v)\right] = \log\Big[\int \mathbb{1}_{[u,v]}(x)f'(x)dx\Big],$$

where $\mathbb{1}_{[u,v]}$ is an indicator function that equals 1 in the interval $[u,v]$, and 0 elsewhere. Furthermore, $\mathbb{1}_{[u,v]}(x)$ is log-concave in (u,v,x), and by Assumption 1, $f'(x)$ is log-concave in x. It follows that $\mathbb{1}_{(u,v)}(x)f'(x)$ is a log-concave in (u,v,x). By Lemma 1, we conclude that $\lambda(u,v)$ is concave in (u,v) where $u > v$. □

Lemma 1 (Corollary 3.5 in Brascamp and Lieb [2002]). *Let $F(x,y):\mathbb{R}^{m+n}\to\mathbb{R}$ be an integrable function where $x\in\mathbb{R}^m, y\in\mathbb{R}^n$. Let*

$$G(x)=\int_{\mathbb{R}^n}F(x,y)dy.$$

If $F(x,y)$ is log concave in (x,y), then $G(x)$ is log concave in x.

1.4 Auxiliary lemmas

This section collects lemmas that are useful for the proofs of the main theorems.

Definition 1 (Atomic M-norm [Ghadermarzy et al., 2019]). Define $T_\pm=\{\mathcal{T}\in\{\pm1\}^{d_1\times\cdots\times d_K}:\operatorname{rank}(\mathcal{T})=1\}$. The atomic M-norm of a tensor $\Theta\in\mathbb{R}^{d_1\times\cdots\times d_K}$ is defined as

$$\begin{aligned}\|\Theta\|_M&=\inf\{t>0:\Theta\in t\operatorname{conv}(T_\pm)\}\\&=\inf\left\{\sum_{\mathcal{X}\in T_\pm}c_x:\Theta=\sum_{\mathcal{X}\in T_\pm}c_x\mathcal{X},\ c_x>0\right\}.\end{aligned}$$

Definition 2 (Spectral norm [Lim, 2005]). The spectral norm of a tensor $\Theta\in\mathbb{R}^{d_1\times\cdots\times d_K}$ is defined as

$$\|\Theta\|_\sigma=\sup\left\{\langle\Theta,\boldsymbol{x}_1\otimes\cdots\otimes\boldsymbol{x}_K\rangle:\|\boldsymbol{x}_k\|_2=1,\ \boldsymbol{x}_k\in\mathbb{R}^{d_k},\ \text{for all } k\in[K]\right\}.$$

Definition 3 (Nuclear norm [Friedland and Lim, 2018]). The nuclear norm of a tensor $\Theta\in\mathbb{R}^{d_1\times\cdots\times d_K}$ is defined as

$$\|\Theta\|_*=\inf\left\{\sum_{i\in[r]}|\lambda_i|:\Theta=\sum_{i=1}^r\lambda_i\boldsymbol{x}_1^{(i)}\otimes\cdots\otimes\boldsymbol{x}_K^{(i)},\ \|\boldsymbol{x}_k^{(i)}\|_2=1,\ \boldsymbol{x}_k^{(i)}\in\mathbb{R}^{d_k},\ \text{for all}k\in[K],\ i\in[r]\right\},$$

where the infimum is taken over all $r\in\mathbb{N}$ and $\left\|\boldsymbol{x}_k^{(i)}\right\|_2=1$ for all $i\in[r]$ and $k\in[K]$.

Lemma 2 (M-norm and infinity norm [Ghadermarzy et al., 2019]). *Let $\Theta\in\mathbb{R}^{d_1\times\cdots\times d_K}$ be an order-K, rank-$(r_1,\ldots,r_K)$ tensor. Then*

$$\|\Theta\|_\infty\le\|\Theta\|_M\le\left(\frac{\prod_k r_k}{r_{\max}}\right)^{\frac{3}{2}}\|\Theta\|_\infty.$$

Lemma 3 (Nuclear norm and F-norm). *Let $\mathcal{A}\in\mathbb{R}^{d_1\times\cdots\times d_K}$ be an order-K tensor with Tucker rank$(\mathcal{A})=(r_1,\ldots,r_K)$. Then*

$$\|\mathcal{A}\|_*\le\sqrt{\frac{\prod_k r_k}{\max_k r_k}}\|\mathcal{A}\|_F,$$

where $\|\cdot\|_*$ denotes the nuclear norm of the tensor.

Proof. Without loss of generality, suppose $r_1=\min_k r_k$. Let $\mathcal{A}_{(k)}$ denote the mode-k matricization of $\mathcal{A}$ for all $k\in[K]$. By Wang et al. [2017, Corollary 4.11], and the invariance relationship between a tensor and its Tucker core [Jiang et al., 2017, Section 6], we have

$$\|\mathcal{A}\|_*\le\sqrt{\frac{\prod_{k\ge2}r_k}{\max_{k\ge2}r_k}}\|\mathcal{A}_{(1)}\|_*,\tag{17}$$

where $\mathcal{A}_{(1)}$ is a d_1-by-$\prod_{k\geq 2} d_k$ matrix with matrix rank r_1. Furthermore, the relationship between the matrix norms implies that $\|\mathcal{A}_{(1)}\|_* \leq \sqrt{r_1}\|\mathcal{A}_{(1)}\|_F = \sqrt{r_1}\|\mathcal{A}\|_F$. Combining this fact with the inequality (17) yields the final claim. □

Lemma 4. *Let $\mathcal{A}$, $\mathcal{B}$ be two order-K tensors of the same dimension. Then*

$$|\langle \mathcal{A}, \mathcal{B}\rangle| \leq \|\mathcal{A}\|_\sigma \|\mathcal{B}\|_*.$$

Proof. By Friedland and Lim [2018, Proposition 3.1], there exists a nuclear norm decomposition of $\mathcal{B}$, such that

$$\mathcal{B} = \sum_r \lambda_r \boldsymbol{a}_r^{(1)} \otimes \cdots \otimes \boldsymbol{a}_r^{(K)}, \quad \boldsymbol{a}_r^{(k)} \in \mathbf{S}^{d_k - 1}(\mathbb{R}), \quad \text{for all } k \in [K],$$

and $\|\mathcal{B}\|_* = \sum_r |\lambda_r|$. Henceforth we have

$$\begin{aligned} |\langle \mathcal{A}, \mathcal{B}\rangle| &= |\langle \mathcal{A}, \sum_r \lambda_r \boldsymbol{a}_r^{(1)} \otimes \cdots \otimes \boldsymbol{a}_r^{(K)}\rangle| \leq \sum_r |\lambda_r| |\langle \mathcal{A}, \boldsymbol{a}_r^{(1)} \otimes \cdots \otimes \boldsymbol{a}_r^{(K)}\rangle| \\ &\leq \sum_r |\lambda_r| \|\mathcal{A}\|_\sigma = \|\mathcal{A}\|_\sigma \|\mathcal{B}\|_*, \end{aligned}$$

which completes the proof. □

The following lemma provides the bound on the spectral norm of random tensors. The result was firstly presented in Nguyen et al. [2015], and we adopt the version from Tomioka and Suzuki [2014].

Lemma 5 (Tomioka and Suzuki [2014])**.** *Suppose that $\mathcal{S} = [\![s_\omega]\!] \in \mathbb{R}^{d_1 \times \cdots \times d_K}$ is an order-K tensor whose entries are independent random variables that satisfy*

$$\mathbb{E}(s_\omega) = 0, \quad \text{and} \quad \mathbb{E}(e^{t s_\omega}) \leq e^{t^2 L^2/2}.$$

Then the spectral norm $\|\mathcal{S}\|_\sigma$ satisfies that,

$$\|\mathcal{S}\|_\sigma \leq \sqrt{8L^2 \log(12K) \sum_k d_k + \log(2/\delta)},$$

with probability at least $1 - \delta$.

Lemma 6. *Suppose that $\mathcal{S} = [\![s_\omega]\!] \in \mathbb{R}^{d_1 \times \cdots \times d_K}$ is an order-K tensor whose entries are independent random variables that satisfy*

$$\mathbb{E}(s_\omega) = 0, \quad \text{and} \quad |s_\omega| \leq U.$$

Then we have

$$\mathbb{P}\left(\|\mathcal{S}\|_\sigma \geq C_2 U \sqrt{\sum_k d_k}\right) \leq \exp\left(-C_1 \log K \sum_k d_k\right)$$

where $C_1 > 0$ is an absolute constant, and $C_2 > 0$ is a constant that depends only on K.

Proof. Note that the random variable $U^{-1} s_\omega$ is zero-mean and supported on $[-1, 1]$. Therefore, $U^{-1} s_\omega$ is sub-Gaussian with parameter $\frac{1-(-1)}{2} = 1$, i.e.

$$\mathbb{E}(U^{-1} s_\omega) = 0, \quad \text{and} \quad \mathbb{E}(e^{t U^{-1} s_\omega}) \leq e^{t^2/2}.$$

It follows from Lemma 5 that, with probability at least $1-\delta$,

$$\|U^{-1}\mathcal{S}\|_\sigma \leq \sqrt{(c_0 \log K + c_1)\sum_k d_k + \log(2/\delta)},$$

where $c_0, c_1 > 0$ are two absolute constants. Taking $\delta = \exp(-C_1 \log K \sum_k d_k)$ yields the final claim, where $C_2 = c_0 \log K + c_1 + 1 > 0$ is another constant. □

Lemma 7. *Let X, Y be two discrete random variables taking values on L possible categories, with category probabilities $\{p_\ell\}_{\ell\in[L]}$ and $\{q_\ell\}_{\ell\in[L]}$, respectively. Suppose p_ℓ, $q_\ell > 0$ for all $i \in [L]$. Then, the Kullback-Leibler (KL) divergence satisfies that*

$$KL(X||Y) \stackrel{def}{=} -\sum_{\ell\in[L]} \mathbb{P}_X(\ell) \log\left\{\frac{\mathbb{P}_Y(\ell)}{\mathbb{P}_X(\ell)}\right\} \leq \sum_{\ell\in[L]} \frac{(p_\ell - q_\ell)^2}{q_\ell}.$$

Proof. Using the fact $\log x \leq x - 1$ for $x > 0$, we have that

$$\begin{aligned}
\mathrm{KL}(X||Y) &= \sum_{\ell\in[L]} p_\ell \log \frac{p_\ell}{q_\ell} \\
&\leq \sum_{\ell\in[L]} \frac{p_\ell}{q_\ell}(p_\ell - q_\ell) \\
&= \sum_{\ell\in[L]} \left(\frac{p_\ell}{q_\ell} - 1\right)(p_\ell - q_\ell) + \sum_{\ell\in[L]} (p_\ell - q_\ell).
\end{aligned}$$

Note that $\sum_{\ell\in[L]}(p_\ell - q_\ell) = 0$. Therefore,

$$\mathrm{KL}(X||Y) \leq \sum_{\ell\in[L]} \left(\frac{p_\ell}{q_\ell} - 1\right)(p_\ell - q_\ell) = \sum_{\ell\in[L]} \frac{(p_\ell - q_\ell)^2}{q_\ell}.$$

□

Lemma 8 (KL divergence and F-norm)**.** *Let $\mathcal{Y} \in [L]^{d_1\times\cdots\times d_K}$ be an ordinal tensor generated from the model (1) with the link function f and parameter tensor Θ. Let $\mathbb{P}_\Theta$ denote the joint categorical distribution of $\mathcal{Y}|\Theta$ induced by the parameter tensor Θ, where $\|\Theta\|_\infty \leq \alpha$. Define*

$$A_\alpha = \min_{\ell\in[L],|\theta|\leq\alpha} \left[f(\theta + b_\ell) - f(\theta + b_{\ell-1})\right]. \tag{18}$$

Then, for any two tensors Θ, Θ^* in the parameter spaces, we have

$$KL(\mathbb{P}_\Theta||\mathbb{P}_{\Theta^*}) \leq \frac{2(2L-3)}{A_\alpha}\dot{f}^2(0)\|\Theta - \Theta^*\|_F^2.$$

Proof. Suppose that the distribution over the ordinal tensor $\mathcal{Y} = [\![y_\omega]\!]$ is induced by $\Theta = [\![\theta_\omega]\!]$. Then, based on the generative model (1),

$$\mathbb{P}(y_\omega = \ell|\theta_\omega) = f(\theta_\omega + b_\ell) - f(\theta_\omega + b_{\ell-1}),$$

for all $\ell \in [L]$ and $\omega \in [d_1] \times \cdots \times [d_K]$. For notational convenience, we suppress the subscribe in θ_ω and simply write θ (and respectively, θ^*). Based on Lemma 7 and Taylor expansion,

$$\mathrm{KL}(\theta||\theta^*) \leq \sum_{\ell\in[L]} \frac{[f(\theta + b_\ell) - f(\theta + b_{\ell-1}) - f(\theta^* + b_\ell) + f(\theta^* + b_{\ell-1})]^2}{f(\theta^* + b_\ell) - f(\theta^* + b_{\ell-1})}$$

$$\leq \sum_{\ell=2}^{L-1} \frac{\left[\dot{f}(\eta_\ell + b_\ell) - \dot{f}(\eta_{\ell-1} + b_{\ell-1})\right]^2}{f(\theta^* + b_\ell) - f(\theta^* + b_{\ell-1})}(\theta - \theta^*)^2 + \frac{\dot{f}^2(\eta_1 + b_1)}{f(\theta^* + b_1)}(\theta - \theta^*)^2$$
$$+ \frac{\dot{f}^2(\eta_{L-1} + b_{L-1})}{1 - f(\theta^* + b_{L-1})}(\theta - \theta^*)^2,$$

where η_ℓ and $\eta_{\ell-1}$ fall between θ and θ^*. Therefore,

$$\mathrm{KL}(\theta||\theta^*) \leq \left(\frac{4(L-2)}{A_\alpha} + \frac{2}{A_\alpha}\right)\dot{f}^2(0)(\theta - \theta^*)^2 = \frac{2(2L-3)}{A_\alpha}\dot{f}^2(0)(\theta - \theta^*)^2, \tag{19}$$

where we have used Taylor expansion, the bound (18), and the fact that $\dot{f}(\cdot)$ peaks at zero for an unimodal and symmetric function. Now summing (19) over the index set $\omega \in [d_1] \times \cdots \times [d_K]$ gives

$$\mathrm{KL}(\mathbb{P}_\Theta||\mathbb{P}_{\Theta^*}) = \sum_{\omega \in [d_1]\times\cdots\times[d_K]} \mathrm{KL}(\theta_\omega||\theta^*_\omega) \leq \frac{2(2L-3)}{A_\alpha}\dot{f}^2(0)\|\Theta - \Theta^*\|_F^2.$$

□

Remark 1. In particular, let $\mathbb{P}_{\mathbf{0}}$ denote the distribution of $\mathcal{Y}|\mathbf{0}$ induced by the zero parameter tensor. Then we have

$$\mathrm{KL}(\mathbb{P}_\Theta||\mathbb{P}_{\mathbf{0}}) \leq \frac{2(2L-3)}{A_\alpha}\dot{f}^2(0)\|\Theta\|_F^2.$$

Lemma 9. *Assume the same setup as in Theorem 4.2. Without loss of generality, suppose* $d_1 = \max_k d_k$. *Define* $R = \max_k r_k$ *and* $d_{total} = \prod_{k\in[K]} d_k$. *For any constant* $0 \leq \gamma \leq 1$, *there exist a finite set of tensors* $\mathcal{X} = \{\Theta_i : i = 1, \ldots\} \subset \mathcal{P}$ *satisfying the following four properties:*

(i) $Card(\mathcal{X}) \geq 2^{Rd_1/8} + 1$, *where Card denotes the cardinality;*

(ii) $\mathcal{X}$ *contains the zero tensor* $\mathbf{0} \in \mathbb{R}^{d_1\times\cdots\times d_K}$;

(iii) $\|\Theta\|_\infty \leq \gamma \min\left\{\alpha,\ C^{-1/2}\sqrt{\frac{Rd_1}{d_{total}}}\right\}$ *for any element* $\Theta \in \mathcal{X}$;

(iv) $\|\Theta_i - \Theta_j\|_F \geq \frac{\gamma}{4}\min\left\{\alpha\sqrt{d_{total}},\ C^{-1/2}\sqrt{Rd_1}\right\}$ *for any two distinct elements* $\Theta_i,\ \Theta_j \in \mathcal{X}$,

Here $C = C(\alpha, L, f, \boldsymbol{b}) = \frac{(4L-6)\dot{f}^2(0)}{A_\alpha} > 0$ *is a constant independent of the tensor dimension and rank.*

Proof. Given a constant $0 \leq \gamma \leq 1$, we define a set of matrices:

$$\mathcal{C} = \left\{\boldsymbol{M} = (m_{ij}) \in \mathbb{R}^{d_1\times R} : a_{ij} \in \left\{0, \gamma\min\left\{\alpha, C^{-1/2}\sqrt{\frac{Rd_1}{d_{\mathrm{total}}}}\right\}\right\},\ \forall (i,j) \in [d_1]\times[R]\right\}.$$

We then consider the associated set of block tensors:

$$\mathcal{B} = \mathcal{B}(\mathcal{C}) = \{\Theta \in \mathbb{R}^{d_1\times\cdots\times d_K} : \Theta = \boldsymbol{A}\otimes\mathbf{1}_{d_3}\otimes\cdots\otimes\mathbf{1}_{d_K},$$
$$\text{where } \boldsymbol{A} = (\boldsymbol{M}|\cdots|\boldsymbol{M}|\boldsymbol{O}) \in \mathbb{R}^{d_1\times d_2},\ \boldsymbol{M} \in \mathcal{C}\},$$

where $\mathbf{1}_d$ denotes a length-d vector with all entries 1, $\boldsymbol{O}$ denotes the $d_1 \times (d_2 - R\lfloor d_2/R\rfloor)$ zero matrix, and $\lfloor d_2/R\rfloor$ is the integer part of d_2/R. In other words, the subtensor $\Theta(\boldsymbol{I}, \boldsymbol{I}, i_3, \ldots, i_K) \in$

$\mathbb{R}^{d_1 \times d_2}$ are the same for all fixed $(i_3, \ldots, i_K) \in [d_3] \times \cdots \times [d_K]$, and furthermore, each subtensor $\Theta(\boldsymbol{I}, \boldsymbol{I}, i_3, \ldots, i_K)$ itself is filled by copying the matrix $\boldsymbol{M} \in \mathbb{R}^{d_1 \times R}$ as many times as would fit. By construction, any element of $\mathcal{B}$, as well as the difference of any two elements of $\mathcal{B}$, has Tucker rank at most $\max_k r_k \leq R$, and the entries of any tensor in $\mathcal{B}$ take values in $[0, \alpha]$. Thus, $\mathcal{B} \subset \mathcal{P}$. By Lemma 10, there exists a subset $\mathcal{X} \subset \mathcal{B}$ with cardinality $\mathrm{Card}(\mathcal{X}) \geq 2^{Rd_1/8} + 1$ containing the zero $d_1 \times \cdots \times d_K$ tensor, such that, for any two distinct elements Θ_i and Θ_j in $\mathcal{X}$,

$$\|\Theta_i - \Theta_j\|_F^2 \geq \frac{Rd_1}{8}\gamma^2 \min\left\{\alpha^2, \frac{C^{-1}Rd_1}{d_{\text{total}}}\right\} \lfloor \frac{d_2}{R} \rfloor \prod_{k \geq 3} d_k \geq \frac{\gamma^2 \min\left\{\alpha^2 d_{\text{total}}, C^{-1}Rd_1\right\}}{16}.$$

In addition, each entry of $\Theta \in \mathcal{X}$ is bounded by $\gamma \min\left\{\alpha, C^{-1/2}\sqrt{\frac{Rd_1}{d_{\text{total}}}}\right\}$. Therefore the Properties (i) to (iv) are satisfied. □

Lemma 10 (Varshamov-Gilbert bound). *Let $\Omega = \{(w_1, \ldots, w_m) \colon w_i \in \{0, 1\}\}$. Suppose $m > 8$. Then there exists a subset $\{w^{(0)}, \ldots, w^{(M)}\}$ of Ω such that $w^{(0)} = (0, \ldots, 0)$ and*

$$\|w^{(j)} - w^{(k)}\|_0 \geq \frac{m}{8}, \quad \text{for } 0 \leq j < k \leq M,$$

where $\|\cdot\|_0$ denotes the Hamming distance, and $M \geq 2^{m/8}$.

Lemma 11 (Theorem 2.5 in Tsybakov [2009]). *Assume that a set $\mathcal{X}$ contains element $\Theta_0, \Theta_1, \ldots, \Theta_M$ ($M \geq 2$) such that*

- *$d(\Theta_j,\ \Theta_k) \geq 2s > 0$, $\forall 0 \leq j \leq k \leq M$;*
- *$\mathbb{P}_j \ll \mathbb{P}_0$, $\forall j = 1, \ldots, M$, and*

$$\frac{1}{M}\sum_{j=1}^{M} KL(\mathbb{P}_j || \mathbb{P}_0) \leq \alpha \log M$$

where $d \colon \mathcal{X} \times \mathcal{X} \mapsto [0, +\infty]$ is a semi-distance function, $0 < \alpha < 1/8$ and $P_j = P_{\Theta_j}$, $j = 0, 1 \ldots, M$.

Then

$$\inf_{\hat{\Theta}} \sup_{\Theta \in \mathcal{X}} \mathbb{P}_\Theta(d(\hat{\Theta}, \Theta) \geq s) \geq \frac{\sqrt{M}}{1 + \sqrt{M}}\left(1 - 2\alpha - \sqrt{\frac{2\alpha}{\log M}}\right) > 0.$$

Lemma 12 (Lemma 28 in Ghadermarzy et al. [2019]). *Define $\mathbb{B}_M(\beta) = \{\Theta \in \mathbb{R}^{d_1 \times \cdots \times d_K} \colon \|\Theta\|_M \leq \beta\}$. Let $\Omega \subset [d_1] \times \cdots \times [d_K]$ be a random set with $m = |\Omega|$, and assume that each entry in Ω is drawn with replacement from $[d_1] \times \cdots \times [d_K]$ using probability Π. Define*

$$\|\Theta\|_{F,\Pi}^2 = \frac{1}{m}\mathbb{E}_{\Omega \in \Pi}\|\Theta\|_{F,\Omega}^2.$$

Then, there exists a universal constant $c > 0$, such that, with probability at least $1 - \exp\left(-\frac{\sum_k d_k}{\sum_k \log d_k}\right)$ over the sampled set Ω,

$$\frac{1}{m}\|\Theta\|_{F,\Omega}^2 \geq \|\Theta\|_{F,\Pi}^2 - c\beta\sqrt{\frac{\sum_k d_k}{m}}$$

holds uniformly for all $\Theta \in \mathbb{B}_M(\beta)$.

2 Extension of Theorem 4.1 to unknown cut-off points

When the cut-off points $\boldsymbol{b}$ is unknown, we estimate $(\hat{\Theta}, \hat{\boldsymbol{b}})$ by

$$(\hat{\Theta}, \hat{\boldsymbol{b}}) = \underset{(\Theta, \boldsymbol{b}) \in \mathcal{P} \times \mathcal{B}}{\arg\max}\, \mathcal{L}_{\mathcal{Y},\Omega}(\Theta, \boldsymbol{b}), \tag{20}$$

where

$$\mathcal{P} = \{\Theta \in \mathbb{R}^{d_1 \times \cdots \times d_K} : \operatorname{rank}(\mathcal{P}) \le \boldsymbol{r},\ \|\Theta\|_\infty \le \alpha\}, \quad \mathcal{B} = \{\boldsymbol{b} \in \mathbb{R}^{L-1} : \|\boldsymbol{b}\|_\infty \le \beta,\ \min_\ell (b_\ell - b_{\ell-1}) \ge \Delta\}.$$

The estimation accuracy is assessed using the mean squared error (MSE):

$$\mathrm{MSE}(\hat{\Theta}, \Theta) = \frac{1}{\prod_k d_k} \|\hat{\Theta} - \Theta\|_F, \quad \mathrm{MSE}(\hat{\boldsymbol{b}}, \boldsymbol{b}) = \frac{1}{L-1} \|\hat{\boldsymbol{b}} - \boldsymbol{b}\|_F.$$

We introduce several quantities that will used in our theory:

1. We make the convention that $b_0 = -\infty$, $b_L = \infty$, $f(-\infty) = 0$, $f(\infty) = 1$, and $\dot{f}(-\infty) = \ddot{f}(-\infty) = \dot{f}(\infty) = \ddot{f}(\infty) = 0$.
2. The difference function $g_\ell(\theta)$ is defined as $g_\ell(\theta) = f(\theta + b_\ell) - f(\theta + b_{\ell-1})$ for all $\theta \in \mathbb{R}$ and $\ell \in [L]$.
3. Define $n_\ell = \sum_{\omega \in \Omega} \mathbb{1}_{\{y_\omega = \ell\}}$, i.e., the number of tensor entries taking value on $\ell \in [L]$.
4. With a little abuse of notation, we re-define the constants in (2) as

$$U_{\alpha,\beta,\Delta} = \max_{\substack{|\theta| \le \alpha, \\ \boldsymbol{b} \in \mathcal{B}}} \max_{\ell \in [L-1]} \frac{|\dot{g}_\ell(\theta)|}{g_\ell(\theta)}, \quad \text{and} \quad L_{\alpha,\beta,\Delta} = \min_{\substack{|\theta| \le \alpha, \\ \boldsymbol{b} \in \mathcal{B}}} \min_{\ell \in [L-1]} \frac{\dot{g}_\ell^2(\theta) - \ddot{g}_\ell(\theta) g_\ell(\theta)}{g_\ell^2(\theta)}. \tag{21}$$

5. We define two additional constants:

$$\begin{aligned}
C_{\alpha,\beta,\Delta} &= \max_{\substack{|\theta| \le \alpha, \\ \boldsymbol{b} \in \mathcal{B}}} \max_{\ell \in [L-1]} \max \left\{ \frac{\dot{f}(\theta + b_\ell)}{g_\ell(\theta)}, \frac{\dot{f}(\theta + b_\ell)}{g_{\ell+1}(\theta)} \right\}, \\
D_{\alpha,\beta,\Delta} &= \min_{\substack{|\theta| \le \alpha, \\ \boldsymbol{b} \in \mathcal{B}}} \min_{\ell \in [L-1]} \min \left\{ -\frac{\partial}{\partial \theta} \left(\frac{\dot{f}(\theta + b_\ell)}{g_\ell(\theta)} \right), \frac{\partial}{\partial \theta} \left(\frac{\dot{f}(\theta + b_\ell)}{g_{\ell+1}(\theta)} \right) \right\} \\
&= \min_{\substack{|\theta| \le \alpha, \\ \boldsymbol{b} \in \mathcal{B}}} \min_{\ell \in [L-1]} \min \left\{ -\frac{\ddot{f}(\theta + b_\ell) g_\ell(\theta) - \dot{f}(\theta + b_\ell) \dot{g}_\ell(\theta)}{g_\ell^2(\theta)}, \right. \\
&\qquad \left. \frac{\ddot{f}(\theta + b_\ell) g_{\ell+1}(\theta) - \dot{f}(\theta + b_\ell) \dot{g}_{\ell+1}(\theta)}{g_{\ell+1}^2(\theta)} \right\}.
\end{aligned} \tag{22}$$

We make the following assumptions about the link function.

Assumption 1. *The link function $f : \mathbb{R} \mapsto [0, 1]$ satisfies the following properties:*

1. *$f(\theta)$ is twice-differentiable and strictly increasing in θ.*
2. *$\dot{f}(\theta)$ is strictly log-concave and symmetric with respect to $\theta = 0$.*

3. *The function* $\frac{\dot{f}(\theta+b_\ell)}{g_\ell(\theta)}$ *is strictly decreasing with respect to* θ *for all* $\boldsymbol{b} \in \mathcal{B}$.

4. *The function* $\frac{\dot{f}(\theta+b_\ell)}{g_{\ell+1}(\theta)}$ *is strictly increasing with respect to* θ *for all* $\boldsymbol{b} \in \mathcal{B}$.

Remark 2. The condition $\Delta = \min_\ell(b_\ell - b_{\ell-1}) > 0$ on the feasible set $\mathcal{B}$ guarantees the strict positiveness of $g_\ell(\theta) = f(\theta + b_\ell) - f(\theta + b_{\ell-1})$. Therefore, the denominators in the above quantities $U_{\alpha,\beta,\Delta}, L_{\alpha,\beta,\Delta}, C_{\alpha,\beta,\Delta}, D_{\alpha,\beta,\Delta}$ are well-defined. Furthermore, by Theorem 1.1, $g_\ell(\theta)$ is strictly log-concave in θ for all $\boldsymbol{b} \in \mathcal{B}$. Based on Assumption 1 and closeness of the feasible set, we have $U_{\alpha,\beta,\Delta} > 0$, $L_{\alpha,\beta,\Delta} > 0$, $C_{\alpha,\beta,\Delta} > 0$, $D_{\alpha,\beta,\Delta} > 0$.

Remark 3. In particular, for logistic link $f(x) = \frac{1}{1+e^{-x}}$, we have

$$C_{\alpha,\beta,\Delta} = \max_{\substack{|\theta|\le\alpha,\\ \boldsymbol{b}\in\mathcal{B}}} \max_{\ell\in[L-1]} \max\left\{ \frac{1}{1-e^{-(b_\ell - b_{\ell-1})}} \frac{1+e^{\theta+b_{\ell-1}}}{1+e^{\theta+b_\ell}}, \ \frac{1}{e^{(b_{\ell+1}-b_\ell)}-1} \frac{1+e^{\theta+b_{\ell+1}}}{1+e^{\theta+b_\ell}} \right\} > 0,$$

$$D_{\alpha,\beta,\Delta} = \min_{\substack{|\theta|\le\alpha,\\ \boldsymbol{b}\in\mathcal{B}}} \min_{\ell\in[L-1]} \frac{1}{(1+e^{\theta+b_\ell})(1+e^{-(\theta+b_\ell)})} > 0.$$

Theorem 2.1 (Statistical convergence with unknown $\boldsymbol{b}$)**.** *Consider an ordinal tensor* $\mathcal{Y} \in [L]^{d_1\times\cdots\times d_K}$ *generated from model* (1) *with the link function* f *and parameters* $(\Theta^{true}, \boldsymbol{b}^{true}) \in \mathcal{P} \times \mathcal{B}$. *Suppose the link function* f *satisfies Assumption 1. Define* $r_{\max} = \max_k r_k$. *Then with very high probability, the estimator in* (20) *satisfies*

$$MSE(\hat{\Theta}, \Theta^{true}) \le \frac{c_1 U_{\alpha,\beta,\Delta}}{L_{\alpha,\beta,\Delta}} r_{\max}^{(K-1)/2} \sqrt{\frac{\sum_k d_k}{\prod_k d_k}},$$

and

$$MSE(\hat{\boldsymbol{b}}, \boldsymbol{b}^{true}) \le \frac{c_2 C_{\alpha,\beta,\Delta}}{D_{\alpha,\beta,\Delta}} \frac{\max_\ell(n_\ell + n_{\ell+1})}{\min_\ell(n_\ell + n_{\ell+1})} \sqrt{\frac{1}{L-1}},$$

where $c_1, c_2 > 0$ *are two constants independent of the tensor dimension and rank, and* $U_{\alpha,\beta,\Delta}$, $L_{\alpha,\beta,\Delta}$, $C_{\alpha,\beta,\Delta}$, $D_{\alpha,\beta,\Delta} > 0$ *are constants defined in* (21) *and* (22).

Remark 4. It may seem counter-intuitive that the MSE of $\hat{\Theta}$ shows no dependence of $\hat{\boldsymbol{b}}$ (and vice versa). In fact, from the proof of Theorem 2.1, the uniform constants $U_{\alpha,\beta,\Delta}, L_{\alpha,\beta,\Delta}$ can be replaced by

$$U_{\alpha,\hat{\boldsymbol{b}}} = \max_{|\theta|\le\alpha} \max_{\ell\in[L-1]} \frac{|\dot{g}_\ell(\theta)|}{g_\ell(\theta)}, \quad \text{and} \quad L_{\alpha,\hat{\boldsymbol{b}}} = \min_{|\theta|\le\alpha} \min_{\ell\in[L-1]} \frac{\dot{g}_\ell^2(\theta) - \ddot{g}_\ell(\theta) g_\ell(\theta)}{g_\ell^2(\theta)}.$$

Therefore, the accuracy of $\hat{\Theta}$ depends on the accuracy of $\hat{\boldsymbol{b}}$ implicitly through $U_{\alpha,\hat{\boldsymbol{b}}}$ and $L_{\alpha,\hat{\boldsymbol{b}}}$.

Proof of Theorem 2.1. Similar to the proof of Theorem 4.1, we suppress Ω in the subscript. Based on the definition of $(\hat{\Theta}, \hat{\boldsymbol{b}})$, we have the following inequalities:

$$\mathcal{L}_{\mathcal{Y}}(\hat{\Theta}, \hat{\boldsymbol{b}}) \ge \mathcal{L}_{\mathcal{Y}}(\Theta^{\text{true}}, \hat{\boldsymbol{b}}) \quad \text{and} \quad \mathcal{L}_{\mathcal{Y}}(\hat{\Theta}, \hat{\boldsymbol{b}}) \ge \mathcal{L}_{\mathcal{Y}}(\hat{\Theta}, \boldsymbol{b}^{\text{true}}). \tag{23}$$

Following the same argument as in Theorem 4.1 and the first inequality in (23), we obtain that

$$\text{MSE}(\hat{\Theta}, \Theta^{\text{true}}) \le \frac{C U_{\alpha,\beta,\Delta}}{L_{\alpha,\beta,\Delta}} r_{\max}^{(K-1)/2} \sqrt{\frac{\sum_k d_k}{\prod_k d_k}},$$

where $U_{\alpha,\beta,\Delta}, L_{\alpha,\beta,\Delta} > 0$ are two universal constants independent of the tensor dimension and rank. Next we bound $\mathrm{MSE}(\hat{\boldsymbol{b}}, \boldsymbol{b}^{\mathrm{true}})$. It follows from the expression of $\mathcal{L}_{\mathcal{Y}}(\Theta, \boldsymbol{b})$ that

$$\frac{\partial \mathcal{L}_{\mathcal{Y}}}{\partial b_\ell} = \sum_{\omega\in\Omega}\left[\mathbb{1}_{\{y_\omega=\ell\}}\frac{\dot f(\theta_\omega+b_\ell)}{g_\ell(\theta_\omega)} - \mathbb{1}_{\{y_\omega=\ell+1\}}\frac{\dot f(\theta_\omega+b_\ell)}{g_{\ell+1}(\theta_\omega)}\right],$$

$$\frac{\partial^2 \mathcal{L}_{\mathcal{Y}}}{\partial b_\ell^2} = \sum_{\omega\in\Omega}\left[\mathbb{1}_{\{y_\omega=\ell\}}\frac{\ddot f(\theta_\omega+b_\ell)g_\ell(\theta_\omega) - \dot f(\theta_\omega+b_\ell)^2}{g_\ell^2(\theta_\omega)} - \mathbb{1}_{\{y_\omega=\ell+1\}}\frac{\ddot f(\theta_\omega+b_\ell)g_{\ell+1}(\theta_\omega) + \dot f(\theta_\omega+b_\ell)^2}{g_{\ell+1}^2(\theta_\omega)}\right],$$

for all $\ell \in [L-1]$,

$$\frac{\partial^2 \mathcal{L}_{\mathcal{Y}}}{\partial b_\ell \partial b_{\ell+1}} = \sum_{\omega\in\Omega}\mathbb{1}_{\{y_\omega=\ell+1\}}\frac{\dot f(\theta_\omega+b_\ell)\dot f(\theta_\omega+b_{\ell+1})}{g_{\ell+1}^2(\theta_\omega)} \quad \text{and} \quad \frac{\partial^2 \mathcal{L}_{\mathcal{Y}}}{\partial b_\ell \partial b_{\ell'}'} = 0 \text{ if } |\ell-\ell'|>1. \tag{24}$$

Therefore, all entries in $\nabla_{\boldsymbol{b}}\mathcal{L}_{\mathcal{Y}}$ are upper bounded by $\{C_{\alpha,\beta,\Delta}\max_\ell(n_\ell+n_{\ell+1})\} > 0$, and $\nabla^2_{\boldsymbol{b}}\mathcal{L}_{\mathcal{Y}}$ is a tridiagonal matrix.

We consider the profile log-likelihood $\mathcal{L}_{\mathcal{Y}}(\boldsymbol{b}, \hat\Theta)$ as a function of $\boldsymbol{b} \in \mathcal{B}$. For notational convenience, we drop $\hat\Theta$ from $\mathcal{L}_{\mathcal{Y}}(\boldsymbol{b}, \hat\Theta)$ and simply write $\mathcal{L}_{\mathcal{Y}}(\boldsymbol{b})$. By the second-order Taylor's expansion of $\mathcal{L}_{\mathcal{Y}}(\boldsymbol{b})$ around $\boldsymbol{b}^{\mathrm{true}}$, we obtain

$$\mathcal{L}_{\mathcal{Y}}(\hat{\boldsymbol{b}}) = \mathcal{L}_{\mathcal{Y}}(\boldsymbol{b}^{\mathrm{true}}) + (\boldsymbol{b}^{\mathrm{true}}-\hat{\boldsymbol{b}})^T\nabla_{\boldsymbol{b}}\mathcal{L}_{\mathcal{Y}}(\boldsymbol{b}^{\mathrm{true}}) + \frac{1}{2}(\boldsymbol{b}^{\mathrm{true}}-\hat{\boldsymbol{b}})^T\nabla^2_{\check{\boldsymbol{b}}}\mathcal{L}_{\mathcal{Y}}(\check{\boldsymbol{b}})(\boldsymbol{b}^{\mathrm{true}}-\hat{\boldsymbol{b}}), \tag{25}$$

where $\check{\boldsymbol{b}} = \gamma\boldsymbol{b}^{\mathrm{true}} + (1-\gamma)\hat{\boldsymbol{b}}$ for some $\gamma \in [0,1]$, and $\nabla^2_{\check{\boldsymbol{b}}}\mathcal{L}_{\mathcal{Y}}(\check{\boldsymbol{b}})$ denotes the $(L-1)$-by-$(L-1)$ Hession matrix evaluated at $\check{\boldsymbol{b}}$.

The linear term in (25) can be bounded by Cauchy-Schwartz inequality,

$$(\boldsymbol{b}^{\mathrm{true}}-\hat{\boldsymbol{b}})^T\nabla_{\boldsymbol{b}}\mathcal{L}_{\mathcal{Y}}(\boldsymbol{b}^{\mathrm{true}}) \le \|\boldsymbol{b}^{\mathrm{true}}-\hat{\boldsymbol{b}}\|_F\|\nabla_{\boldsymbol{b}}\mathcal{L}_{\mathcal{Y}}(\boldsymbol{b}^{\mathrm{true}})\|_F \le \|\boldsymbol{b}^{\mathrm{true}}-\hat{\boldsymbol{b}}\|_F\sqrt{L-1}C_{\alpha,\beta,\Delta}\max_{\ell\in[L-1]}(n_\ell+n_{\ell+1}), \tag{26}$$

where the last inequality is followed from

$$\left|\frac{\partial \mathcal{L}_{\mathcal{Y}}}{\partial b_\ell}\Big|_{\boldsymbol{b}=\boldsymbol{b}^{\mathrm{true}}}\right| \le C_{\alpha,\beta,\Delta}\max_{\ell\in[L-1]}(n_\ell+n_{\ell+1}), \quad \text{for all } \ell\in[L-1].$$

We next bound the quadratic term in (25). Note that

$$\begin{aligned}
&(\boldsymbol{b}^{\mathrm{true}}-\hat{\boldsymbol{b}})^T\nabla^2_{\check{\boldsymbol{b}}}\mathcal{L}_{\mathcal{Y}}(\check{\boldsymbol{b}})(\boldsymbol{b}^{\mathrm{true}}-\hat{\boldsymbol{b}}) \qquad (27)\\
&= \sum_{\ell\in[L-1]}\left(\frac{\partial^2\mathcal{L}_{\mathcal{Y}}}{\partial b_\ell^2}\Big|_{\boldsymbol{b}=\check{\boldsymbol{b}}}\right)(\hat b_\ell - b_\ell^{\mathrm{true}})^2 + 2\sum_{\ell\in[L-1]/\{1\}}\left(\frac{\partial^2\mathcal{L}_{\mathcal{Y}}}{\partial b_\ell\partial b_{\ell-1}}\Big|_{\boldsymbol{b}=\check{\boldsymbol{b}}}\right)(\hat b_\ell - b_\ell^{\mathrm{true}})(\hat b_{\ell-1} - b_{\ell-1}^{\mathrm{true}})\\
&\le \sum_{\ell\in[L-1]}\left(\frac{\partial^2\mathcal{L}_{\mathcal{Y}}}{\partial b_\ell^2}\Big|_{\boldsymbol{b}=\check{\boldsymbol{b}}}\right)(\hat b_\ell - b_\ell^{\mathrm{true}})^2 + \sum_{\ell\in[L-1]/\{1\}}\left(\frac{\partial^2\mathcal{L}_{\mathcal{Y}}}{\partial b_\ell\partial b_{\ell-1}}\Big|_{\boldsymbol{b}=\check{\boldsymbol{b}}}\right)\left[(\hat b_\ell - b_\ell^{\mathrm{true}})^2 + (\hat b_{\ell-1} - b_{\ell-1}^{\mathrm{true}})^2\right]\\
&= \left(\frac{\partial^2\mathcal{L}_{\mathcal{Y}}}{\partial b_1^2} + \frac{\partial^2\mathcal{L}_{\mathcal{Y}}}{\partial b_1\partial b_2}\right)\Big|_{\boldsymbol{b}=\check{\boldsymbol{b}}}(\hat b_1 - b_1^{\mathrm{true}})^2 + \left(\frac{\partial^2\mathcal{L}_{\mathcal{Y}}}{\partial b_{L-1}^2} + \frac{\partial^2\mathcal{L}_{\mathcal{Y}}}{\partial b_{L-2}\partial b_{L-1}}\right)\Big|_{\boldsymbol{b}=\check{\boldsymbol{b}}}(\hat b_{L-1} - b_{L-1}^{\mathrm{true}})^2\\
&\quad + \sum_{\ell\in[L-2]/\{1\}}\left(\frac{\partial^2\mathcal{L}_{\mathcal{Y}}}{\partial b_\ell^2} + \frac{\partial^2\mathcal{L}_{\mathcal{Y}}}{\partial b_\ell\partial b_{\ell-1}} + \frac{\partial^2\mathcal{L}_{\mathcal{Y}}}{\partial b_{\ell+1}\partial b_\ell}\right)\Big|_{\boldsymbol{b}=\check{\boldsymbol{b}}}(\hat b_\ell - b_\ell^{\mathrm{true}})^2\\
&\le -\tilde D_{\alpha,\beta,\Delta}\sum_{\ell\in[L-1]}(\hat b_\ell - b_{\mathrm{true},\ell})^2
\end{aligned}$$

$$= -\tilde{D}_{\alpha,\beta,\Delta}\|\hat{\boldsymbol{b}} - \boldsymbol{b}^{\text{true}}\|_F^2,$$

where

$$\begin{aligned}
\tilde{D}_{\alpha,\beta,\Delta} &= \min_{\substack{|\theta|\le\alpha,\\ \boldsymbol{b}\in\mathcal{B}}} \min_{\ell\in[L-1]} -\left(\frac{\partial^2\mathcal{L}_{\mathcal{Y}}}{\partial b_\ell^2} + \frac{\partial^2\mathcal{L}_{\mathcal{Y}}}{\partial b_\ell \partial b_{\ell-1}} + \frac{\partial^2\mathcal{L}_{\mathcal{Y}}}{\partial b_{\ell+1}\partial b_\ell}\right) \\
&= \min_{\substack{|\theta|\le\alpha,\\ \boldsymbol{b}\in\mathcal{B}}} \min_{\ell\in[L-1]} \left\{\sum_{\omega\in\Omega} -\mathbb{1}_{\{y_\omega=\ell\}}\left(\frac{\ddot{f}(\theta_\omega+b_\ell)g_\ell(\theta_\omega) - \dot{f}(\theta_\omega+b_\ell)\dot{g}_\ell(\theta_\omega)}{g_\ell^2(\theta_\omega)}\right)\right. \\
&\qquad\qquad \left. + \sum_{\omega\in\Omega}\mathbb{1}_{\{y_\omega=\ell+1\}}\left(\frac{\ddot{f}(\theta_\omega+b_\ell)g_{\ell+1}(\theta_\omega) - \dot{f}_{\theta_\omega+b_\ell}\dot{g}_{\ell+1}(\theta_\omega)}{g_{\ell+1}^2(\theta_\omega)}\right)\right\} \\
&= \min_{\substack{|\theta|\le\alpha,\\ \boldsymbol{b}\in\mathcal{B}}} \min_{\ell\in[L-1]} \left\{\underbrace{-\sum_{\omega\in\Omega}\mathbb{1}_{\{y_\omega=\ell\}}\frac{\partial}{\partial\theta_\omega}\left(\frac{\dot{f}(\theta_\omega+b_\ell)}{g_\ell(\theta_\omega)}\right)}_{>0} + \underbrace{\sum_{\omega\in\Omega}\mathbb{1}_{\{y_\omega=\ell+1\}}\frac{\partial}{\partial\theta_\omega}\left(\frac{\dot{f}(\theta_\omega+b_\ell)}{g_{\ell+1}(\theta_\omega)}\right)}_{>0}\right\} \\
&\ge D_{\alpha,\beta,\Delta}\min_{\ell\in[L-1]}(n_\ell+n_{\ell+1})\,.
\end{aligned}$$

Combining inequalities (25), (26) and (27) yields

$$\mathcal{L}_{\mathcal{Y}}(\hat{\boldsymbol{b}}) \le \mathcal{L}_{\mathcal{Y}}(\boldsymbol{b}^{\text{true}}) + C_{\alpha,\beta,\Delta}\sqrt{L-1}\max_\ell(n_\ell+n_{\ell+1})\|\hat{\boldsymbol{b}}-\boldsymbol{b}^{\text{true}}\|_F - \frac{D_{\alpha,\beta,\Delta}}{2}\min_\ell(n_\ell+n_{\ell+1})\|\hat{\boldsymbol{b}}-\boldsymbol{b}^{\text{true}}\|_F^2.$$

Since $\hat{\boldsymbol{b}}$ satisfies $\mathcal{L}_{\mathcal{Y}}(\hat{\boldsymbol{b}}) - \mathcal{L}_{\mathcal{Y}}(\boldsymbol{b}^{\text{true}}) \ge 0$, we have that

$$C_{\alpha,\beta,\Delta}\sqrt{L-1}\max_\ell(n_\ell+n_{\ell+1})\|\hat{\boldsymbol{b}}-\boldsymbol{b}^{\text{true}}\|_F - \frac{D_{\alpha,\beta,\Delta}}{2}\min_\ell(n_\ell+n_{\ell+1})\|\hat{\boldsymbol{b}}-\boldsymbol{b}^{\text{true}}\|_F^2 \ge 0$$

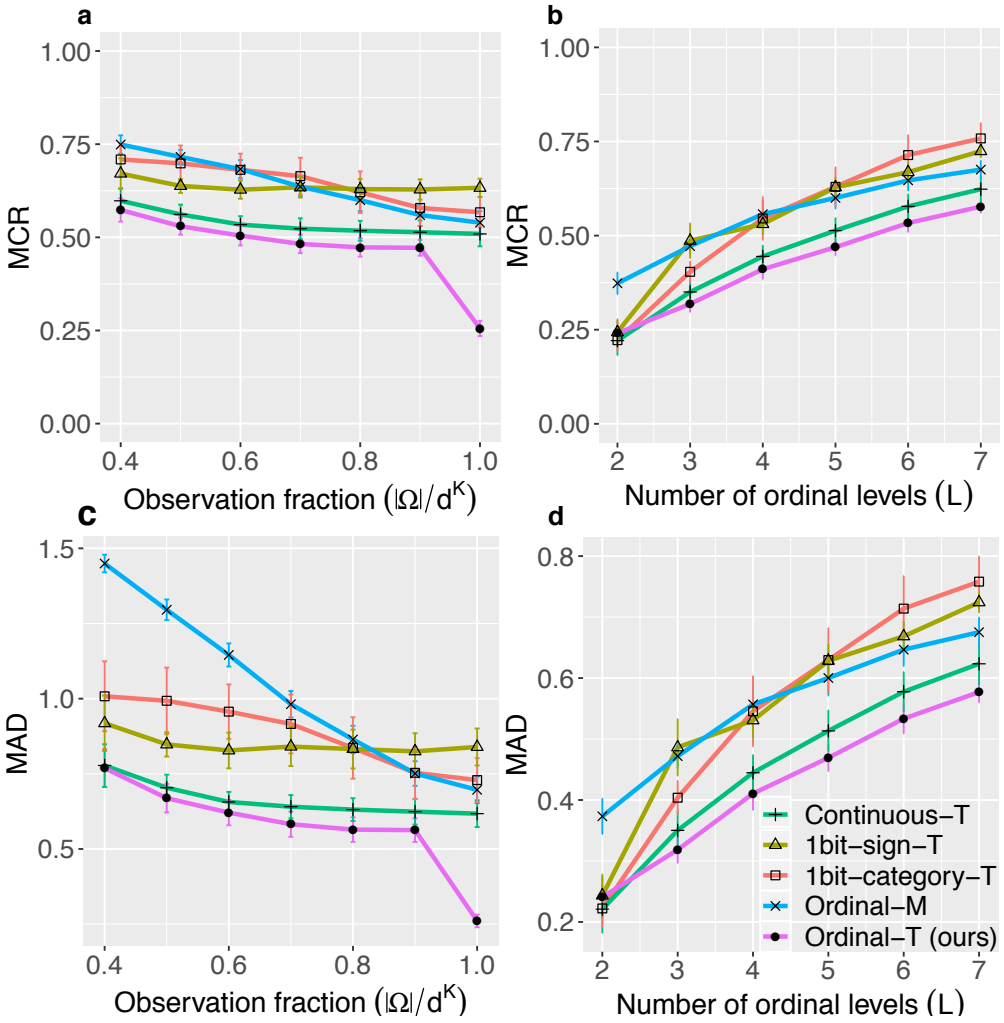
Finally,

$$\text{MSE}(\hat{\boldsymbol{b}}, \boldsymbol{b}^{\text{true}}) = \frac{1}{L-1}\|\hat{\boldsymbol{b}}-\boldsymbol{b}^{\text{true}}\|_F \le \frac{2C_{\alpha,\beta,\Delta}}{D_{\alpha,\beta,\Delta}}\frac{\max_\ell(n_\ell+n_{\ell+1})}{\min_\ell(n_\ell+n_{\ell+1})}\sqrt{\frac{1}{L-1}},$$

which completes the proof. □

3 Additional results on simulations

In the main paper, we compare our method (**Ordinal-T**) with other four methods (**Ordinal-M**, **1bit-category-T**,**1bit-sign-T**, and **Continuous-T**). The methods are evaluated by their capabilities in predicting the most likely label for each entry, i.e., $y_\omega^{\text{mode}} = \arg\max_\ell \mathbb{P}(y_\omega = \ell)$. Here we perform additional simulations to compare the methods by their performance in predicting the median label, $y_\omega^{\text{median}} = \min\{\ell\colon \mathbb{P}(y_\omega = \ell) \ge 0.5\}$. Under the latent variable model (4) and Assumption 1, the median label is the quantized Θ without noise; i.e. $y_\omega^{\text{median}} = \sum_\ell \mathbb{1}_{\theta_\omega\in(b_{\ell-1},b_\ell]}$.
[FIXME (miaoyan): Please add a few sentence to describe Figure S]



Supplementary Figure S1: Performance comparison for predicting the median label. (a,c) Prediction errors versus sample complexity $\rho = |\Omega|/d^k$ when $L = 5$. (b,d) Prediction errors versus the number of ordinal levels L, when $\rho = 0.8$.

4 Additional results on HCP analysis

4.1 Clustering based on Tucker representation

We perform clustering analyses based on the Tucker representation of the estimated tensor parameter $\hat{\Theta}$. The procedure is motivated from the higher-order extension of Principal Component Analysis (PCA) or Singular Value Decomposition (SVD). Recall that, in the matrix case, we usually perform clustering on an $m \times n$ (normalized) matrix X based on the following procedure. First, we factorize X into

$$X = U\Sigma V^T,$$

where Σ is a diagonal matrix and U, V are factor matrices with orthogonal columns. Second, we take each column of V as a principal axis and each row in $U\Sigma$ as principal component. A subsequent multivariate clustering method (such as K-means) is then applied to the m rows of $U\Sigma$.

We apply a similar clustering procedure to the estimated parameter tensor $\hat{\Theta}$. Based on Tucker representation of $\hat{\Theta}$, we have

$$\hat{\Theta} = \hat{\mathcal{C}} \times_1 \hat{\boldsymbol{M}}_1 \times_2 \cdots \times_K \hat{\boldsymbol{M}}_K, \tag{28}$$

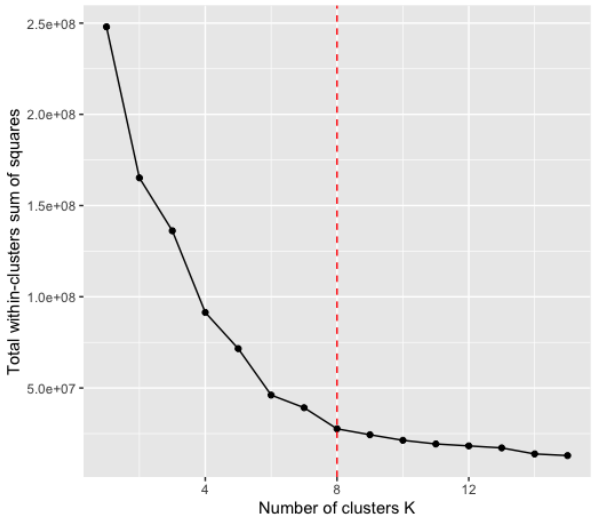
where $\hat{\mathcal{C}} \in \mathbb{R}^{r_1 \times \cdots \times r_K}$ is the estimated core tensor, $\hat{\boldsymbol{M}}_k \in \mathbb{R}^{d_k \times r_k}$ are estimated factor matrices with orthogonal columns, and $\times_k$ denotes the tensor-by-matrix multiplication [Kolda and Bader, 2009]. The mode-k matricization of (28) gives

$$\hat{\Theta}_{(k)} = \hat{\boldsymbol{M}}_k \hat{\mathcal{C}}_{(k)} \left(\hat{\boldsymbol{M}}_K \otimes \cdots \otimes \hat{\boldsymbol{M}}_1 \right),$$

where $\hat{\Theta}_{(k)}, \hat{\mathcal{C}}_{(k)}$ denote the mode-k unfolding of $\hat{\Theta}$ and $\hat{\mathcal{C}}$, respectively. Then, the mode-k clustering can be performed as follows. First, we take colums in $\left(\hat{\boldsymbol{M}}_K \otimes \cdots \otimes \hat{\boldsymbol{M}}_1 \right)$ as principal axes and rows in $\hat{\boldsymbol{M}}_k \hat{\mathcal{C}}_{(k)}$ as principal components. Then, we perform K-means clustering method to the d_k rows of the matrix $\hat{\boldsymbol{M}}_k \hat{\mathcal{C}}_{(k)}$.

4.2 Clustering results on HCP

We perform clustering analysis on the 68 brain nodes using the the produce described in Section 4.1. Recall that our ordinal tensor method outputs the estimated parameter tensor $\hat{\Theta} \in \mathbb{R}^{68 \times 68 \times 136}$ with rank $(23, 23, 8)$. We apply K-means to the mode-1 principal component matrix of size 68×184 ($184 = 23 \times 8$). The elbow method suggests eight clusters among the 68 nodes (see Figure 4.2). Table 4.2 shows the eight clusters and their compositions.



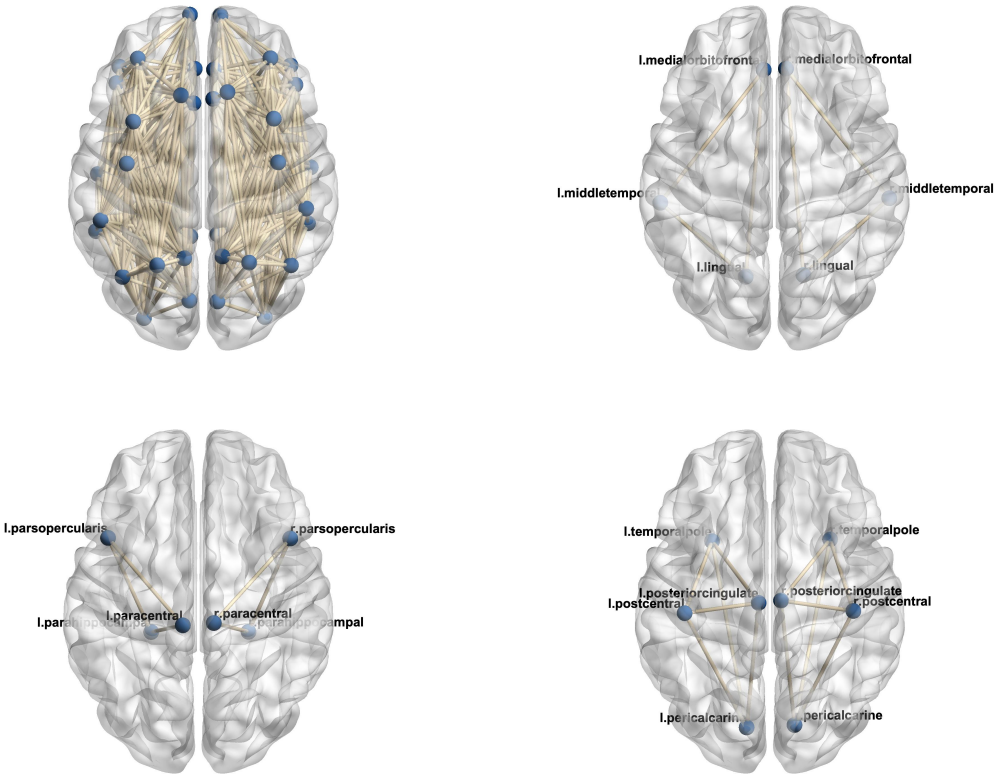
Supplementary Figure S2: The elbow plot for the number of clusters

We find that most brain nodes fall into cluster I and the cluster II, which can be represented as the left and right hemispheres, respectively (Figure 4.2a). From cluster III to the cluster VIII, entries within each cluster share the same name encoded in the data. For example, cluster III and cluster IV represent the *Supramarginal gyrus* region on the left brain and right brain. These regions are known to be in charge of interpreting tactile sensory data and involved in perception of space and limbs location Carlson [2012], Reed and Caselli [1994]. **[FIXME (Miaoyan): (I do not understand**

the figure and text. Why are there four panels in the Figure? Did you just plot all pairwise connections in Cluster I? Please change the edge weights using averaged connection strengths. Step 1. take average of connections over individuals. 2. plot the averaged strength between nodes within a same cluster. You also need to tune the sparsity level to make the figure interpretable...) (Texts need to be rewritten...) Figure 4.2 displays the averaged connection within each cluster brain nodes. Each node is connected by edges within the same cluster and the groups can be paired as the right part and the left part of the brain.] The results demonstrate that our clustering successfully groups similar brain nodes together without knowledge of external annotation.

Cluster	I			II		
Brain nodes	"RMF_L","'Fpole_L","Insula_L", "SupF_L","CaudMF_L", "parsTriangularis_L","parsOpercularis_L", "preCentral_L","Tpole_L","SupT_L", "SupT_L","postCentral_L", "SupP_L", "IP_L","LO_L","MOF_L","SupF_L", "IsthmusC_L","Precuneus_L", "Cuneus_L","paraHippo_L", "Lingual_L","SupT_L","LO_L"			"RMF_R","Fpole_R","Insula_R", "SupF_R", "CaudMF_R", "parsTriangularis_R" ,"parsOpercularis_R", "preCentral_R", "Tpole_R","SupT_R", "SupT_R","postCentral_R","SupP_R", "IP_R","LO_R","MOF_R", "SupF_R", "IsthmusC_R","Precuneus_R", "Cuneus_R","paraHippo_R", "Lingual_R","SupT_R","LO_R"		
Cluster	III	IV	V	VI	VII	VIII
Brain nodes	"SupraM_L" "SupraM_L" "SupraM_L" "SupraM_L"	"SupraM_R" "SupraM_R" "SupraM_R" "SupraM_R"	"IT_L" "IT_L" "IT_L"	"IT_R" "IT_R" "IT_R"	"MT_L" "MT_L" "MT_L"	"MT_R" "MT_R" "MT_R"

Supplementary Table S1: Clustering result of brain nodes. The last alphabet in the node name indicates the left brain ('L') and the right brain ('R') Please use node names from Github.../../BrainNetViewer_20171031/Data/miaoyan/Desikan_miaoyan.node.txt. Please remove the quote.



Supplementary Figure S3: Averaged connection within each of the eight clusters. [**FIXME (Miaoyan): Nodes within each cluster are connected by edges?**]

References

Herm Jan Brascamp and Elliott H Lieb. On extensions of the brunn-minkowski and prékopa-leindler theorems, including inequalities for log concave functions, and with an application to the diffusion equation. In *Inequalities*, pages 441–464. Springer, 2002.

Neil R Carlson. *Physiology of behavior 11th edition*. Pearson, 2012.

Shmuel Friedland and Lek-Heng Lim. Nuclear norm of higher-order tensors. *Mathematics of Computation*, 87(311):1255–1281, 2018.

Navid Ghadermarzy, Yaniv Plan, and Özgür Yilmaz. Near-optimal sample complexity for convex tensor completion. *Information and Inference: A Journal of the IMA*, 8(3):577–619, 2019.

Bo Jiang, Fan Yang, and Shuzhong Zhang. Tensor and its Tucker core: the invariance relationships. *Numerical Linear Algebra with Applications*, 24(3):e2086, 2017.

Tamara G Kolda and Brett W Bader. Tensor decompositions and applications. *SIAM Review*, 51 (3):455–500, 2009.

Lek-Heng Lim. Singular values and eigenvalues of tensors: a variational approach. In *1st IEEE International Workshop on Computational Advances in Multi-Sensor Adaptive Processing, 2005.*, pages 129–132. IEEE, 2005.

Nam H Nguyen, Petros Drineas, and Trac D Tran. Tensor sparsification via a bound on the spectral norm of random tensors. *Information and Inference: A Journal of the IMA*, 4(3):195–229, 2015.

Catherine L Reed and Richard J Caselli. The nature of tactile agnosia: a case study. *Neuropsychologia*, 32(5):527–539, 1994.

Ryota Tomioka and Taiji Suzuki. Spectral norm of random tensors. *arXiv preprint arXiv:1407.1870*, 2014.

Alexandre B Tsybakov. Introduction to nonparametric estimation. revised and extended from the 2004 french original. translated by Vladimir Zaiats, 2009.

Miaoyan Wang, Khanh Dao Duc, Jonathan Fischer, and Yun S Song. Operator norm inequalities between tensor unfoldings on the partition lattice. *Linear Algebra and Its Applications*, 520: 44–66, 2017.