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

A. Proofs

Here, we provide proofs of the theoretical results presented in Sections 4.

A.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

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

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 (1), $\nabla^2_\Theta \mathcal{L}_{\mathcal{Y}}$ is a diagonal matrix. Recall that

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

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

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

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

We first bound the linear term in (2). Note that, by Lemma 4,

$$|\langle \text{Vec}(\nabla_\Theta \mathcal{L}_{\mathcal{Y}}(\Theta^{\text{true}}), \text{Vec}(\Theta - \Theta^{\text{true}}) \rangle| \leq \|\nabla_\Theta \mathcal{L}_{\mathcal{Y}}(\Theta^{\text{true}})\|_\sigma \|\Theta - \Theta^{\text{true}}\|_*, \tag{3}$$

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 (1) 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

$$\mathbb{E}(s_\omega) = 0, \quad |s_\omega| \leq U_\alpha, \quad \text{for all } \omega \in [d_1] \times \cdots \times [d_K]. \tag{4}$$

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

$$\|\nabla_\Theta \mathcal{L}_{\mathcal{Y}}(\Theta^{\text{true}})\|_\sigma \leq C_2 U_\alpha \sqrt{\sum_k d_k}, \tag{5}$$

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 (3), (4) and (5), we have that, with probability at least $1 - \exp(-C_1 \sum_k d_k)$,

$$|\langle \text{Vec}(\nabla_\Theta \mathcal{L}_\mathcal{Y}(\Theta^{\text{true}})), \text{Vec}(\Theta - \Theta^{\text{true}})\rangle| \leq C_2 U_\alpha \sqrt{r_{\max}^{K-1} \sum_k d_k} \|\Theta - \Theta^{\text{true}}\|_F. \tag{6}$$

We next bound the quadratic term in (2). Note that

$$\begin{aligned}
\text{Vec}(\Theta - \Theta^{\text{true}})^T \nabla^2_\Theta \mathcal{L}_\mathcal{Y}(\check{\Theta}) \, \text{Vec}(\Theta - \Theta^{\text{true}}) &= \sum_\omega \left(\frac{\partial^2 \mathcal{L}_\mathcal{Y}}{\partial \theta_\omega^2} \Big|_{\Theta = \check{\Theta}} \right) (\theta_\omega - \theta_{\text{true},\omega})^2 \\
&\leq -L_\alpha \sum_\omega (\Theta_\omega - \Theta_{\text{true},\omega})^2 \\
&= -L_\alpha \|\Theta - \Theta^{\text{true}}\|_F^2,
\end{aligned} \tag{7}$$

where the second line comes from the fact that $\|\check{\Theta}\|_\infty \leq \alpha$ and the definition of L_α.

Combining (2), (6) and (7), 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,

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

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 $\mathcal{X} = \{\Theta_i \colon i = 1, \ldots\} \subset \mathcal{P}$ 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

$$\text{KL}(\mathbb{P}_\Theta || \mathbb{P}_{\mathbf{0}}) \leq C \|\Theta\|_F^2, \tag{8}$$

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 (8) with property (iii) of $\mathcal{X}$, we have

$$\text{KL}(\mathbb{P}_\Theta || \mathbb{P}_{\mathbf{0}}) \leq \gamma^2 r_{\max} d_{\max}. \tag{9}$$

From (9) and the property (i), we deduce that the condition

$$\frac{1}{\text{Card}(\mathcal{X})-1}\sum_{\Theta\in\mathcal{X}}\text{KL}(\mathbb{P}_{\Theta},\mathbb{P}_{\mathbf{0}})\leq\varepsilon\log_2\{\text{Card}(\mathcal{X})-1\}\tag{10}$$

holds for any $\varepsilon\geq 0$ when $\gamma\in[0,1]$ is chosen to be sufficiently small depending on ε, e.g., $\gamma\leq\sqrt{\frac{\varepsilon\log 2}{8}}$. By applying Lemma 11 to (10), and in view of the property (iv), we obtain that

$$\inf_{\hat\Theta}\sup_{\Theta^{\text{true}}\in\mathcal{X}}\mathbb{P}\left(\|\hat\Theta-\Theta^{\text{true}}\|_F\geq\frac{\gamma}{8}\min\left\{\alpha\sqrt{d_{\text{total}}},C^{-1/2}\sqrt{r_{\max}d_{\max}}\right\}\right)\geq\frac{1}{2}\left(1-2\varepsilon-\sqrt{\frac{16\varepsilon}{r_{\max}d_{\max}\log 2}}\right).\tag{11}$$

Note that $\text{Loss}(\hat\Theta,\Theta^{\text{true}})=\|\hat\Theta-\Theta^{\text{true}}\|_F^2/d_{\text{total}}$ and $\mathcal{X}\subset\mathcal{P}$. By taking $\varepsilon=1/10$ and $\gamma=1/11$, we conclude from (11) that

$$\inf_{\hat\Theta}\sup_{\Theta^{\text{true}}\in\mathcal{P}}\mathbb{P}\left(\text{Loss}(\hat\Theta,\Theta^{\text{true}})\geq c\min\left\{\alpha^2,\frac{C^{-1}r_{\max}d_{\max}}{d_{\text{total}}}\right\}\right)\geq\frac{1}{2}\left(\frac{4}{5}-\sqrt{\frac{1.6}{r_{\max}d_{\max}\log 2}}\right)\geq\frac{1}{8},$$

where $c=\frac{1}{88^2}$ and the last inequality comes from the condition for $d_{\max}$. This completes the proof. □

A.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

$$\mathcal{L}_{\mathcal{Y},\Omega}(\Theta)=\mathcal{L}_{\mathcal{Y},\Omega}(\Theta^{\text{true}})+\langle\text{Vec}(\nabla_\Theta\mathcal{L}_{\mathcal{Y},\Omega}),\ \text{Vec}(\Theta-\Theta^{\text{true}})\rangle+\frac{1}{2}\text{Vec}(\Theta-\Theta^{\text{true}})^T\nabla_\Theta^2\mathcal{L}_{\mathcal{Y},\Omega}(\check\Theta)\,\text{Vec}(\Theta-\Theta^{\text{true}}),\tag{12}$$

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 (3) and (7), we have

$$|\langle\text{Vec}(\nabla_\Theta\mathcal{L}_{\mathcal{Y},\Omega}),\ \text{Vec}(\Theta-\Theta^{\text{true}})\rangle|\leq C_2U_\alpha\sqrt{r_{\max}^{K-1}\sum_k d_k}\|\Theta-\Theta^{\text{true}}\|_{F,\Omega}$$

and

$$\text{Vec}(\Theta-\Theta^{\text{true}})^T\nabla_\Theta^2\mathcal{L}_{\mathcal{Y}}(\check\Theta)\,\text{Vec}(\Theta-\Theta^{\text{true}})\leq -L_\alpha\|\Theta-\Theta^{\text{true}}\|_{F,\Omega}^2.\tag{13}$$

Combining (12)-(13) with the fact that $\mathcal{L}_{\mathcal{Y},\Omega}(\hat\Theta)\geq\mathcal{L}_{\mathcal{Y},\Omega}(\Theta^{\text{true}})$, we have

$$\|\hat\Theta-\Theta^{\text{true}}\|_{F,\Omega}\leq\frac{2C_2U_\alpha r_{\max}^{(K-1)/2}}{L_\alpha}\sqrt{\sum_k d_k},\tag{14}$$

with probability at least $1-\exp(-C_1\sum_k d_k)$. Lastly, we invoke the result regarding the closeness of Θ to its sampled version Θ_Ω, under the entrywise bound condition. Note that $\|\hat\Theta-\Theta^{\text{true}}\|_\infty\leq 2\alpha$ and $\text{rank}(\hat\Theta-\Theta^{\text{true}})\leq 2\boldsymbol{r}$. By Lemma 2, $\|\hat\Theta-\Theta^{\text{true}}\|_M\leq 2^{(3K-1)/2}\alpha\left(\frac{\prod r_k}{r_{\max}}\right)^{3/2}$. 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 (14) gives

$$\begin{aligned}\|\hat\Theta-\Theta^{\text{true}}\|_{F,\Pi}^2&\leq\frac{1}{m}\|\hat\Theta-\Theta^{\text{true}}\|_{F,\Omega}^2+c\beta\sqrt{\frac{\sum_k d_k}{|\Omega|}}\\&\leq C_2r_{\max}^{K-1}\frac{\sum_k d_k}{|\Omega|}+C_1\alpha r_{\max}^{3(K-1)/2}\sqrt{\frac{\sum_k d_k}{|\Omega|}},\end{aligned}$$

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 $r_{\max} = O(1)$. $\square$

A.3. Convexity of the log-likelihood function

Theorem A.1. *Define the function*

$$\mathcal{L}_{\mathcal{Y},\Omega}(\Theta, \boldsymbol{b}) = \sum_{\omega\in\Omega}\sum_{\ell\in[L]} \Big(\mathbb{1}\{y_\omega=\ell\}\log\left[f(b_\ell-\theta_\omega)-f(b_{\ell-1}-\theta_\omega)\right]\Big), \tag{15}$$

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

Proof. Define $d_{\text{total}} = \prod_k d_k$. By 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

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

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 (15), 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

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

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 $u>v$. $\square$

Lemma 1 (Corollary 3.5 in Brascamp & Lieb (2002))**.** *Let $F(x,y)\colon \mathbb{R}^{m+n} \to \mathbb{R}$ 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.

A.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 $T_{\pm} = \{\mathcal{T} \in \{\pm 1\}^{d_1\times\cdots\times d_K}\colon \text{rank}(\mathcal{T}) = 1\}$. The atomic M-norm of a tensor $\Theta \in \mathbb{R}^{d_1\times\cdots\times d_K}$ is defined as

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

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 & 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 $\|\boldsymbol{x}_k^{(i)}\|_2 = 1$ 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

$$\|\mathcal{A}\|_* \le \sqrt{\frac{\prod_{k\ge 2} r_k}{\max_{k\ge 2} r_k}}\|\mathcal{A}_{(1)}\|_*, \tag{16}$$

where $\mathcal{A}_{(1)}$ is a d_1-by-$\prod_{k\ge 2} d_k$ matrix with matrix rank r_1. Furthermore, the relationship between the matrix norms implies that $\|\mathcal{A}_{(1)}\|_* \le \sqrt{r_1}\|\mathcal{A}_{(1)}\|_F = \sqrt{r_1}\|\mathcal{A}\|_F$. Combining this fact with the inequality (16) yields the final claim. □

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

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

Proof. By Friedland & 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

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

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 & Suzuki (2014).

Lemma 5 (Spectral norm of random tensors (Tomioka & 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*

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

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

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

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*

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

Then we have

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

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.

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

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

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

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. $\square$

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

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

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

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

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

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

$\square$

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 $\|\Theta\|_\infty \le \alpha$. Define*

$$A_\alpha = \min_{\ell\in[L],|\theta|\le\alpha} \left[f(b_\ell-\theta)-f(b_{\ell-1}-\theta)\right]. \tag{17}$$

Then, for any two tensors $\Theta,\ \Theta^*$ in the parameter spaces, we have

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

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),

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

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,

$$\begin{aligned}
\mathrm{KL}(\theta||\theta^*) &\le \sum_{\ell\in[L]} \frac{\left[f(b_\ell-\theta)-f(b_{\ell-1}-\theta)-f(b_\ell-\theta^*)+f(b_{\ell-1}-\theta^*)\right]^2}{f(b_\ell-\theta^*)-f(b_{\ell-1}-\theta^*)}\\
&\le \sum_{\ell=2}^{L-1} \frac{\left[\dot{f}(b_\ell-\eta_\ell)-\dot{f}(b_{\ell-1}-\eta_{\ell-1})\right]^2}{f(b_\ell-\theta^*)-f(b_{\ell-1}-\theta^*)}(\theta-\theta^*)^2 + \frac{\dot{f}^2(b_1-\eta_1)}{f(b_1-\theta^*)}(\theta-\theta^*)^2\\
&\quad + \frac{\dot{f}^2(b_{L-1}-\eta_{L-1})}{1-f(b_{L-1}-\theta^*)}(\theta-\theta^*)^2,
\end{aligned}$$

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

$$\mathrm{KL}(\theta||\theta^*) \le \left(\frac{4(L-2)}{A_\alpha}+\frac{2}{A_\alpha}\right)\dot{f}^2(0)(\theta-\theta^*)^2 = \frac{2(2L-3)}{A_\alpha}\dot{f}^2(0)(\theta-\theta^*)^2, \tag{18}$$

where we have used Taylor expansion, the bound (17), and the fact that $\dot{f}(\cdot)$ peaks at zero for an unimodal and symmetric function. Now summing (18) over the index set $\omega\in[d_1]\times\cdots\times[d_K]$ gives

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

□

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

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

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 $0\le\gamma\le 1$, there exist a finite set of tensors $\mathcal{X}=\{\Theta_i : i=1,\ldots\}\subset\mathcal{P}$ satisfying the following four properties:*

1. *Card($\mathcal{X}$) $\ge 2^{Rd_1/8}+1$, where Card denotes the cardinality;*
2. *$\mathcal{X}$ contains the zero tensor $\mathbf{0}\in\mathbb{R}^{d_1\times\cdots\times d_K}$;*
3. *$\|\Theta\|_\infty \le \gamma\min\left\{\alpha,\ C^{-1/2}\sqrt{\frac{Rd_1}{d_{total}}}\right\}$ for any element $\Theta\in\mathcal{X}$;*
4. *$\|\Theta_i-\Theta_j\|_F \ge \frac{\gamma}{4}\min\left\{\alpha\sqrt{d_{total}},\ C^{-1/2}\sqrt{Rd_1}\right\}$ 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 $0 \le \gamma \le 1$, we define a set of matrices:

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

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 $\max_k r_k \le R$, 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 $\text{Card}(\mathcal{X}) \ge 2^{Rd_1/8} + 1$ containing the zero $d_1 \times \cdots \times d_K$ tensor, such that, for any two distinct elements Θ_i and Θ_j in $\mathcal{X}$,

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

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*

$$\|w^{(j)} - w^{(k)}\|_0 \ge \frac{m}{8}, \quad \textit{for } 0 \le j < k \le M,$$

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

Lemma 11 (Theorem 2.5 in Tsybakov (2008))**.** *Assume that a set* $\mathcal{X}$ *contains element* $\Theta_0, \Theta_1, \ldots, \Theta_M$ *(*$M \ge 2$*) such that*

- $d(\Theta_j, \ \Theta_k) \ge 2s > 0$*,* $\forall 0 \le j \le k \le M$*;*
- $\mathbb{P}_0$ *is absolutely continuous with respect to* $\mathbb{P}_j$*,* $\forall j = 1, \ldots, M$*, and*

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

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

Then

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

Lemma 12 (Lemma 28 in Ghadermarzy et al. (2019))**.** *Define* $\mathbb{B}_M(\beta) = \{\Theta \in \mathbb{R}^{d_1 \times \cdots \times d_K} \colon \|\Theta\|_M \le \beta\}$*. 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*,*

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

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

B. Extension of Theorem 4.1 to unknown cut-off points

We now extend Theorem 4.1 to the case of unknown cut-off points $\boldsymbol{b}$. Assume that the true parameters $(\Theta^{\text{true}}, \boldsymbol{b}^{\text{true}}) \in \mathcal{P} \times \mathcal{B}$, where the feasible sets are defined as

$$\mathcal{P} = \{\Theta \in \mathbb{R}^{d_1 \times \cdots \times d_K} : \text{rank}(\mathcal{P}) \leq \boldsymbol{r},\ \langle \Theta, \mathcal{J} \rangle = 0,\ \|\Theta\|_\infty \leq \alpha\},$$
$$\mathcal{B} = \{\boldsymbol{b} \in \mathbb{R}^{L-1} : \|\boldsymbol{b}\|_\infty \leq \beta,\ \min_\ell (b_\ell - b_{\ell-1}) \geq \Delta\},$$

with positive constants $\alpha, \beta, \Delta > 0$ and a given rank $\boldsymbol{r} \in \mathbb{N}_+^K$. Here, $\mathcal{J} = [\![1]\!] \in \mathbb{R}^{d_1 \times \cdots \times d_K}$ denotes a tensor of all ones. The constraint $\langle \Theta, \mathcal{J} \rangle = 0$ is imposed to ensure the identifiability of Θ and $\boldsymbol{b}$. We propose the constrained M-estimator

$$(\hat{\Theta}, \hat{\boldsymbol{b}}) = \underset{(\Theta, \boldsymbol{b}) \in \mathcal{P} \times \mathcal{B}}{\arg\max}\ \mathcal{L}_{\mathcal{Y}}(\Theta, \boldsymbol{b}). \tag{19}$$

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

$$\text{MSE}\left(\hat{\Theta}, \Theta^{\text{true}}\right) = \frac{1}{\prod_k d_k} \|\hat{\Theta} - \Theta^{\text{true}}\|_F^2, \quad \text{MSE}\left(\hat{\boldsymbol{b}}, \boldsymbol{b}^{\text{true}}\right) = \frac{1}{L-1} \|\hat{\boldsymbol{b}} - \boldsymbol{b}^{\text{true}}\|_F^2.$$

To facilitate the examination of MSE, we define an order-$(K+1)$ tensor, $\mathcal{Z} = [\![z_{\omega,\ell}]\!] \in \mathbb{R}^{d_1 \times \cdots \times d_K \times (L-1)}$, by stacking the parameters $\Theta = [\![\theta_\omega]\!]$ and $\boldsymbol{b} = [\![b_\ell]\!]$ together. Specifically, let $z_{\omega,\ell} = -\theta_\omega + b_\ell$ for all $\omega \in [d_1] \times \cdots \times [d_K]$ and $\ell \in [L-1]$; that is,

$$\mathcal{Z} = -\Theta \otimes \mathbf{1} + \mathcal{J} \otimes \boldsymbol{b},$$

where $\mathbf{1}$ denotes a length-$(L-1)$ vector of all ones. Under the identifiability constraint $\langle \Theta, \mathcal{J} \rangle = 0$, there is an one-to-one mapping between $\mathcal{Z}$ and $(\Theta, \boldsymbol{b})$, with $\text{rank}(\mathcal{Z}) \leq (\text{rank}(\Theta) + 1,\ 2)^T$. Furthermore,

$$\|\hat{\mathcal{Z}} - \mathcal{Z}^{\text{true}}\|_F^2 = \|\hat{\Theta} - \Theta^{\text{true}}\|_F^2 (L-1) + \|\hat{\boldsymbol{b}} - \boldsymbol{b}^{\text{true}}\|_F^2 \left(\prod_k d_k\right), \tag{20}$$

where $\mathcal{Z}^{\text{true}} = -\Theta^{\text{true}} \otimes \mathbf{1} + \mathcal{J} \otimes \boldsymbol{b}^{\text{true}}$ and $\hat{\mathcal{Z}} = -\hat{\Theta} \otimes \mathbf{1} + \mathcal{J} \otimes \hat{\boldsymbol{b}}$.

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

We define the following constants that will be used in the theory:

$$C_{\alpha,\beta,\Delta} = \max_{|z| \leq \alpha+\beta} \max_{\substack{z' \leq z-\Delta \\ z'' \geq z+\Delta}} \max \left\{ \frac{\dot{f}(z)}{f(z) - f(z')},\ \frac{\dot{f}(z)}{f(z'') - f(z)} \right\}, \tag{21}$$
$$D_{\alpha,\beta,\Delta} = \min_{|z| \leq \alpha+\beta} \min_{\substack{z' \leq z-\Delta \\ z'' \geq z+\Delta}} \min \left\{ -\frac{\partial}{\partial z} \left(\frac{\dot{f}(z)}{f(z) - f(z')} \right),\ \frac{\partial}{\partial z} \left(\frac{\dot{f}(z)}{f(z'') - f(z)} \right) \right\},$$
$$A_{\alpha,\beta,\Delta} = \min_{|z| \leq \alpha+\beta} \min_{z' \leq z-\Delta} \left(f(z) - f(z') \right).$$

Remark 2. The condition $\Delta = \min_\ell (b_\ell - b_{\ell-1}) > 0$ on the feasible set $\mathcal{B}$ guarantees the strict positiveness of $f(z) - f(z')$ and $f(z'') - f(z)$. Therefore, the denominators in the above quantities $C_{\alpha,\beta,\Delta},\ D_{\alpha,\beta,\Delta}$ are well-defined. Furthermore, by Theorem A.1, $f(z) - f(z')$ is strictly log-concave in (z, z') for $z \leq z' - \Delta,\ z, z' \in [-\alpha - \beta,\ \alpha + \beta]$. Based on Assumption 1 and closeness of the feasible set, we have $C_{\alpha,\beta,\Delta} > 0$, $D_{\alpha,\beta,\Delta} > 0$, $A_{\alpha,\beta,\Delta} > 0$.

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

$$C_{\alpha,\beta,\Delta} = \max_{|z|\leq\alpha+\beta} \max_{\substack{z'\leq z-\Delta \\ z''\geq z+\Delta}} \max\left\{ \frac{1}{e^{\Delta}-1}\left(\frac{1+e^{-z'}}{1+e^{-z}}\right),\ \frac{1}{1-e^{-\Delta}}\left(\frac{1+e^{-z''}}{1+e^{-z}}\right)\right\} > 0,$$

$$D_{\alpha,\beta,\Delta} = \min_{|z|\leq\alpha+\beta} \frac{e^{-z}}{(1+e^{-z})^2} > 0.$$

Theorem B.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 $(\Theta^{\text{true}}, \boldsymbol{b}^{\text{true}}) \in \mathcal{P}\times\mathcal{B}$. Suppose the link function f satisfies Assumption 1. Define $r_{\max} = \max_k r_k + 1$, and assume $r_{\max} = \mathcal{O}(1)$.*

Then with very high probability, the estimator in (19) satisfies

$$\|\hat{\mathcal{Z}} - \mathcal{Z}^{\text{true}}\|_F^2 \leq \frac{c_1 r_{\max}^K C^2_{\alpha,\beta,\Delta}}{A^2_{\alpha,\beta,\Delta} D^2_{\alpha,\beta,\Delta}}\left(L-1+\sum_k d_k\right), \tag{22}$$

In particular,

$$\text{MSE}\left(\hat{\Theta}, \Theta^{\text{true}}\right) \leq \min\left\{4\alpha^2,\ \frac{c_1 r_{\max}^K C^2_{\alpha,\beta,\Delta}}{A^2_{\alpha,\beta,\Delta} D^2_{\alpha,\beta,\Delta}}\left(\frac{L-1+\sum_k d_k}{\prod_k d_k}\right)\right\},$$

and

$$\text{MSE}\left(\hat{\boldsymbol{b}}, \boldsymbol{b}^{\text{true}}\right) \leq \min\left\{4\beta^2,\ \frac{c_1 r_{\max}^K C^2_{\alpha,\beta,\Delta}}{A^2_{\alpha,\beta,\Delta} D^2_{\alpha,\beta,\Delta}}\left(\frac{L-1+\sum_k d_k}{\prod_k d_K}\right)\right\},$$

where $c_1, C_{\alpha,\beta,\Delta}, D_{\alpha,\beta,\Delta}$ are positive constants independent of the tensor dimension, rank, and number of ordinal levels.

Proof. The log-likelihood associated with the observed entries in terms of $\mathcal{Z}$ is

$$\mathcal{L}_{\mathcal{Y}}(\mathcal{Z}) = \sum_{\omega\in\Omega}\sum_{\ell\in[L]} \mathbb{1}\{y_\omega = \ell\}\log\left[f(z_{\omega,\ell}) - f(z_{\omega,\ell-1})\right].$$

Let $\nabla_{\mathcal{Z}}\mathcal{L}_{\mathcal{Y}} = [\![\frac{\partial \mathcal{L}_{\mathcal{Y}}}{\partial z_{\omega,\ell}}]\!] \in \mathbb{R}^{d_1\times\cdots\times d_K\times[L-1]}$ denote the score function, and $\boldsymbol{H} = \nabla^2_{\mathcal{Z}}\mathcal{L}_{\mathcal{Y}}$ the Hession matrix. Based on the definition of $\hat{\mathcal{Z}}$, we have the following inequality:

$$\mathcal{L}_{\mathcal{Y}}(\hat{\mathcal{Z}}) \geq \mathcal{L}_{\mathcal{Y}}(\mathcal{Z}^{\text{true}}). \tag{23}$$

Following the similar argument in Theorem 4.1 and the inequality (23), we obtain that

$$\|\hat{\mathcal{Z}} - \mathcal{Z}^{\text{true}}\|_F^2 \leq c_1 r_{\max}^K \frac{\|\nabla_{\mathcal{Z}}\mathcal{L}_{\mathcal{Y}}(\mathcal{Z}^{\text{true}})\|_\sigma^2}{\lambda_1^2\left(\boldsymbol{H}(\check{\mathcal{Z}})\right)}, \tag{24}$$

where $\nabla_{\mathcal{Z}}\mathcal{L}_{\mathcal{Y}}(\mathcal{Z}^{\text{true}})$ is the score evaluated at $\mathcal{Z}^{\text{true}}$, $\boldsymbol{H}(\check{\mathcal{Z}})$ is the Hession evaluated at $\check{\mathcal{Z}}$, for some $\check{\mathcal{Z}}$ between $\hat{\mathcal{Z}}$ and $\mathcal{Z}^{\text{true}}$, and $\lambda_1(\cdot)$ is the largest matrix eigenvalue.

We bound the score and the Hessian to obtain (22).

1. (Score.) The (ω,ℓ)-th entry in $\nabla_{\mathcal{Z}}\mathcal{L}_{\mathcal{Y}}$ is

$$\frac{\partial \mathcal{L}_{\mathcal{Y}}}{\partial z_{\omega,\ell}} = \mathbb{1}\{y_\omega=\ell\}\frac{\dot{f}(z)}{f(z)-f(z')}\Bigg|_{(z,\ z')=(z_{\omega,\ell},\ z_{\omega,\ell-1})} - \mathbb{1}\{y_\omega=\ell+1\}\frac{\dot{f}(z)}{f(z'')-f(z)}\Bigg|_{(z'',\ z)=(z_{\omega,\ell+1},\ z_{\omega,\ell})},$$

   which is upper bounded in magnitude by $C_{\alpha,\beta,\Delta} > 0$ with zero mean. By Lemma 6, with probability at least $1-\exp\left(-c_2'\left(\sum_k d_k + L - 1\right)\right)$, we have

$$\|\nabla_{\mathcal{Z}}\mathcal{L}_{\mathcal{Y}}(\mathcal{Z}^{\text{true}})\|_\sigma \leq c_2 C_{\alpha,\beta,\Delta}\sqrt{L-1+\sum_k d_k}, \tag{25}$$

   where c_2, c_2' are two positive constants that depend only on K.

2. (Hession.) The entries in the Hession matrix are

$$\text{Diagonal: } \frac{\partial^2 \mathcal{L}_{\mathcal{Y}}}{\partial z_{\omega,\ell}^2} = \mathbb{1}\{y_\omega = \ell\} \frac{\ddot{f}(z)\left(f(z) - f(z')\right) - \dot{f}^2(z)}{\left(f(z) - f(z')\right)^2}\Bigg|_{(z,\, z')=(z_{\omega,\ell},\, z_{\omega,\ell-1})} -$$

$$\mathbb{1}\{y_\omega = \ell+1\} \frac{\ddot{f}(z)\left(f(z'') - f(z)\right) + \dot{f}^2(z)}{\left(f(z'') - f(z)\right)^2}\Bigg|_{(z'',\, z)=(z_{\omega,\ell+1},\, z_{\omega,\ell})},$$

$$\text{Off-diagonal: } \frac{\partial^2 \mathcal{L}_{\mathcal{Y}}}{\partial z_{\omega,\ell} z_{\omega,\ell+1}} = \mathbb{1}\{y_\omega = \ell+1\} \frac{\dot{f}(z_{\omega,\ell})\dot{f}(z_{\omega,\ell+1})}{\left(f(z_{\omega,\ell+1}) - f(z_{\omega,\ell})\right)^2} \quad \text{and} \quad \frac{\partial^2 \mathcal{L}_{\mathcal{Y}}}{\partial z_{\omega,\ell} z_{\omega',\ell'}} = 0 \text{ otherwise.}$$

Based on Assumption 1, the Hession matrix $\boldsymbol{H}$ has the following three properties:

(a) The Hession matrix is a block matrix, $\boldsymbol{H} = \operatorname{diag}\{\boldsymbol{H}_\omega \colon \omega \in [d_1] \times \cdots \times [d_K]\}$, and each block $\boldsymbol{H}_\omega \in \mathbb{R}^{(L-1)\times(L-1)}$ is a tridiagonal matrix.

(b) The off-diagonal entries are either zero or strictly positive.

(c) The diagonal entries are either zero or strictly negative. Furthermore,

$$\begin{aligned}
&\boldsymbol{H}_\omega(\ell,\ell) + \boldsymbol{H}_\omega(\ell,\ell-1) + \boldsymbol{H}_\omega(\ell,\ell+1) \\
=&\frac{\partial^2 \mathcal{L}_{\mathcal{Y}}}{\partial z_{\omega,\ell}^2} + \frac{\partial^2 \mathcal{L}_{\mathcal{Y}}}{\partial z_{\omega,\ell} z_{\omega,\ell+1}} + \frac{\partial^2 \mathcal{L}_{\mathcal{Y}}}{\partial z_{\omega,\ell-1} z_{\omega,\ell}} \\
=&\mathbb{1}\{y_\omega = \ell\} \frac{\partial}{\partial z}\left(\frac{\dot{f}(z)}{f(z) - f(z')}\right)\Bigg|_{(z,\, z')=(z_{\omega,\ell},\, z_{\omega,\ell-1})} - \mathbb{1}\{y_\omega = \ell+1\} \frac{\partial}{\partial z}\left(\frac{\dot{f}(z)}{f(z) - f(z')}\right)\Bigg|_{(z'',\, z)=(z_{\omega,\ell+1},\, z_{\omega,\ell})} \\
\le& - D_{\alpha,\beta,\Delta} \mathbb{1}\{y_\omega = \ell \text{ or } \ell+1\}.
\end{aligned}$$

We will show that, with very high probability over $\mathcal{Y}$, $\boldsymbol{H}$ is negative definite in that

$$\lambda_1(\boldsymbol{H}) = \max_{\boldsymbol{z}\neq 0} \frac{\boldsymbol{z}^T \boldsymbol{H} \boldsymbol{z}}{\|\boldsymbol{z}\|_F^2} \le -c A_{\alpha,\beta,\Delta} D_{\alpha,\beta,\Delta}, \tag{26}$$

where $A_{\alpha,\beta,\Delta},\ D_{\alpha,\beta,\Delta} > 0$ are constants defined in (21), and $c > 0$ is a constant.

Let $\boldsymbol{z}_\omega = (z_{\omega,1}, \ldots, z_{\omega,L-1})^T \in \mathbb{R}^{L-1}$ and $\boldsymbol{z} = (\boldsymbol{z}_{1,\ldots,1,1}, \ldots, \boldsymbol{z}_{d_1,\ldots,d_K,L-1})^T \in \mathbb{R}^{(L-1)\prod_k d_k}$. It follows from property (a) that

$$\boldsymbol{z}^T \boldsymbol{H} \boldsymbol{z} = \sum_\omega \boldsymbol{z}_\omega^T \boldsymbol{H}_\omega \boldsymbol{z}_\omega.$$

Furthermore, from properties (b) and (c) we have

$$\begin{aligned}
\boldsymbol{z}_\omega^T \boldsymbol{H}_\omega \boldsymbol{z}_\omega &= \sum_{\ell\in[L-1]} \boldsymbol{H}_\omega(\ell,\ell) z_{\omega,\ell}^2 + \sum_{\ell\in[L-1]/\{1\}} 2\boldsymbol{H}_\omega(\ell,\ell-1) z_{\omega,\ell} z_{\omega,\ell-1} \\
&\le \sum_{\ell\in[L-1]} \boldsymbol{H}(\ell,\ell) z_{\omega,\ell}^2 + \sum_{\ell\in[L-1]/\{1\}} \boldsymbol{H}(\ell,\ell-1)\left[z_{\omega,\ell}^2 + z_{\omega,\ell-1}^2\right] \\
&= \left(\boldsymbol{H}(1,1) + \boldsymbol{H}(1,2)\right) z_{\omega,1}^2 + \left(\boldsymbol{H}(L-1,L-1) + \boldsymbol{H}(L-1,L-2)\right) z_{\omega,L-1}^2 \\
&\quad + \sum_{\ell\in[L-2]/\{1\}} \left(\boldsymbol{H}(\ell,\ell) + \boldsymbol{H}(\ell,\ell-1) + \boldsymbol{H}(\ell,\ell+1)\right) z_{\omega,\ell}^2 \\
&\le -D_{\alpha,\beta,\Delta} \sum_\ell z_{\omega,\ell}^2 \mathbb{1}\{y_\omega = \ell \text{ or } \ell+1\}.
\end{aligned}$$

Therefore,

$$\boldsymbol{z}^T \boldsymbol{H} \boldsymbol{z} = \sum_\omega \boldsymbol{z}_\omega^T \boldsymbol{H}_\omega \boldsymbol{z}_\omega \le -D_{\alpha,\beta,\Delta} \sum_\omega \sum_\ell z_{\omega,\ell}^2 \mathbb{1}\{y_\omega = \ell \text{ or } \ell+1\}. \tag{27}$$

Define the subspace:

$$\mathcal{S} = \{\text{Vec}(\mathcal{Z}) : \mathcal{Z} = -\Theta \otimes \mathbf{1} + \mathcal{J} \otimes \boldsymbol{b},\ (\Theta, \boldsymbol{b}) \in (\mathcal{P}, \mathcal{B})\}.$$

It suffices to prove the negative definiteness of Hession when restricted in the subspace $\mathcal{S}$. Specifically, for any vector $\boldsymbol{z} = [\![z_{\omega,\ell}]\!] \in \mathcal{S}$,

$$\begin{aligned}
\sum_{\omega,\ell} z_{\omega,\ell}^2 \mathbb{1}\{y_\omega = \ell \text{ or } \ell+1\} &= \sum_{\omega,\ell} (-\theta_\omega + b_\ell)^2 \mathbb{1}\{y_\omega = \ell \text{ or } \ell+1\} \\
&= \sum_{\omega,\ell} (\theta_\omega^2 - 2\theta_\omega b_\ell + b_\ell^2) \mathbb{1}\{y_\omega = \ell \text{ or } \ell+1\} \\
&= \sum_{\omega,\ell} \theta_\omega^2 \mathbb{1}\{y_\omega = \ell \text{ or } \ell+1\} - 2\sum_{\omega,\ell} \theta_\omega b_\ell \mathbb{1}\{y_\omega = \ell \text{ or } \ell+1\} + \sum_{\omega,\ell} b_\ell^2 \mathbb{1}\{y_\omega = \ell \text{ or } \ell+1\} \\
&\geq \sum_{\omega} \theta_\omega^2 - 2\sum_{\omega,\ell} \theta_\omega b_\ell + \sum_{\ell} b_\ell^2 \left(n_\ell + n_{\ell+1}\right) \\
&\geq \sum_{\omega} \theta_\omega^2 + \min_\ell \left(n_\ell + n_{\ell+1}\right) \sum_{\ell} b_\ell^2
\end{aligned}$$

On the other hand,

$$\|\boldsymbol{z}\|_F^2 = \sum_{\omega,\ell} z_{\omega,\ell}^2 = \sum_{\omega,\ell} (-\theta_\omega + b_\ell)^2 = L_{\text{total}} \sum_{\omega} \theta_\omega^2 + d_{\text{total}} \sum_{\ell} b_\ell^2,$$

where $L_{\text{total}} := (L-1)$ and $d_{\text{total}} := \prod_k d_k$.

Therefore, we have

$$\begin{aligned}
\max_{\boldsymbol{z}\in\mathcal{S}, \boldsymbol{z}\neq \mathbf{0}} \frac{\sum_{\omega,\ell} z_{\omega,\ell}^2 \mathbb{1}\{\{y_\omega = \ell \text{ or } \ell+1\}\}}{\|\boldsymbol{z}\|_F^2} &\geq \frac{\sum_\omega \theta_\omega^2 + \min_\ell \left(n_\ell + n_{\ell+1}\right) \sum_\ell b_\ell^2}{L_{\text{total}} \sum_\omega \theta_\omega^2 + d_{\text{total}} \sum_\ell b_\ell^2} \geq \frac{\min_\ell (n_\ell + n_{\ell+1})}{(1 + \frac{\alpha^2}{c\Delta^2}) d_{\text{total}}} \\
&\geq \frac{2A_{\alpha,\beta,\Delta}}{1 + \frac{\alpha^2}{c\Delta^2}} \quad \text{in high probability as } d_{\min} \to \infty. \qquad (28)
\end{aligned}$$

The second inequality in (28) is from the conditions that

$$\sum_\omega \theta_\omega^2 \in [0,\ \alpha^2 d_{\text{total}}] \quad \text{and} \quad \sum_\ell b_\ell^2 \in [cL_{\text{total}}\Delta^2,\ L_{\text{total}}\beta^2],$$

for some universal constant $c > 0$. The last inequality in (28) follows by applying the law of large numbers and the uniform bound $\min_{z_{\omega,\ell}} \mathbb{P}(y_\omega = \ell \text{ or } \ell+1 | z_{\omega,\ell}) \geq 2A_{\alpha,\beta,\Delta}$ to the empirical ratio:

$$\frac{\min_\ell (n_\ell + n_{\ell+1})}{d_{\text{total}}} \xrightarrow{p} \min_\ell \mathbb{P}(y_\omega = \ell \text{ or } \ell+1 | z_{\omega,\ell}) \geq 2A_{\alpha,\beta,\Delta}, \quad \text{in high probability as } d_{\min} \to \infty.$$

By (27) and (28), we have

$$\boldsymbol{z}^T \boldsymbol{H} \boldsymbol{z} \leq -c' A_{\alpha,\beta,\Delta} D_{\alpha,\beta,\Delta} \|\boldsymbol{z}\|_F^2,$$

for some constant $c' > 0$, therefore (26) is proved.

Plugging (25) and (26) into (24) yields

$$\|\hat{\mathcal{Z}} - \mathcal{Z}^{\text{true}}\|_F^2 \leq \frac{c_1 r_{\max}^K C_{\alpha,\beta,\Delta}^2}{A_{\alpha,\beta,\Delta}^2 D_{\alpha,\beta,\Delta}^2} \left(L - 1 + \sum_k d_k\right).$$

The MSEs for $\hat{\Theta}$ and $\hat{\boldsymbol{b}}$ readily follow from (20).

□

C. Additional explanations of HCP analysis

We perform clustering analyses based on the Tucker representation of the estimated signal tensor $\hat{\Theta}$. 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 perform clustering on an $m \times n$ (normalized) matrix $\boldsymbol{X}$ based on the following procedure. First, we factorize X into

$$\boldsymbol{X} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^T,$$

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

We apply a similar clustering procedure to the estimated signal tensor $\hat{\Theta}$. We factorize $\hat{\Theta}$ based on Tucker decomposition.

$$\hat{\Theta} = \hat{\mathcal{C}} \times_1 \hat{\boldsymbol{M}}_1 \times_2 \cdots \times_K \hat{\boldsymbol{M}}_K, \tag{29}$$

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 & Bader, 2009). The mode-k matricization of (29) 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. We conduct clustering on this the mode-k unfolded signal tensor. We take columns 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 apply K-means clustering method to the d_k rows of the matrix $\hat{\boldsymbol{M}}_k \hat{\mathcal{C}}_{(k)}$.

We perform a clustering analysis on the 68 brain nodes using the procedure described above. 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 11 clusters among the 68 nodes (see Figure 1). The clustering result is presented in Section 7.

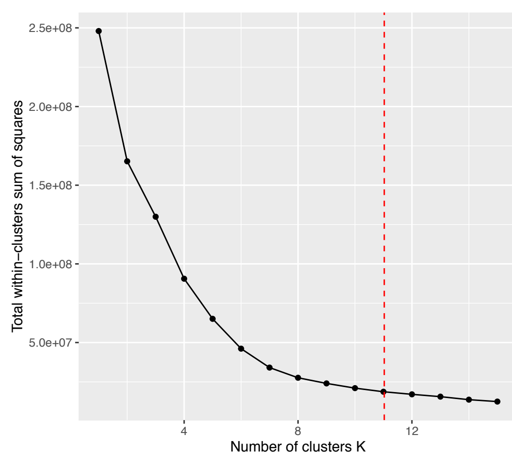

Figure 1. Elbow plot for determining the number of clusters in K-means.

References

Brascamp, H. J. and Lieb, E. H. 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*, pp. 441–464. Springer, 2002.

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

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

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

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

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

Nguyen, N. H., Drineas, P., and Tran, T. D. 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.

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

Tsybakov, A. B. *Introduction to nonparametric estimation*. Springer Science & Business Media, 2008.

Wang, M., Duc, K. D., Fischer, J., and Song, Y. S. Operator norm inequalities between tensor unfoldings on the partition lattice. *Linear Algebra and Its Applications*, 520:44–66, 2017.