

## Correction of Theorem 0.1

Miaoyan Wang, March 1, 2020

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

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

Here,  $\mathcal{J} = \llbracket 1 \rrbracket \in \mathbb{R}^{d_1 \times \dots \times d_K}$  denotes a tensor of all ones. The constraint  $\langle \Theta, \mathcal{J} \rangle = 0$  is imposed to ensure the identifiability of  $\Theta$  and  $\mathbf{b}$ . We propose the constrained M-estimator

$$(\hat{\Theta}, \hat{\mathbf{b}}) = \arg \max_{(\Theta, \mathbf{b}) \in \mathcal{P} \times \mathcal{B}} \mathcal{L}_{\mathcal{Y}}(\Theta, \mathbf{b}). \quad (1)$$

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

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

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

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

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  $(\Theta, \mathbf{b})$ , with  $\text{rank}(\mathcal{Z}) \leq (\text{rank}(\Theta), 2)^T$ . Furthermore,

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

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

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:

$$\begin{aligned}C_{\alpha, \beta, \Delta} &= \max_{|z| \leq \alpha + \beta} \max_{\substack{z' \leq z - \Delta \\ z'' \geq z + \Delta}} \max \left\{ \frac{\dot{f}(z)}{f(z) - f(z')}, \frac{\dot{f}(z)}{f(z'') - f(z)} \right\}, \\ D_{\alpha, \beta, \Delta} &= \max_{|z| \leq \alpha + \beta} \max_{\substack{z' \leq z - \Delta \\ z'' \geq z + \Delta}} \max \left\{ -\frac{\partial}{\partial z} \left( \frac{\dot{f}(z)}{f(z) - f(z')} \right), \frac{\partial}{\partial z} \left( \frac{\dot{f}(z)}{f(z'') - f(z)} \right) \right\}, \\ A_{\alpha, \beta, \Delta} &= \min_{|z| \leq \alpha + \beta} \min_{z' \leq z - \Delta} (f(z) - f(z')).\end{aligned} \quad (3)$$

**Remark 1.** 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 8.1,  $f(z) - f(z')$  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 2.** Add the specific bound for logistic link.

**Theorem 0.1** (Statistical convergence with unknown  $\mathbf{b}$ ). *Consider an ordinal tensor  $\mathcal{Y} \in [L]^{d_1 \times \dots \times d_K}$  generated from model (1) with the link function  $f$  and parameters  $(\Theta^{\text{true}}, \mathbf{b}^{\text{true}}) \in \mathcal{P} \times \mathcal{B}$ . Suppose the link function  $f$  satisfies Assumption 1. Define  $r_{\max} = \max_k r_k$ , and assume  $r_{\max} = \mathcal{O}(1)$ .*

*Then with very high probability, the estimator in (1) satisfies*

$$\|\hat{\mathcal{Z}} - \mathcal{Z}^{\text{true}}\|_F^2 \leq \frac{c_1 r_{\max}^{K-1} C_{\alpha,\beta,\Delta}^2}{A_{\alpha,\beta,\Delta}^2 D_{\alpha,\beta,\Delta}^2} \left( L - 1 + \sum_k d_k \right), \quad (4)$$

*In particular,*

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

*and*

$$\text{MSE}(\hat{\mathbf{b}}, \mathbf{b}^{\text{true}}) \leq \min \left( 4\beta^2, \frac{c_1 r_{\max}^{K-1} C_{\alpha,\beta,\Delta}^2}{A_{\alpha,\beta,\Delta}^2 D_{\alpha,\beta,\Delta}^2} \frac{L - 1 + \sum_k d_k}{(L - 1) \prod_k d_k} \right),$$

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.* (sketch)

Let  $\nabla_{\mathcal{Z}} \mathcal{L}_{\mathcal{Y}} = \llbracket \frac{\partial \mathcal{L}_{\mathcal{Y}}}{\partial z_{\omega,\ell}} \rrbracket \in \mathbb{R}^{d_1 \times \dots \times d_K \times [L-1]}$  denote the score function, and  $\mathbf{H} = \nabla_{\mathcal{Z}}^2 \mathcal{L}_{\mathcal{Y}}$  the Hessian matrix. Following the same argument in the previous version (Taylor expansion,  $r_{\max}(\mathcal{Z}) = r_{\max}(\Theta)$ , etc), we have

$$\|\hat{\mathcal{Z}} - \mathcal{Z}^{\text{true}}\|_F^2 \leq c_1 r_{\max}^{K-1} \frac{\|\nabla_{\mathcal{Z}} \mathcal{L}_{\mathcal{Y}}(\mathcal{Z}^{\text{true}})\|_{\sigma}^2}{\lambda_1^2(\mathbf{H}(\check{\mathcal{Z}}))}, \quad (5)$$

where  $\nabla_{\mathcal{Z}} \mathcal{L}_{\mathcal{Y}}(\mathcal{Z}^{\text{true}})$  is the score evaluated at  $\mathcal{Z}^{\text{true}}$ ,  $\mathbf{H}(\check{\mathcal{Z}})$  is the Hessian 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.

Hence, it suffices to bound the score and the Hessian.

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

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

which is upper bounded in magnitude by  $C_{\alpha,\beta,\Delta} > 0$ . Therefore, with very high probability,

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

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

$$\begin{aligned}
\text{Diagonal: } \frac{\partial^2 \mathcal{L}_Y}{\partial z_{\omega, \ell}^2} &= \mathbb{1}_{\{y_\omega = \ell\}} \frac{\ddot{f}(z) (f(z) - f(z')) - \dot{f}^2(z)}{(f(z) - f(z'))^2} \Big|_{(z, z') = (z_{\omega, \ell}, z_{\omega, \ell-1})} - \\
&\quad \mathbb{1}_{\{y_\omega = \ell+1\}} \frac{\ddot{f}(z) (f(z'') - f(z)) + \dot{f}^2(z)}{(f(z'') - f(z))^2} \Big|_{(z'', z) = (z_{\omega, \ell+1}, z_{\omega, \ell})}, \\
\text{Off-diagonal: } \frac{\partial^2 \mathcal{L}_Y}{\partial z_{\omega, \ell} \partial z_{\omega, \ell+1}} &= \mathbb{1}_{\{y_\omega = \ell+1\}} \frac{\dot{f}(z_{\omega, \ell}) \dot{f}(z_{\omega, \ell+1})}{(f(z_{\omega, \ell+1}) - f(z_{\omega, \ell}))^2} \quad \text{and} \quad \frac{\partial^2 \mathcal{L}_Y}{\partial z_{\omega, \ell} \partial z_{\omega', \ell'}} = 0 \text{ otherwise.}
\end{aligned}$$

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

- (a) The Hession matrix is a block matrix,  $\mathbf{H} = \text{diag}\{\mathbf{H}_\omega : \omega \in [d_1] \times \dots \times [d_K]\}$ , and each block  $\mathbf{H}_\omega \in \mathbb{R}^{(L-1) \times (L-1)}$  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,

$$\begin{aligned}
&\mathbf{H}_\omega(\ell, \ell) + \mathbf{H}_\omega(\ell, \ell-1) + \mathbf{H}_\omega(\ell, \ell+1) \\
&= \frac{\partial^2 \mathcal{L}_Y}{\partial z_{\omega, \ell}^2} + \frac{\partial^2 \mathcal{L}_Y}{\partial z_{\omega, \ell} \partial z_{\omega, \ell+1}} + \frac{\partial^2 \mathcal{L}_Y}{\partial z_{\omega, \ell-1} \partial z_{\omega, \ell}} \\
&= \mathbb{1}_{\{y_\omega = \ell\}} \frac{\partial}{\partial z} \left( \frac{\dot{f}(z)}{f(z) - f(z')} \right) \Big|_{(z, z') = (z_{\omega, \ell}, z_{\omega, \ell-1})} - \mathbb{1}_{\{y_\omega = \ell+1\}} \frac{\partial}{\partial z} \left( \frac{\dot{f}(z)}{f(z) - f(z')} \right) \Big|_{(z'', z) = (z_{\omega, \ell+1}, z_{\omega, \ell})} \\
&\leq -D_{\alpha, \beta, \Delta} < 0.
\end{aligned}$$

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

$$\lambda_1(\mathbf{H}) = \max_{\mathbf{z}} \frac{\mathbf{z}^T \mathbf{H} \mathbf{z}}{\|\mathbf{z}\|_F^2} \leq -c_2 A_{\alpha, \beta, \Delta} D_{\alpha, \beta, \Delta}, \quad (6)$$

where  $A_{\alpha, \beta, \Delta}, D_{\alpha, \beta, \Delta} > 0$  are constants defined in (3), and  $c_1 > 0$  is a constant.

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

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

Furthermore, properties (b) and (c) (or similar arguments as in page 29, arXiv preprint) imply that

$$\mathbf{z}_\omega^T \mathbf{H}_\omega \mathbf{z}_\omega \leq -D_{\alpha, \beta, \Delta} \sum_{\ell} z_{\omega, \ell}^2 \underbrace{\mathbb{1}_{\{y_\omega = \ell \text{ or } \ell+1\}}}_{\text{Bernoulli r.v. with probability bounded by } A_{\alpha, \beta, \Delta}}.$$

Therefore,

$$\mathbf{z}^T \mathbf{H} \mathbf{z} = \sum_{\omega} \mathbf{z}_\omega^T \mathbf{H}_\omega \mathbf{z}_\omega \leq -D_{\alpha, \beta, \Delta} \sum_{\omega} \sum_{\ell} z_{\omega, \ell}^2 \mathbb{1}_{\{y_\omega = \ell \text{ or } \ell+1\}}. \quad (7)$$

Based on central limit theorem (and concentration properties of Bernoulli r.v.'s), as the tensor dimension goes to infinity,

$$\sum_{\omega} \sum_{\ell} z_{\omega,\ell}^2 \mathbb{1}_{\{y_{\omega}=\ell \text{ or } \ell+1\}} \rightarrow \sum_{\omega} \sum_{\ell} z_{\omega,\ell}^2 \mathbb{P}(y_{\omega} = \ell \text{ or } \ell+1) \geq c_2 A_{\alpha,\beta,\Delta} \|\mathbf{z}\|_F^2 \quad (8)$$

holds with very high probability.

By (7) and (8), we have

$$\mathbf{z}^T \mathbf{H} \mathbf{z} \leq -c_2 A_{\alpha,\beta,\Delta} D_{\alpha,\beta,\Delta} \|\mathbf{z}\|_F^2,$$

and therefore (6) is proved. Plugging (4) and (6) into (5) yields

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

The MSEs for  $\hat{\Theta}$  and  $\hat{\mathbf{b}}$  readily follow from (2).

Correction of (8). Define the subspace:

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

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

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

On the other hand,

$$\|\mathbf{z}\|_F^2 = \sum_{\omega,\ell} z_{\omega,\ell}^2 = \sum_{\omega,\ell} (-\theta_{\omega} + b_{\ell})^2 = (L-1) \sum_{\omega} \theta_{\omega}^2 + \left( \prod_k d_k \right) \sum_{\ell} b_{\ell}^2.$$

Therefore, there exists a positive constant  $c_1 > 0$  (? perhaps depending on  $\alpha, \beta, \Delta$  etc...needs some calculation..) such that

$$\max_{\mathbf{z} \in \mathcal{S}, \mathbf{z} \neq \mathbf{0}} \frac{\sum_{\omega,\ell} z_{\omega,\ell}^2 \mathbb{1}_{\{y_{\omega}=\ell \text{ or } \ell+1\}}}{\|\mathbf{z}\|_F^2} \geq c_1 \frac{\min_{\ell} (n_{\ell} + n_{\ell+1})}{(L-1) \prod_k d_k} \geq c_1 \frac{\min_{\ell} (n_{\ell} + n_{\ell+1})}{(L-1) \prod_k d_k}.$$

The conclusion (8) then follows by applying the law of large number 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 ratio:

$$\frac{\min_{\ell} (n_{\ell} + n_{\ell+1})}{\prod_k d_k} \geq 2A_{\alpha,\beta,\Delta}, \quad \text{in high probability as } d_{\min} \rightarrow \infty.$$

