Supplementary files to "Community Detection in General Hypergraph via Graph Embedding"

Yaoming Zhen and Junhui Wang
School of Data Science
City University of Hong Kong

Proof of Proposition 1. Denote $S = \{\Theta \in \Omega | e_L(\Theta, \Theta^*) \geq \epsilon_n\}$. Let

$$
\begin{aligned}
I &:= P\Big(\sup_S \big(\mathcal{L}_\lambda(\Theta^*; \mathcal{A}) - \mathcal{L}_\lambda(\Theta; \mathcal{A})\big) \geq 0\Big) \\
&= P\Big(\sup_S \frac{1}{\varphi(n,m)} \sum_{\delta^{n+1,ord}_{i_1\dots i_m}=0} \big(L(\theta^*_{i_1\dots i_m}; a_{i_1\dots i_m}) - L(\theta_{i_1\dots i_m}; a_{i_1\dots i_m})\big) + \lambda_n\big(J(\boldsymbol{\alpha}^*) - J(\boldsymbol{\alpha})\big) \geq 0\Big).
\end{aligned}
$$

We now decompose S as follows. Let $S_u = \{\Theta \in \Omega | 2^{u-1}\epsilon_n \leq e_L(\Theta, \Theta^*) < 2^u \epsilon_n\}$, for $u = 1, 2, \dots$. It immediately follows that $S = \bigcup_{u=1}^{+\infty} S_u$. Consequently, by the union bound, we have

$$
\begin{aligned}
I &\leq \sum_{u=1}^{+\infty} P\Big(\sup_{S_u} \frac{1}{\varphi(n,m)} \sum_{\delta^{n+1,ord}_{i_1\dots i_m}=0} \big(L(\theta^*_{i_1\dots i_m}; a_{i_1\dots i_m}) - L(\theta_{i_1\dots i_m}; a_{i_1\dots i_m})\big) + \lambda_n\big(J(\boldsymbol{\alpha}^*) - J(\boldsymbol{\alpha})\big) \geq 0\Big) \\
&:= \sum_{u=1}^{+\infty} I_u.
\end{aligned}
$$

Define an empirical process

$$
\begin{aligned}
\nu_{n,m}(\Theta, \mathcal{A}) = \frac{1}{\varphi(n,m)} \sum_{\delta^{n+1,ord}_{i_1\dots i_m}=0} \Big(&L(\theta^*_{i_1\dots i_m}; a_{i_1\dots i_m}) - \\
&L(\theta_{i_1\dots i_m}; a_{i_1\dots i_m}) - E\big(L(\theta^*_{i_1\dots i_m}; a_{i_1\dots i_m}) - L(\theta_{i_1\dots i_m}; a_{i_1\dots i_m})\big)\Big),
\end{aligned}
$$

for some independent but not identical data. It then follows that

$$I_u \leq P\Big(\sup_{S_u} \nu_{n,m}(\Theta, \mathcal{A}) \geq \inf_{S_u} \big(e_L(\Theta, \Theta^*) + \lambda_n J(\boldsymbol{\alpha}) - \lambda_n J(\boldsymbol{\alpha}^*) \big) \Big).$$

Since $\inf_{S_u} \big(e_L(\Theta, \Theta^*) + \lambda_n J(\boldsymbol{\alpha}) - \lambda_n J(\boldsymbol{\alpha}^*) \big) \geq 2^{u-1}\epsilon_n - \frac{1}{2}\epsilon_n \geq 2^{u-2}\epsilon_n := M_u$ and Lemma 1 shows that $E \sup_{S_u} \nu_{n,m}(\Theta, \mathcal{A}) \leq M_u/2$ when n is large enough, we obtain

$$I_u \leq P\big(\sup_{S_u} \nu_{n,m}(\Theta, \mathcal{A}) \geq M_u \big) \leq P\big(\sup_{S_u} \nu_{n,m}(\Theta, \mathcal{A}) \geq E \sup_{S_u} \nu_{n,m}(\Theta, \mathcal{A}) + M_u/2 \big).$$

Note that for any $Y \sim \text{Bernoulli}(p)$ with $p = s_n\big(1 + \exp(-\theta)\big)^{-1}$, we have

$$\begin{aligned} E\big(L(\theta; Y) - L(\theta^*; Y)\big) &= -2p^* \log(\frac{p}{p^*})^{1/2} - 2(1-p^*) \log\big(\frac{1-p}{1-p^*}\big)^{1/2} \\ &\geq -2p^*\big(\frac{p^{1/2}}{(p^*)^{1/2}} - 1\big) - 2(1-p^*)\big(\frac{(1-p)^{1/2}}{(1-p^*)^{1/2}} - 1\big) \\ &= \big(p^{1/2} - (p^*)^{1/2}\big)^2 + \big((1-p)^{1/2} - (1-p^*)^{1/2}\big)^2, \end{aligned}$$

where $p^* = s_n\big(1 + \exp(-\theta^*)\big)^{-1}$. It immediately follows that on the set S_u, $D^2(\Theta, \Theta^*) \leq e_L(\Theta, \Theta^*) < 2^u \epsilon_n$.

We now turn to bound the variance of the empirical process by the averaged squared Hellinger distance $D^2(\Theta, \Theta^*)$ over S_u. Denote $\widetilde{\mathcal{P}}, \widetilde{\mathcal{P}}^*, \bar{\mathcal{P}}, \bar{\mathcal{P}}^*$ to be some tensors sharing the same shape as $\mathcal{P}$ with entries $\tilde{p}_{i_1 \ldots i_m} = p_{i_1 \ldots i_m}^{1/2}$, $\tilde{p}^*_{i_1 \ldots i_m} = (p^*_{i_1 \ldots i_m})^{1/2}$, $\bar{p}_{i_1 \ldots i_m} = (1 - p_{i_1 \ldots i_m})^{1/2}$, $\bar{p}^*_{i_1 \ldots i_m} = (1 - p^*_{i_1 \ldots i_m})^{1/2}$, respectively, if $\delta^{n+1,ord}_{i_1 \ldots i_m} = 0$ and zero otherwise. It is clear that

$$\frac{1}{\varphi(n,m)} \big(||\widetilde{\mathcal{P}} - \widetilde{\mathcal{P}}^*||_F^2 + ||\bar{\mathcal{P}} - \bar{\mathcal{P}}^*||_F^2 \big) = D^2(\Theta, \Theta^*) < 2^u \epsilon_n.$$

Applying these shorthand notations, we have

$$\frac{1}{\varphi(n,m)}\sum_{\delta^{n+1,ord}_{i_1,,,im}=0} E\big(L(\theta_{i_1\ldots i_m};a_{i_1\ldots i_m})-L(\theta^*_{i_1\ldots i_m};a_{i_1\ldots i_m})\big)^2$$
$$=\frac{4}{\varphi(n,m)}\sum_{\delta^{n+1,ord}_{i_1,,,im}=0}\Big((\tilde{p}^*_{i_1\ldots i_m})^2\big(\log\tilde{p}_{i_1\ldots i_m}-\log\tilde{p}^*_{i_1\ldots i_m}\big)^2+(\bar{p}^*_{i_1\ldots i_m})^2\big(\log\bar{p}_{i_1\ldots i_m}-\log\bar{p}^*_{i_1\ldots i_m}\big)^2\Big)$$
$$:=\tilde{g}(\widetilde{\mathcal{P}})+\bar{g}(\bar{\mathcal{P}}).$$

By Taylor's expansion of $\tilde{g}(\widetilde{\mathcal{P}})$ up to the second order at $\widetilde{P}^*$ over S_u, it holds true that

$$\tilde{g}(\widetilde{\mathcal{P}})=\frac{4}{\varphi(n,m)}||\widetilde{\mathcal{P}}-\widetilde{\mathcal{P}}^*||_F^2+o(\frac{1}{\varphi(n,m)}||\widetilde{\mathcal{P}}-\widetilde{\mathcal{P}}^*||_F^2).$$

Similarly,

$$\bar{g}(\bar{\mathcal{P}})=\frac{4}{\varphi(n,m)}||\bar{\mathcal{P}}-\bar{\mathcal{P}}^*||_F^2+o(\frac{1}{\varphi(n,m)}||\bar{\mathcal{P}}-\bar{\mathcal{P}}^*||_F^2).$$

Therefore,

$$\frac{1}{\varphi(n,m)}\sum_{\delta^{n+1,ord}_{i_1,,,im}=0} E\big(L(\theta_{i_1\ldots i_m};a_{i_1\ldots i_m})-L(\theta^*_{i_1\ldots i_m};a_{i_1\ldots i_m})\big)^2=4D^2(\Theta,\Theta^*)+o(D^2(\Theta,\Theta^*)).$$

On the set S_u, the variance of $\nu_{n,m}(\Theta,\mathcal{A})$ can be bounded as

$$Var\big(\nu_{n,m}(\Theta,\mathcal{A})\big)\le\frac{1}{\varphi^2(n,m)}\sum_{\delta^{n+1,ord}_{i_1\ldots i_m}=0} E\big(L(\theta_{i_1\ldots i_m};a_{i_1\ldots i_m})-L(\theta^*_{i_1..i_m};a_{i_1\ldots i_m})\big)^2$$
$$<\frac{5D^2(\Theta,\Theta^*)}{\varphi(n,m)}\le\frac{5\times 2^u\epsilon_n}{\varphi(n,m)}=\frac{20M_u}{\varphi(n,m)}.$$

Also note that $|L(\theta;Y)-L(\theta^*;Y)|$ can be upper bounded as

$$\big|L(\theta;Y)-L(\theta^*;Y)\big|\le\max\{\big|\log\frac{1+\exp(-\theta)}{1+\exp(-\theta^*)}\big|,\big|\log\frac{1-s_n\big(1+\exp(-\theta)\big)^{-1}}{1-s_n\big(1+\exp(-\theta^*)\big)^{-1}}\big|\}\le\log 2+c_0^m,$$

where the last inequality comes from the fact that

$$\frac{1}{2}\exp(-c_0^m) \le \frac{1+\exp(-\theta)}{1+\exp(-\theta^*)}, \frac{1-s_n\big(1+\exp(-\theta)\big)^{-1}}{1-s_n\big(1+\exp(-\theta^*)\big)^{-1}} \le 2\exp(c_0^m).$$

It then follows that

$$\frac{L(\theta^*_{i_1\ldots i_m}; a_{i_1\ldots i_m}) - L(\theta_{i_1\ldots i_m}; a_{i_1\ldots i_m}) - E\big(L(\theta^*_{i_1\ldots i_m}; a_{i_1\ldots i_m}) - L(\theta_{i_1\ldots i_m}; a_{i_1\ldots i_m})\big)}{2(\log 2 + c_0^m)} \in [-1, 1].$$

Finally, let $\tilde{\nu}_{n,m}(\Theta, \mathcal{A}) = \varphi(n,m)(2\log 2 + 2c_0^m)^{-1}\nu_{n,m}(\Theta, \mathcal{A})$ and $\tilde{M}_u = \varphi(n,m)(2\log 2 + 2c_0^m)^{-1}M_u$. By the concentration inequality in Theorem 1.1 of Klein et al. (2005), we obtain

$$\begin{aligned}
I_u &\le \exp\bigg(-\frac{(\tilde{M}_u/2)^2}{2\Big(2E\sup_{S_u}\tilde{\nu}_{n,m}(\Theta,\mathcal{A}) + \sup_{S_u} Var\big(\tilde{\nu}_{n,m}(\Theta,\mathcal{A})\big)\Big) + 3\tilde{M}_u/2}\bigg) \\
&< \exp\bigg(-\frac{\big(\varphi(n,M)M_u/(4\log 2 + 4c_0^m)\big)^2}{2\Big(\frac{\varphi(n,M)}{2(\log 2+c_0^m)}M_u + \frac{5\varphi(n,m)}{(\log 2+c_0^m)^2}M_u\Big) + 3\frac{\varphi(n,M)}{4(\log 2+c_0^m)}M_u}\bigg) \\
&= \exp\bigg(-\frac{2^u\varphi(n,M)\epsilon_n}{16\big(7c_0^m + 7\log 2 + 40\big)}\bigg) \le \exp\Big(-c_2 u\varphi(n,M)\epsilon_n\Big) := \xi^u,
\end{aligned}$$

where $c_2 = \frac{1}{16}(7c_0^m + 7\log 2 + 40)^{-1}$. It immediately follows that

$$I \le \sum_{u=1}^{+\infty}\xi^u = \frac{\xi}{1-\xi}.$$

Simple algebra immediately implies that $I \le (1+I)\xi \le 2\xi$, which completes the proof. □

Lemma 1. *Under the conditions of Proposition 1,* $E\big(\sup_{S_u}\nu_{n,m}(\Theta,\mathcal{A})\big) \le M_u/2$ *for any integer* $u \ge 1$ *when* n *is large enough.*

Proof of Lemma 1. Denote $f(\theta_{i_1\ldots i_m}; a_{i_1\ldots i_m}) = L(\theta^*_{i_1\ldots i_m}; a_{i_1\ldots i_m}) - L(\theta_{i_1\ldots i_m}; a_{i_1\ldots i_m})$, and hence $\nu_{n,m}(\Theta,\mathcal{A}) = \varphi^{-1}(n,m)\sum_{\delta^{n+1,ord}_{i_1\ldots i_m}=0}\big(f(\theta_{i_1\ldots i_m}; a_{i_1\ldots i_m}) - Ef(\theta_{i_1\ldots i_m}; a_{i_1\ldots i_m})\big)$. Let $\mathcal{A}' = (a'_{i_1\ldots i_m})$ be an independent copy of $\mathcal{A}$ and $\boldsymbol{\tau} = (\tau_{i_1\ldots i_m})_{\delta^{n+1,ord}_{i_1\ldots i_m}=0}$ be a collection of independent

Rademacher random variables. By the standard symmetrization argument, we have

$$
\begin{aligned}
E_{\mathcal{A}} \sup_{S_u} \nu_{n,m}(\Theta, \mathcal{A}) &= \frac{1}{\varphi(n,m)} E_{\mathcal{A}} \sup_{S_u} \sum_{\delta_{i_1 \ldots i_m}^{n+1,ord}=0} \big(f(\theta_{i_1 \ldots i_m}; a_{i_1 \ldots i_m}) - E_{\mathcal{A}'} f(\theta_{i_1 \ldots i_m}; a'_{i_1 \ldots i_m}) \big) \\
&\leq \frac{1}{\varphi(n,m)} E_{\mathcal{A},\mathcal{A}'} \sup_{S_u} \sum_{\delta_{i_1 \ldots i_m}^{n+1,ord}=0} \big(f(\theta_{i_1 \ldots i_m}; a_{i_1 \ldots i_m}) - f(\theta_{i_1 \ldots i_m}; a'_{i_1 \ldots i_m}) \big) \\
&= \frac{1}{\varphi(n,m)} E_{\mathcal{A},\mathcal{A}',\boldsymbol{\tau}} \sup_{S_u} \sum_{\delta_{i_1 \ldots i_m}^{n+1,ord}=0} \tau_{i_1 \ldots i_m} \big(f(\theta_{i_1 \ldots i_m}; a_{i_1 \ldots i_m}) - f(\theta_{i_1 \ldots i_m}; a'_{i_1 \ldots i_m}) \big) \\
&\leq \frac{2}{\varphi(n,m)} E_{\mathcal{A},\boldsymbol{\tau}} \sup_{S_u} \Big| \sum_{\delta_{i_1 \ldots i_m}^{n+1,ord}=0} \tau_{i_1 \ldots i_m} f(\theta_{i_1 \ldots i_m}; a_{i_1 \ldots i_m}) \Big|,
\end{aligned}
$$

where the second equality comes from the fact that $f(\theta_{i_1 \ldots i_m}; a_{i_1 \ldots i_m}) - f(\theta_{i_1 \ldots i_m}; a'_{i_1 \ldots i_m})$ and $\tau_{i_1 \ldots i_m} \big(f(\theta_{i_1 \ldots i_m}; a_{i_1 \ldots i_m}) - f(\theta_{i_1 \ldots i_m}; a'_{i_1 \ldots i_m}) \big)$ share the same distribution.

Denote $X(\Theta; \mathcal{A}) = \varphi^{-1/2}(n,m) \sum_{\delta_{i_1 \ldots i_m}^{n+1,ord}=0} \tau_{i_1 \ldots i_m} f(\theta_{i_1 \ldots i_m}; a_{i_1 \ldots i_m})$ as the Rademacher process. For any $\Theta^{(1)}, \Theta^{(2)} \in S_u$, and $\zeta \in \mathbb{R}$, we have $E_{\boldsymbol{\tau}|\mathcal{A}} \exp\Big(\zeta \big(X(\Theta^{(1)}; \mathcal{A}) - X(\Theta^{(2)}; \mathcal{A}) \big) \Big) \leq \exp\big(\frac{1}{2} \zeta^2 \rho^2(\Theta^{(1)}, \Theta^{(2)}; \mathcal{A}) \big)$, where

$$
\rho^2(\Theta^{(1)}, \Theta^{(2)}; \mathcal{A}) = \varphi^{-1}(n,m) \sum_{\delta_{i_1 \ldots i_m}^{n+1,ord}=0} \big(f(\theta_{i_1 \ldots i_m}^{(1)}; a_{i_1 \ldots i_m}) - f(\theta_{i_1 \ldots i_m}^{(2)}; a_{i_1 \ldots i_m}) \big)^2,
$$

showing that $X(\Theta; \mathcal{A})$ is a sub-Gaussian process with respect to the metric ρ when $\mathcal{A}$ is given. Thus, by Theorem 3.11 of Koltchinskii (2011), there exists a positive constant c_5, such that

$$
\varphi^{-1/2}(n,m) E_{\mathcal{A},\boldsymbol{\tau}} \sup_{S_u} \Big| \sum_{\delta_{i_1 \ldots i_m}^{n+1,ord}=1} \tau_{i_1 \ldots i_m} f(\theta_{i_1 \ldots i_m}; a_{i_1 \ldots i_m}) \Big| \leq \frac{c_5}{2} E_{\mathcal{A}} \int_0^{\mathrm{diam}(S_u)} H^{1/2}(\varepsilon; S_u, \rho) d\varepsilon,
$$

where $\mathrm{diam}(S_u)$ is the diameter of S_u and $H(\varepsilon; S_u, \rho)$ is the metric entropy. Note that $\big| \frac{dL(\theta_{i_1 \ldots i_m}; a_{i_1 \ldots i_m})}{d\theta_{i_1 \ldots i_m}} \big| = \big| \frac{\exp(-\theta_{i_1 \ldots i_m})}{1 - s_n + \exp(-\theta_{i_1 \ldots i_m})} (p_{i_1 \ldots i_m} - a_{i_1 \ldots i_m}) \big| < 1$. Thus, both $L(\theta_{i_1 \ldots i_m}; a_{i_1 \ldots i_m})$ and $f(\theta_{i_1 \ldots i_m}; a_{i_1 \ldots i_m})$ are Lipschitz continuous with Lipschitz constant 1. For any $\Theta^{(1)}, \Theta^{(2)} \in S_u$,

by triangle inequality, and Cauchy-Schwarz inequality, we have

$$
\begin{aligned}
&\rho^2(\Theta^{(1)}, \Theta^{(2)}; \mathcal{A}) \\
&\leq \varphi^{-1}(n,m) \sum_{\delta^{n+1,ord}_{i_1 \dots i_m}=0} |\theta^{(1)}_{i_1 \dots i_m} - \theta^{(2)}_{i_1 \dots i_m}|^2 \\
&= \varphi^{-1}(n,m) \sum_{\delta^{n+1,ord}_{i_1 \dots i_m}=0} |\mathcal{I} \times_1 (\boldsymbol{\alpha}^{(1)}_{i_1})^T \times_2 \dots \times_m (\boldsymbol{\alpha}^{(1)}_{i_m})^T - \mathcal{I} \times_1 (\boldsymbol{\alpha}^{(2)}_{i_1})^T \times_2 \dots \times_m (\boldsymbol{\alpha}^{(2)}_{i_m})^T|^2 \\
&\leq \varphi^{-1}(n,m) m \sum_{\delta^{n+1,ord}_{i_1 \dots i_m}=0} \sum_{j=1}^{m} ||\boldsymbol{\alpha}^{(1)}_{i_j} - \boldsymbol{\alpha}^{(2)}_{i_j}||_2^2 (\Pi_{l=1}^{j-1} ||\boldsymbol{\alpha}^{(1)}_{i_l}||_2^2)(\Pi_{l=j+1}^{m} ||\boldsymbol{\alpha}^{(2)}_{i_l}||_2^2) \\
&\leq \varphi^{-1}(n,m) m^2 n^{m-1} c_0^{2(m-1)} ||\boldsymbol{\alpha}^{(1)} - \boldsymbol{\alpha}^{(2)}||_F^2.
\end{aligned}
$$

As a result,

$$
\rho(\Theta^{(1)}, \Theta^{(2)}; \mathcal{A}) \leq \frac{m n^{m/2} c_0^m}{\varphi^{1/2}(n,m)} || \frac{1}{\sqrt{n} c_0} (\boldsymbol{\alpha}^{(1)} - \boldsymbol{\alpha}^{(2)}) ||_F,
$$

which directly leads to

$$
H(\varepsilon; S_u, \rho) \leq H(\frac{\varepsilon \varphi^{1/2}(n,m)}{m n^{m/2} c_0^m}; B(n \times r), || \cdot ||_F) \leq nr \log \frac{3 m n^{m/2} c_0^m}{\varepsilon \varphi^{1/2}(n,m)} \leq c_6 nr \log \frac{1}{\varepsilon},
$$

for some constant c_6 that depends on m, where $B(n \times r)$ is the unit ball in $\mathbb{R}^{n \times r}$ with respect to F-norm and the last inequality comes form the fact that $n^{m/2}$ and $\varphi^{1/2}(n,m)$ are of the same order. As such,

$$
E_{\mathcal{A}} \sup_{S_u} \nu_{n,m}(\Theta, \mathcal{A}) \leq \frac{c_5}{\varphi^{1/2}(n,m)} E_{\mathcal{A}} \int_0^{\mathrm{diam}(S_u)} \sqrt{c_6 nr \log \frac{1}{\varepsilon}} d\varepsilon.
$$

By concavity,

$$E_{\mathcal{A}} \int_0^{\mathrm{diam}(S_u)} \sqrt{c_6 nr \log \frac{1}{\varepsilon}} d\varepsilon \leq \int_0^{E_{\mathcal{A}}\mathrm{diam}(S_u)} \sqrt{c_6 nr \log \frac{1}{\varepsilon}} d\varepsilon$$
$$\leq \int_0^{\sqrt{E_{\mathcal{A}}\mathrm{diam}^2(S_u)}} \sqrt{c_6 nr \log \frac{1}{\varepsilon}} d\varepsilon.$$

According to the same argument of bounding the variance of $\nu_{n,m}(\Theta, \mathcal{A})$, we have

$$E_{\mathcal{A}}\rho^2(\Theta^{(1)}, \Theta^{(2)}; \mathcal{A}) \leq 2\big(E_{\mathcal{A}}\rho^2(\Theta^{(1)}, \Theta^*; \mathcal{A}) + E_{\mathcal{A}}\rho^2(\Theta^{(2)}, \Theta^*; \mathcal{A})\big)$$
$$\leq 2(5 \times 2^u \epsilon_n + 5 \times 2^u \epsilon_n) = 5 \times 2^{u+2}\epsilon_n,$$

implying that $\sqrt{E_{\mathcal{A}}\mathrm{diam}^2(S_u)} \leq (5 \times 2^{u+2}\epsilon_n)^{1/2}$. Putting the pieces together, we have

$$\begin{aligned}
E_{\mathcal{A}} \sup_{S_u} \nu_{n,m}(\Theta, \mathcal{A}) &\leq \frac{c_5}{\varphi^{1/2}(n,m)} \int_0^{(5\times 2^{u+2}\epsilon_n)^{1/2}} \sqrt{c_6 nr \log \frac{1}{\varepsilon}} d\varepsilon \\
&= \frac{c_5\sqrt{c_6 nr}}{\varphi^{1/2}(n,m)} \int_{(5\times 2^{u+2}\epsilon_n)^{-1/2}}^{+\infty} \frac{\sqrt{\log \varepsilon}}{\varepsilon^2} d\varepsilon \\
&\leq \frac{c_5\sqrt{c_6 nr}}{\varphi^{1/2}(n,m)\sqrt{\log\left(5 \times 2^{u+2}\epsilon_n\right)^{-1/2}}} \int_{(5\times 2^{u+2}\epsilon_n)^{-1/2}}^{+\infty} \frac{\log \varepsilon}{\varepsilon^2} d\varepsilon \\
&= \frac{c_5\sqrt{5 \times 2^{u+2} c_6 nr\epsilon_n}}{\varphi^{1/2}(n,m)} \frac{1 + \log\left(5 \times 2^{u+2}\epsilon_n\right)^{-1/2}}{\sqrt{\log\left(5 \times 2^{u+2}\epsilon_n\right)^{-1/2}}} \\
&\leq c_7 \sqrt{\frac{2^{u-2} nr\epsilon_n}{\varphi(n,m)} \log \frac{1}{\sqrt{\epsilon_n}}},
\end{aligned}$$

where c_7 is a positive constant. Since $M_u = 2^{u-2}\epsilon_n$, for $u = 1, 2, ...$, we have

$$E_{\mathcal{A}} \sup_{S_u} \nu_{n,m}(\Theta, \mathcal{A}) \leq c_7 \epsilon_n \sqrt{\frac{2^{u-2} nr}{\varphi(n,m)\epsilon_n} \log \frac{1}{\sqrt{\epsilon_n}}} \leq \frac{1}{2} M_u,$$

where the second inequality follows from the assumption that $\frac{nr}{\varphi(n,m)\epsilon_n} \log \frac{1}{\sqrt{\epsilon_n}} \leq c_1$ with c_1

taking to be $1/(4c_7^2)$. $\square$

Proof of Theorem 1. By the definition of $\widehat{\Theta}$, it follows from Proposition 1 that

$$\begin{aligned}
&P(D^2(\widehat{\Theta},\Theta^*)\geq\epsilon_n)\leq P\big(e_L(\widehat{\Theta},\Theta^*)\geq\epsilon_n\big)\\
&\leq P\big(\sup_{\{\Theta\in\Omega|e_L(\Theta,\Theta^*)\geq\epsilon_n\}}\mathcal{L}_\lambda(\Theta^*)-\mathcal{L}_\lambda(\Theta)\geq 0\big)\leq 2\exp\Big(-c_2\varphi(n,M)\epsilon_n\Big),
\end{aligned}$$

which implies that $D^2(\widehat{\Theta},\Theta^*)=O_p(\epsilon_n)$ as $\varphi(n,m)\epsilon_n$ diverges.

Next, we bound the F-norm of the different between $\widehat{\Theta}$ and Θ^*,

$$\begin{aligned}
\frac{1}{n^m}\|\widehat{\Theta}-\Theta^*\|_F^2&=\frac{1}{n^m}\sum_{\delta^{n+1}_{i_1\ldots i_m}=0}(\hat{\theta}_{i_1\ldots i_m}-\theta^*_{i_1\ldots i_m})^2+\frac{1}{n^m}\sum_{\delta^{n+1}_{i_1\ldots i_m}=1}(\hat{\theta}_{i_1\ldots i_m}-\theta^*_{i_1\ldots i_m})^2\\
&\leq\frac{m!}{n^m}\sum_{\delta^{n+1,ord}_{i_1\ldots i_m}=0}(\hat{\theta}_{i_1\ldots i_m}-\theta^*_{i_1\ldots i_m})^2+\frac{1}{n^m}\sum_{\delta^{n+1}_{i_1\ldots i_m}=1}(\hat{\theta}_{i_1\ldots i_m}-\theta^*_{i_1\ldots i_m})^2.
\end{aligned}$$

As the first term usually dominates the second term on diagonal elements, it suffices to bound the first term.

Let $g(x)=\log\frac{x^2}{s_n-x^2}$. By Lagrange's mean value theorem, there exists a constant $c_8=\max\{\frac{2(1+\exp(c_0^m))^{3/2}}{\exp(c_0^m)},\frac{2(1+\exp(-c_0^m))^{3/2}}{\exp(-c_0^m)}\}$, such that

$$|\theta_{i_1\ldots i_m}-\theta^*_{i_1\ldots i_m}|=|g(p^{1/2}_{i_1\ldots i_m})-g\big((p^*_{i_1\ldots i_m})^{1/2}\big)|\leq\frac{c_8}{\sqrt{s_n}}|p^{1/2}_{i_1\ldots i_m}-(p^*_{i_1\ldots i_m})^{1/2}|,$$

where the inequality follows from the assumption that $\max_{i_1,\ldots,i_m}\theta_{i_1\ldots i_m}\leq c_0^m$. It further implies that

$$\varphi^{-1}(n,m)\sum_{\delta^{n+1,ord}_{i_1\ldots i_m}=0}(\theta_{i_1\ldots i_m}-\theta^*_{i_1\ldots i_m})^2\leq\frac{c_8^2}{s_n}D^2(\Theta,\Theta^*).$$

Therefore,

$$\frac{m!}{n^m}\sum_{\delta^{n+1,ord}_{i_1\ldots i_m}=0}(\hat{\theta}_{i_1\ldots i_m}-\theta^*_{i_1\ldots i_m})^2\leq\frac{c_8^2m!\varphi(n,m)}{n^m s_n}D^2(\widehat{\Theta},\Theta^*).$$

Since $D^2(\widehat{\Theta}, \Theta^*) = O_p(\epsilon_n)$, it immediately follows that $\frac{1}{n^{m/2}}\|\widehat{\Theta} - \Theta^*\|_F = O_p(\sqrt{\epsilon_n/s_n})$. $\square$

Lemma 2. *Under the conditions of Theorem 1, if* $\lim_{n\to+\infty} \lambda_n \epsilon_n s_n^{-2}(\log s_n^{-1})^{-1} > 0$, *then with probability at least* $1 - n^{-2}$, *there exists a positive constant* c_9 *such that*

$$\frac{1}{n^{m/2}}||\mathcal{I} \times_1 \hat{Z}\hat{C} \times_2 ... \times_m \hat{Z}\hat{C} - \mathcal{I} \times_1 Z^*C^* \times_2 ... \times_m Z^*C^*||_F \leq c_9\sqrt{\epsilon_n/s_n}.$$

Proof of Lemma 2. We first prove a claim that $J(\hat{\boldsymbol{\alpha}}) = O_p(\epsilon_n/s_n)$. By the definition of $\hat{\boldsymbol{\alpha}}$, we have $\lambda_n J(\hat{\boldsymbol{\alpha}}) < \mathcal{L}_\lambda(\hat{\boldsymbol{\alpha}}; \mathcal{A}) \leq \mathcal{L}_\lambda(\boldsymbol{\alpha}^*; \mathcal{A}) \leq \mathcal{L}(\boldsymbol{\alpha}^*, \mathcal{A}) + \frac{1}{2}\epsilon_n$. Note that

$$\varphi(n,m)\mathcal{L}(\boldsymbol{\alpha}^*, \mathcal{A}) = \sum_{\delta_{i_1\ldots i_m}^{n+1,ord}=0} \big(a_{i_1\ldots i_m} \log \frac{1 - p^*_{i_1\ldots i_m}}{p^*_{i_1..i_m}} + \log \frac{1}{1 - p^*_{i_1\ldots i_m}}\big) := \sum_{\delta_{i_1\ldots i_m}^{n+1,ord}=0} X_{i_1\ldots i_m},$$

which is the sum of $\varphi(n,m)$ independent two-value random variables with $EX_{i_1\ldots i_m}$ of the order $s_n \log \frac{1}{s_n}$ and $EX^2_{i_1\ldots i_m}$ of the order $s_n(\log \frac{1}{s_n})^2$. Since $|X_{i_1\ldots i_m}| = O(\log \frac{1}{s_n})$, it follows from Bernstein inequality that there exists some constant c_{10}, such that

$$P\big(\frac{1}{\varphi(n,m)} \sum_{\delta_{i_1\ldots i_m}^{n+1,ord}=0} (X_{i_1\ldots i_m} - EX_{i_1\ldots i_m}) > t\big) \leq \exp\Big\{-\frac{c_{10}\varphi^2(n,m)t^2}{\varphi(n,m)s_n(\log \frac{1}{s_n})^2 + \varphi(n,m)t \log \frac{1}{s_n}}\Big\}.$$

Taking $t = \frac{2}{\sqrt{c_{10}}}\varphi^{-1/2}(n,m)s_n^{1/2}(\log \frac{1}{s_n})(\log n)^{1/2}$, with probability at least $1 - n^{-2}$ it holds true that

$$\mathcal{L}(\boldsymbol{\alpha}^*, \mathcal{A}) = \frac{1}{\varphi(n,m)} \sum_{\delta_{i_1\ldots i_m}^{n+1,ord}=0} X_{i_1\ldots i_m} \leq c_{11} s_n \log \frac{1}{s_n},$$

for some positive constant c_{11}. The desired claim follows immediately after the assumption that $\lim_{n\to+\infty} \lambda_n \epsilon_n s_n^{-2}(\log s_n^{-1})^{-1} > 0$ and the fact that $\epsilon_n = o(s_n \log \frac{1}{s_n})$.

We now turn to quantify the difference between $\mathcal{I} \times_1 (\hat{Z}\hat{C}) \times_2 ... \times_m (\hat{Z}\hat{C})$ and $\mathcal{I} \times_1$

$Z^*C^* \times_2 ... \times_m Z^*C^*$. Applying the triangle inequality yields that

$$\begin{aligned}&\frac{1}{n^{m/2}}\left\|\mathcal{I}\times_1\hat{\boldsymbol{\alpha}}\times_2...\times_m\hat{\boldsymbol{\alpha}}-\mathcal{I}\times_1(\hat{Z}\hat{C})\times_2...\times_m(\hat{Z}\hat{C})\right\|_F\\ \leq&\frac{1}{n^{m/2}}\sum_{j=0}^{m-1}\left\|\hat{\boldsymbol{\alpha}}-\hat{Z}\hat{C}\right\|_F\left\|\hat{\boldsymbol{\alpha}}\right\|_F^j\left\|\hat{Z}\hat{C}\right\|_F^{m-1-j}=mc_0^{m-1}\sqrt{J(\hat{\boldsymbol{\alpha}})}=O_p(\sqrt{\epsilon_n/s_n}),\end{aligned}\tag{1}$$

where the last equality follows form the claim above. Similarly,

$$\frac{1}{n^{m/2}}\left\|\mathcal{I}\times_1\boldsymbol{\alpha^*}\times_2...\times_m\boldsymbol{\alpha^*}-\mathcal{I}\times_1(Z^*C^*)\times_2...\times_m(Z^*C^*)\right\|_F\leq mc_0^{m-1}\sqrt{J(\boldsymbol{\alpha}^*)}=O(\sqrt{\epsilon_n}),\tag{2}$$

Finally, combing (1), (2) and Theorem 1 immediately leads to the desired probability bound.

□

Lemma 3. *Let $\widehat{\mathcal{B}}=\mathcal{I}\times_1\hat{C}\times_2...\times_m\hat{C}$ be the estimation counterpart of $\mathcal{B}^*$. Denote $\mathcal{M}^*=\mathcal{B}^*\times_2 Z^*\times_3...\times_m Z^*$ and $\widehat{\mathcal{M}}=\widehat{B}\times_2\hat{Z}\times_3...\times_m\hat{Z}$. Under the conditions of Theorem 1 as well as Assumptions A and B, if $\lim_{n\to+\infty}\lambda_n\epsilon_n s_n^{-2}(\log s_n^{-1})^{-1}>0$ and $K=o(\gamma_n^2 s_n/\epsilon_n)$, then with probability at least $1-n^{-2}$, the following event F holds. F : for any $k\in[K]$, there exists an unique $k'\in[K]$, such that*

$$\frac{1}{n^{(m-1)/2}}\left\|\widehat{\mathcal{M}}_{k'}-\mathcal{M}_k^*\right\|_F=o(\gamma_n).$$

Proof of Lemma 3. We first prove existence. Denote F_0 be the event that there exists $k\in[K]$ and a constant c_{12} such that $\frac{1}{n^{(m-1)/2}}\left\|\widehat{\mathcal{M}}_{k'}-\mathcal{M}_k^*\right\|_F\geq c_{12}\gamma_n$, for any $k'\in[K]$. It follows that

$$\frac{1}{n^{m/2}}\left\|\widehat{\mathcal{B}}\times_1\hat{Z}\times_2...\times_m\hat{Z}-\mathcal{B}^*\times_1 Z^*\times_2...\times_m Z^*\right\|_F\geq\frac{1}{n^{m/2}}(\sum_{i\in N_k^*}\|(\widehat{\mathcal{M}}\times_1\hat{Z})_i-\mathcal{M}_k^*\|_F^2)^{1/2}\geq c_{12}\gamma_n\sqrt{\frac{n_k}{n}}.$$

Moreover, by Assumption B and the condition $K = o(\gamma_n^2 s_n/\epsilon_n)$, we have

$$\gamma_n\sqrt{\frac{n_k}{n}} \geq \gamma_n\sqrt{\frac{\min_{k\in[K]} n_k}{n}} \geq \gamma_n\sqrt{\frac{\max_{k\in[K]} n_k}{c_4 n}} \geq \gamma_n\sqrt{\frac{1}{c_4 K}} \gg \sqrt{\frac{\epsilon_n}{s_n}}.$$

Hence, it follows form Lemma 2 that

$$P(F_0) \leq P(\frac{1}{n^{m/2}}\left\|\widehat{\mathcal{B}} \times_1 \hat{Z} \times_2 ... \times_m \hat{Z} - \mathcal{B}^* \times_1 Z^* \times_2 ... \times_m Z^*\right\|_F \geq c_{12}\gamma_n\sqrt{\frac{n_k}{n}}) \leq n^{-2}.$$

Therefore, with probability at least $1 - n^{-2}$, F_0^C, the complement of F_0 holds; that is, the existence holds with high probability.

We now prove uniqueness condition on F_0^C. Assume there exist $k_1 \neq k_2 \in [K]$ such that $\frac{1}{n^{(m-1)/2}}\left\|\widehat{\mathcal{M}}_{k_i} - \mathcal{M}_k^*\right\|_F = o(\gamma_n)$ for $i \in [2]$. By existence, there exists $a \in [K]$ and $b_1 \neq b_2 \in [K]$ such that $\frac{1}{n^{(m-1)/2}}\left\|\widehat{\mathcal{M}}_a - \mathcal{M}_{b_j}^*\right\|_F = o(\gamma_n)$ for $j \in [2]$. On one hand, the triangle inequality yields that

$$\frac{1}{n^{(m-1)/2}}\|\mathcal{M}_{b_1}^* - \mathcal{M}_{b_2}^*\|_F \leq \frac{1}{n^{(m-1)/2}}\left\|\widehat{\mathcal{M}}_a - \mathcal{M}_{b_1}^*\right\|_F + \frac{1}{n^{(m-1)/2}}\left\|\widehat{\mathcal{M}}_a - \mathcal{M}_{b_2}^*\right\|_F = o(\gamma_n).$$

On the other hand, Assumption A and B imply that

$$\frac{1}{n^{(m-1)/2}}\|\mathcal{M}_{b_1}^* - \mathcal{M}_{b_2}^*\|_F \geq \sqrt{\frac{\min_{k\in[K]} n_k^{m-1}}{n^{m-1}}}\left\|\mathcal{B}_{b_1}^* - \mathcal{B}_{b_2}^*\right\|_F \geq c_3 c_4^{(1-m)/2}\gamma_n,$$

which is a contradiction. Hence, F_0^C also implies uniqueness, showing that F holds with probability at least $1 - n^{-2}$. □

Proof of Theorem 2. Based on Lemma 3, with probability at least $1 - n^{-2}$, there exists a permutation $\pi^* \in S_K$ such that for each $k \in [K]$, $k' = \pi^*(k)$ satisfies $n^{(1-m)/2}\|\widehat{\mathcal{M}}_{k'} - \mathcal{M}_k^*\|_F = o(\gamma_n)$. We want to show $\lim_{n\to+\infty}\min_{\pi\in S_K}\frac{1}{n}\sum_{i=1}^{n}\mathbf{1}\{\psi_i^* \neq \pi(\hat{\psi}_i)\} = O_p(\epsilon_n s_n^{-1}\gamma_n^{-2})$. Let

$\hat{N}_k = \{i : \hat{\psi}_i = k\}$, for $k \in [K]$. Note that

$$\min_{\pi \in S_K} \sum_{i=1}^{n} \mathbf{1}\{\psi_i^* \neq \pi(\hat{\psi}_i)\} = \min_{\pi \in S_K} |\bigcup_{k=1}^{K} \{i : \psi_i^* = k, \hat{\psi}_i \neq \pi^{-1}(k)\}| = \min_{\pi \in S_K} \sum_{k=1}^{K} |N_k^* \setminus \hat{N}_{\pi^{-1}(k)}|$$
$$= \min_{\pi \in S_K} \sum_{k=1}^{K} |N_k^* \setminus \hat{N}_{\pi(k)}|,$$

where the last equality follows from the fact that π^{-1} is also a permutation in S_K. It then suffices to show that for the particular permutation π^*,

$$\sum_{k=1}^{K} \frac{|N_k^* \setminus \hat{N}_{\pi^*(k)}|}{n} = O_p(\epsilon_n s_n^{-1} \gamma_n^{-2}).$$

In fact, by Lemma 2, we have

$$P\big(\sum_{k=1}^{K} \frac{|N_k^* \setminus \hat{N}_{\pi^*(k)}|}{n} = O(\epsilon_n s_n^{-1} \gamma_n^{-2})\big) = P\big(\sum_{k=1}^{K} \frac{c_3^2 \gamma_n^2 |N_k^* \setminus \hat{N}_{\pi^*(k)}|}{4c_4^{m-1} n} = O(\epsilon_n s_n^{-1})\big)$$
$$\geq P\big(\sum_{k=1}^{K} \frac{c_3^2 \gamma_n^2 |N_k^* \setminus \hat{N}_{\pi^*(k)}|}{4c_4^{m-1} n} \leq \frac{1}{n^m} \big\|\widehat{\mathcal{B}} \times_1 \hat{Z} \times_2 ... \times_m \hat{Z} - \mathcal{B}^* \times_1 Z^* \times_2 ... \times_m Z^*\big\|_F^2\big) +$$
$$P\big(\frac{1}{n^m} \big\|\widehat{\mathcal{B}} \times_1 \hat{Z} \times_2 ... \times_m \hat{Z} - \mathcal{B}^* \times_1 Z^* \times_2 ... \times_m Z^*\big\|_F^2 = O(\epsilon_n s_n^{-1})\big) - 1$$
$$\geq P\big(\sum_{k=1}^{K} \frac{c_3^2 \gamma_n^2 |N_k^* \setminus \hat{N}_{\pi^*(k)}|}{4c_4^{m-1} n} \leq \frac{1}{n^m} \big\|\widehat{\mathcal{B}} \times_1 \hat{Z} \times_2 ... \times_m \hat{Z} - \mathcal{B}^* \times_1 Z^* \times_2 ... \times_m Z^*\big\|_F^2\big) - n^{-2}.$$

For any tensors $\mathcal{T}_1, \mathcal{T}_2, \mathcal{T}_3$ sharing the same shape, it is easy to verify the following by triangle

inequality, $\|\mathcal{T}_1 - \mathcal{T}_2\|_F^2 \geq \frac{1}{2}\|\mathcal{T}_3 - \mathcal{T}_2\|_F^2 - \|\mathcal{T}_1 - \mathcal{T}_3\|_F^2$. Consequently,

$$\frac{1}{n^m}\big\|\widehat{\mathcal{B}} \times_1 \hat{Z} \times_2 \ldots \times_m \hat{Z} - \mathcal{B}^* \times_1 Z^* \times_2 \ldots \times_m Z^*\big\|_F^2 \geq \frac{1}{n^m}\sum_{k=1}^{K}\sum_{i\in N_k^*\setminus \hat{N}_{\pi^*(k)}} \big\|(\widehat{\mathcal{M}} \times_1 \hat{Z})_i - \mathcal{M}_k^*\big\|_F^2$$

$$\geq \frac{1}{n^m}\sum_{k=1}^{K}\sum_{i\in N_k^*\setminus \hat{N}_{\pi^*(k)}} \Big(\frac{1}{2}\big\|\mathcal{M}^*_{(\pi^*)^{-1}(\hat{\psi}_i)} - \mathcal{M}_k^*\big\|_F^2 - \|\widehat{\mathcal{M}}_{\hat{\psi}_i} - \mathcal{M}^*_{(\pi^*)^{-1}(\hat{\psi}_i)}\|_F^2\Big)$$

$$\geq \sum_{k=1}^{K}\frac{|N_k^* \setminus \hat{N}_{\pi^*(k)}|}{n^m}\Big(\min_{i\in N_k^*\setminus \hat{N}_{\pi^*(k)}} \frac{1}{2}\big\|\mathcal{M}^*_{(\pi^*)^{-1}(\hat{\psi}_i)} - \mathcal{M}_k^*\big\|_F^2 - \max_{i\in N_k^*\setminus \hat{N}_{\pi^*(k)}} \big\|\widehat{\mathcal{M}}_{\hat{\psi}_i} - \mathcal{M}^*_{(\pi^*)^{-1}(\hat{\psi}_i)}\big\|_F^2\Big)$$

$$\geq \sum_{k=1}^{K}\frac{|N_k^* \setminus \hat{N}_{\pi^*(k)}|}{n^m}\Big(\frac{c_3^2\gamma_n^2 n^{m-1}}{2c_4^{m-1}} - \max_{i\in N_k^*\setminus \hat{N}_{\pi^*(k)}} \big\|\widehat{\mathcal{M}}_{\hat{\psi}_i} - \mathcal{M}^*_{(\pi^*)^{-1}(\hat{\psi}_i)}\big\|_F^2\Big),$$

where the last inequality follows from Assumption A and B. Combining these with Lemma 3, we have

$$P\Big(\sum_{k=1}^{K}\frac{|N_k^* \setminus \hat{N}_{\pi^*(k)}|}{n} = O(\epsilon_n s_n^{-1}\gamma_n^{-2})\Big) + n^{-2}$$

$$\geq P\Big(\sum_{k=1}^{K}\frac{c_3^2\gamma_n^2|N_k^* \setminus \hat{N}_{\pi^*(k)}|}{4c_4^{m-1}n} \leq \sum_{k=1}^{K}\frac{|N_k^* \setminus \hat{N}_{\pi^*(k)}|}{n^m}\Big(\frac{c_3^2\gamma_n^2 n^{m-1}}{2c_4^{n-1}} - \max_{i\in N_k^*\setminus \hat{N}_{\pi^*(k)}} \big\|\widehat{\mathcal{M}}_{\hat{\psi}_i} - \mathcal{M}^*_{(\pi^*)^{-1}(\hat{\psi}_i)}\big\|_F^2\Big)\Big)$$

$$\geq P\Big(\bigcap_{k=1}^{K}\Big\{\frac{c_3^2\gamma_n^2}{4c_4^{m-1}} \leq \frac{c_3^2\gamma_n^2}{2c_4^{m-1}} - \max_{i\in N_k^*\setminus \hat{N}_{\pi^*(k)}} \frac{1}{n^{m-1}}\big\|\widehat{\mathcal{M}}_{\hat{\psi}_i} - \mathcal{M}^*_{(\pi^*)^{-1}(\hat{\psi}_i)}\big\|_F^2\Big\}\Big)$$

$$\geq P\Big(\bigcap_{k=1}^{K}\Big\{\max_{i\in N_k^*\setminus \hat{N}_{\pi^*(k)}} \frac{1}{n^{m-1}}\big\|\widehat{\mathcal{M}}_{\hat{\psi}_i} - \mathcal{M}^*_{(\pi^*)^{-1}(\hat{\psi}_i)}\big\|_F^2 = o(\gamma_n^2)\Big\}\Big).$$

Finally, the definition of π^* implies that

$$P\Big(\sum_{k=1}^{K}\frac{|N_k^* \setminus \hat{N}_{\pi^*(k)}|}{n} = O(\epsilon_n s_n^{-1}\gamma_n^{-2})\Big) \geq 1 - n^{-2} - n^{-2} = 1 - 2n^{-2},$$

and thus the desired consistency result follows immediately. □

References

Klein, T., Rio, E., et al. (2005). Concentration around the mean for maxima of empirical processes. *The Annals of Probability*, 33(3):1060–1077.

Koltchinskii, V. (2011). *Oracle Inequalities in Empirical Risk Minimization and Sparse Recovery Problems: Ecole d'Eté de Probabilités de Saint-Flour XXXVIII-2008*, volume 2033. Springer Science & Business Media.