Robust Covariance Assisted Tensor Response Regression

Ning Wang and Xin Zhang

This document contains the supplementary materials for the paper "Robust Covariance Assisted Tensor Response Regression". It provides additional technical lemmas and proofs for Theorem 1 of the paper.

1 Additional Lemmas

Let $\mathbf{E}_i \sim t(0, \boldsymbol{\Sigma}, \nu)$. By definition, $\mathbf{E}_i$ can be written as $1/\sqrt{G_i}\mathbf{Q}_i$, where $\mathbf{Q}_i \sim N(0, \boldsymbol{\Sigma})$.

Lemma 1 (Wang et al. (2022))**.** *Let $\gamma_i = 1/\sqrt{G_i}$ and $t_i = \gamma_i^2 \mathrm{tr}(\boldsymbol{\Sigma})/p$. Then we have*

$$\mathbb{P}(|\frac{1}{p}\mathrm{vec}(\mathbf{E}_i)^T\mathrm{vec}(\mathbf{E}_i) - t_i| \geq \epsilon_i/2|\gamma_i) \leq 2\exp\{-cp\min(\frac{c_1\epsilon_i^2}{\gamma_i^4}, \frac{c_2\epsilon_i}{\gamma_i^2})\}. \tag{1}$$

Let $\widetilde{t}_i = \|\mathbf{Y}_i - \widehat{\mathbf{B}}^{\mathrm{OLS}}\bar{\times}_{M+1}\mathbf{X}_i\|_F/p$, which is the inverse of $\widehat{\omega}_i$.

Lemma 2. *We have $\max_i|\widetilde{t}_iG_i - \mathrm{tr}(\boldsymbol{\Sigma})/p| = O(\sqrt{\log(n)/p})$ with probability at least $1-4n^{-C_1}$.*

Proof. Let $\mathbf{G} = \mathbf{B}_{(M+1)}^T$, and $\widehat{\mathbf{G}}$ and $\widehat{\mathbf{B}}^{\mathrm{OLS}}$ be defined similarly, and $t_i = \gamma_i^2\mathrm{tr}(\boldsymbol{\Sigma})/p$. We first decompose $\widetilde{t}_i$ as follows.

$$\begin{aligned}
\widetilde{t}_i &= \frac{1}{p}\mathrm{vec}(\mathbf{E}_i)^T\mathrm{vec}(\mathbf{E}_i) + \frac{2}{p}\mathbf{E}_i^T(\widehat{\mathbf{G}}^{\mathrm{OLS}} - \mathbf{G})\mathbf{X}_i + \frac{1}{p}((\widehat{\mathbf{G}}^{\mathrm{OLS}} - \mathbf{G})\mathbf{X}_i)^T(\widehat{\mathbf{G}}^{\mathrm{OLS}} - \mathbf{G})\mathbf{X}_i \\
&\leq \frac{1}{p}\mathrm{vec}(\mathbf{E}_i)^T\mathrm{vec}(\mathbf{E}_i) + \frac{2}{p}\sqrt{\mathrm{vec}(\mathbf{E}_i)^T\mathrm{vec}(\mathbf{E}_i)}\sqrt{((\widehat{\mathbf{G}}^{\mathrm{OLS}} - \mathbf{G})\mathbf{X}_i)^T(\widehat{\mathbf{G}}^{\mathrm{OLS}} - \mathbf{G})\mathbf{X}_i} \\
&+ \frac{1}{p}((\widehat{\mathbf{G}}^{\mathrm{OLS}} - \mathbf{G})\mathbf{X}_i)^T(\widehat{\mathbf{G}}^{\mathrm{OLS}} - \mathbf{G})\mathbf{X}_i.
\end{aligned}$$

It follows that

$$
\begin{aligned}
\mathbb{P}(|\widetilde{t}_i - t_i| \geq 2\epsilon) \leq\ & \mathbb{P}(|\frac{1}{p}\mathrm{vec}(\mathbf{E}_i)^T\mathrm{vec}(\mathbf{E}_i) - t_i| \geq \epsilon/2) \\
& + \mathbb{P}(\frac{2}{p}\sqrt{\mathrm{vec}(\mathbf{E}_i)^T\mathrm{vec}(\mathbf{E}_i)}\sqrt{((\widehat{\mathbf{G}}^{\mathrm{OLS}} - \mathbf{G})\mathbf{X}_i)^T(\widehat{\mathbf{G}}^{\mathrm{OLS}} - \mathbf{G})\mathbf{X}_i} \geq \epsilon) \\
& + \mathbb{P}(\frac{1}{p}((\widehat{\mathbf{G}}^{\mathrm{OLS}} - \mathbf{G})\mathbf{X}_i)^T(\widehat{\mathbf{G}}^{\mathrm{OLS}} - \mathbf{G})\mathbf{X}_i \geq \epsilon/2) \\
\leq\ & 2\mathbb{P}(|\frac{1}{p}\mathrm{vec}(\mathbf{E}_i)^T\mathrm{vec}(\mathbf{E}_i) - t_i| \geq \epsilon/2) + 2\mathbb{P}(\frac{1}{p}((\widehat{\mathbf{G}}^{\mathrm{OLS}} - \mathbf{G})\mathbf{X}_i)^T(\widehat{\mathbf{G}}^{\mathrm{OLS}} - \mathbf{G})\mathbf{X}_i \geq \epsilon/2).
\end{aligned}
$$

Let $\epsilon_i^* = A\sqrt{\log(n)/p}t_i$, where $A > 1/(\sqrt{cc_1}\mathrm{tr}(\boldsymbol{\Sigma})/p)$. By Lemma 1, we have

$$
n\mathbb{P}(|\frac{1}{p}\mathbf{E}_i^T\mathbf{E}_i - t_i| \geq \epsilon_i^*/2 \mid G_1, \cdots, G_n) \leq 2n^{-C_1}
$$

for a positive constant C_1. By taking expectation with respect to $G_1, \cdots, G_n$ on both sides of last inequality, we have

$$
n\mathbb{P}(|\frac{1}{p}\mathbf{E}_i^T\mathbf{E}_i - t_i| \geq \epsilon_i/2) \leq 2n^{-C_1}.
$$

Then note that

$$
\sqrt{n}(\widehat{\mathbf{G}}^{\mathrm{OLS}} - \mathbf{G}) \mid (G_1, \cdots, G_n) \sim \mathrm{TN}(0, \boldsymbol{\Sigma}, \boldsymbol{\Sigma}_{\mathbf{X},G}^{-1}).
$$

It follows that

$$
(\widehat{\mathbf{G}}^{\mathrm{OLS}} - \mathbf{G})\mathbf{X}_i \mid (G_1, \cdots, G_n) \sim N(0, \mathbf{X}_i^T\boldsymbol{\Sigma}_{\mathbf{X},G}^{-1}\mathbf{X}_i/n \cdot \boldsymbol{\Sigma})
$$

Then by Lemma 1, we have

$$
\mathbb{P}(|\frac{1}{p}((\widehat{\mathbf{G}}^{\mathrm{OLS}} - \mathbf{G})\mathbf{X}_i)^T(\widehat{\mathbf{G}}^{\mathrm{OLS}} - \mathbf{G})\mathbf{X}_i - \mathbf{X}_i\boldsymbol{\Sigma}_{\mathbf{X}}^{-1}\mathbf{X}_i t_i/n| \geq \epsilon_i/2 \mid G_1, \cdots, G_n) \leq 2\exp\{-\tilde{c}p\min(\frac{\tilde{c}_1\epsilon_i^2}{\gamma_i^4}, \frac{\tilde{c}_2\epsilon_i}{\gamma_i^2})\}
$$

Let $\epsilon_i^* = A\sqrt{\log(n)/p}t_i$, where $A > 1/(\sqrt{cc_1}\mathrm{tr}(\boldsymbol{\Sigma})/p)$, we have

$$
n\mathbb{P}(\frac{1}{p}((\widehat{\mathbf{G}}^{\mathrm{OLS}} - \mathbf{G})\mathbf{X}_i)^T(\widehat{\mathbf{G}}^{\mathrm{OLS}} - \mathbf{G})\mathbf{X}_i \geq \epsilon_i^*/2 + C_x M_x q t_i/n) \leq 2n^{-C_1}.
$$

Then let $\epsilon = \epsilon_i^* + 2C_x M_x q t_i/n$, $\mathcal{D}_i = \{|\widetilde{t}_i - t_i| \leq \epsilon_i/2\}$ and $\mathcal{D} = \bigcap_i \mathcal{D}_i$. By Union bound, we have

$$
\mathbb{P}(\mathcal{D}^c) \leq 2n\mathbb{P}(|\frac{1}{p}\mathbf{E}_i^T\mathbf{E}_i - t_i| \geq \epsilon_i/2) + 2n\mathbb{P}(\frac{1}{p}((\widehat{\mathbf{G}}^{\mathrm{OLS}} - \mathbf{G})\mathbf{X}_i)^T(\widehat{\mathbf{G}}^{\mathrm{OLS}} - \mathbf{G})\mathbf{X}_i \geq \epsilon_i/2) \leq 4n^{-C_1}
$$

It follows that $\max_i |\widetilde{t}_i G_i - \mathrm{tr}(\boldsymbol{\Sigma})/p| = O(\sqrt{\log(n)/p} + 1/n)$, with probability at least $1 - 4n^{-C_1}$ for a constant $C_1 > 0$. □

A direct conclusion follows from Lemma 2 is $|\widehat{\omega}_i/G_i - p/\mathrm{tr}(\mathbf{\Sigma})| = O(\sqrt{\log(n)/p} + 1/n)$ with probability at least $1 - 4n^{-C_1}$.

Lemma 3. *Let $\mathbf{Z}_i \in \mathbb{R}^{p_1 \times \cdots \times p_M}$, $i = 1, \cdots, n$, are i.i.d random varibales from $\mathrm{TN}(0, \mathbf{I}_{p_1}, \cdots, \mathbf{I}_{p_M})$, $\mathbf{A} \in \mathbb{R}^{p_{-m} \times p_{-m}}$ be a positive and symmetric definite matrix whose eigenvalues are bounded by a constant c, and α_i, $i = 1, \cdots, n$, are positive constants. Then for any fixed vector $\mathbf{x} \in \mathcal{S}^{p_m - 1} = \{\mathbf{x} : \mathbf{x} \in \mathbb{R}^{p_m}, \|\mathbf{x}\| = 1\}$and any $\epsilon \geq 0$, we have*

$$
\begin{aligned}
&\mathbb{P}(|\frac{1}{np_{-m}} \sum_{i=1}^{n} \alpha_i \mathbf{x}^T (\mathbf{Z}_i)_{(m)} \mathbf{A} (\mathbf{Z}_i)_{(m)}^T \mathbf{x} - \frac{\mathrm{tr}(\mathbf{A}) \sum_{i=1}^{n} \alpha_i}{np_{-m}}| \geq \epsilon) \\
&\leq 2 \exp\{-Cnp_{-m} \min(\frac{\epsilon^2}{16c^2 \sum_{i=1}^{n} \alpha_i^2/n}, \frac{\epsilon}{4c \max_i \alpha_i})\}.
\end{aligned}
$$

Proof. Let $\mathbf{P}\mathbf{\Delta}\mathbf{P}^T$ be the eigenvalue decomposition of $\mathbf{A}$, where $\mathbf{\Delta}$ is a $p_{-m} \times p_{-m}$ diagonal matrix whose i-th diagonal elements is δ_i. Let $(\mathbf{W}_i)_{(m)} = (\mathbf{Z}_i)_{(m)} \mathbf{P}$. We know that $\mathrm{vec}((\mathbf{W}_i)_{(m)}) \sim N(0, \mathbf{I}_p)$. Then we have

$$
\begin{aligned}
\frac{1}{np_{-m}} \sum_{i=1}^{n} \alpha_i \mathbf{x}^T (\mathbf{Z}_i)_{(m)} \mathbf{A} (\mathbf{Z}_i)_{(m)}^T \mathbf{x} &= \frac{1}{np_{-m}} \sum_{i=1}^{n} \alpha_i \mathbf{x}^T (\mathbf{W}_i)_{(m)} \mathbf{\Delta} (\mathbf{W}_i)_{(m)}^T \mathbf{x} \\
&= \frac{1}{np_{-m}} \sum_{i=1}^{n} \sum_{l=1}^{p_{-m}} \alpha_i \delta_l \mathbf{x}^T (\mathbf{W}_i)_{(m),l} (\mathbf{W}_i)_{(m),l}^T \\
&= \frac{1}{np_{-m}} \sum_{i=1}^{n} \sum_{l=1}^{p_{-m}} \alpha_i \delta_l \mathbf{x}^T (\mathbf{W}_i)_{(m),l} (\mathbf{W}_i)_{(m),l}^T \mathbf{x},
\end{aligned}
$$

where $\mathbf{W}_{(m),l}$ is the l-th column of $\mathbf{W}_{(m)}$. Note that $\mathbf{x}^T (\mathbf{W}_i)_{(m),l} \sim N(0, 1)$ and $\mathbf{x}^T (\mathbf{W}_i)_{(m),l}$ are independent for $l = 1, \cdots, p_{-m}$ and $i = 1, \cdots, n$.

By Bernstein's inequality, we have

$$
\begin{aligned}
&\mathbb{P}(|\frac{1}{np_{-m}} \sum_{i=1}^{n} \sum_{l=1}^{p_{-m}} \alpha_i \delta_l \mathbf{x}^T (\mathbf{W}_i)_{(m),l} (\mathbf{W}_i)_{(m),l}^T \mathbf{x} - \frac{1}{p_{-m}} \mathrm{tr}(\mathbf{A})| \leq \epsilon) \\
&\leq 2 \exp\{-Cp_{-m} \min(\frac{\epsilon^2}{16 \sum_{i=1}^{n} \sum_{l=1}^{p_{-m}} \alpha_i^2 \delta_l^2/(np_{-m})}, \frac{\epsilon}{4 \max_l \delta_l \max_i \alpha_i})\}.
\end{aligned}
$$

Since $\max_l \delta_l \leq c$, we have the desired conclusion. $\square$

Lemma 4. *Let $\mathbf{Z}_i \sim \mathrm{TN}(0, \mathbf{I}_{p_1}, \cdots, \mathbf{I}_{p_M})$, for $i = 1, \cdots, n$, independently, and*

$$
\mathbf{L} = \mathbf{\Sigma}_m^{1/2} \{\frac{p}{np_{-m} \mathrm{tr}(\mathbf{\Sigma})} \sum_{i=1}^{n} \alpha_i (\mathbf{Z}_i)_{(m)} (\bigotimes_{m' \neq m} \mathbf{\Sigma}_{m'}) (\mathbf{Z}_i)_{(m)}^T\} \mathbf{\Sigma}_m^{1/2}.
$$

We have

$$\mathbb{P}(\|\mathbf{L} - \frac{p_m \sum_{i=1}^n \alpha_i}{n\mathrm{tr}(\boldsymbol{\Sigma}_m)}\boldsymbol{\Sigma}_m\|_2 \geq \epsilon) \leq \exp(C_1 p_m - C_2 n p_{-m} \min(\epsilon^2/(\sum_{i=1}^n \alpha_i^2/n), \epsilon/\max_i \alpha_i)),$$

for some constant C_1 and C_2.

Lemma 5. *For $m = 1, \cdots, M$, $\|\widehat{\boldsymbol{\Sigma}}_m - \frac{p_m}{\mathrm{tr}(\boldsymbol{\Sigma}_m)}\boldsymbol{\Sigma}_m\|_2 = O(\sqrt{\frac{p_m}{np_{-m}}} + \sqrt{\log(n)/p} + \sqrt{1/n})$ with probability at least $1 - C_1 n^{-C_2} - C_3 \exp(-p_m)$.*

Proof. We decompose $\widehat{\boldsymbol{\Sigma}}_m$ in the following ways,

$$\begin{aligned}
\widehat{\boldsymbol{\Sigma}}_m &= \frac{1}{np_{-m}} \sum_{i=1}^n \widehat{\omega}_i (\mathbf{Y}_i - \mathbf{B}\bar{\times}_{(M+1)}\mathbf{X}_i)_{(m)} (\mathbf{Y}_i - \mathbf{B}\bar{\times}_{(M+1)}\mathbf{X}_i)_{(m)}^T \\
&+ \frac{1}{np_{-m}} \sum_{i=1}^n \widehat{\omega}_i (\mathbf{Y}_i - \mathbf{B}\bar{\times}_{(M+1)}\mathbf{X}_i)_{(m)} \{(\mathbf{B} - \widehat{\mathbf{B}}^{\mathrm{OLS}})\bar{\times}_{(M+1)}\mathbf{X}_i\}_{(m)}^T \\
&+ \frac{1}{np_{-m}} \sum_{i=1}^n \widehat{\omega}_i \{(\mathbf{B} - \widehat{\mathbf{B}}^{\mathrm{OLS}})\bar{\times}_{(M+1)}\mathbf{X}_i\}_{(m)} (\mathbf{Y}_i - \mathbf{B}\bar{\times}_{(M+1)}\mathbf{X}_i)_{(m)}^T \\
&+ \frac{1}{np_{-m}} \sum_{i=1}^n \widehat{\omega}_i \{(\mathbf{B} - \widehat{\mathbf{B}}^{\mathrm{OLS}})\bar{\times}_{(M+1)}\mathbf{X}_i\}_{(m)} \{(\mathbf{B} - \widehat{\mathbf{B}}^{\mathrm{OLS}})\bar{\times}_{(M+1)}\mathbf{X}_i\}_{(m)}^T \\
&= \mathbf{L}_1 + \mathbf{L}_2 + \mathbf{L}_3 + \mathbf{L}_4.
\end{aligned}$$

Note that $\mathbf{Y}_i - \mathbf{B}\bar{\times}_{(M+1)}\mathbf{X}_i = \mathbf{E}_i = \frac{1}{G_i}\boldsymbol{\Sigma}^{1/2}\mathbf{Z}_i$, where $\mathbf{Z}_i \sim \mathrm{TN}(0, \mathbf{I}_{p_1}, \cdots, \mathbf{I}_{p_M})$. We can further decompose $\mathbf{L}_1$ in the following ways.

$$\begin{aligned}
\mathbf{L}_1 &= \boldsymbol{\Sigma}_m^{1/2}\{\frac{1}{np_{-m}} \sum_{i=1}^n \widehat{\omega}_i/G_i (\mathbf{Z}_i)_{(m)} (\bigotimes_{m' \neq m} \boldsymbol{\Sigma}_{m'}) (\mathbf{Z}_i)_{(m)}^T\}\boldsymbol{\Sigma}_m^{1/2} \\
&= \boldsymbol{\Sigma}_m^{1/2}\{\frac{p}{np_{-m}\mathrm{tr}(\boldsymbol{\Sigma})} \sum_{i=1}^n (\mathbf{Z}_i)_{(m)} (\bigotimes_{m' \neq m} \boldsymbol{\Sigma}_{m'}) (\mathbf{Z}_i)_{(m)}^T\}\boldsymbol{\Sigma}_m^{1/2} \\
&\quad + \boldsymbol{\Sigma}_m^{1/2} \frac{1}{np_{-m}} \sum_{i=1}^n \{(\widehat{\omega}_i/G_i - \frac{p}{\mathrm{tr}(\boldsymbol{\Sigma})})(\mathbf{Z}_i)_{(m)} (\bigotimes_{m' \neq m} \boldsymbol{\Sigma}_{m'}) (\mathbf{Z}_i)_{(m)}^T\}\boldsymbol{\Sigma}_m^{1/2} \\
&= \mathbf{L}_{11} + \mathbf{L}_{12}.
\end{aligned}$$

By Lemma 4, we know that

$$\mathbb{P}(\|\mathbf{L}_{11} - \frac{p_m}{\mathrm{tr}(\boldsymbol{\Sigma}_m)}\boldsymbol{\Sigma}_m\|_2 \geq \epsilon) \leq \exp(C_1 p_m - C_2 n p_{-m} \epsilon^2),$$

Let $\epsilon^2 = (C_1 + 1)p_m/(np_{-m}C_2)$, we have

$$\mathbb{P}(\|\mathbf{L}_{11} - \frac{p_m}{\mathrm{tr}(\boldsymbol{\Sigma}_m)}\boldsymbol{\Sigma}_m\|_2 \geq \epsilon) \leq \exp(-p_m). \tag{2}$$

For $\mathbf{L}_{12}$, we have

$$\|\mathbf{L}_{12}\|_2 \leq \max_i |\widetilde{\omega}_i/G_i - \frac{p}{\mathrm{tr}(\mathbf{\Sigma})}| \|\mathbf{\Sigma}_m^{1/2}\{\frac{1}{np_{-m}}\sum_{i=1}^n (\mathbf{Z}_i)_{(m)}(\bigotimes_{m'\neq m}\mathbf{\Sigma}_{m'})(\mathbf{Z}_i)_{(m)}^T\}\mathbf{\Sigma}_m^{1/2}\|$$
$$= \frac{\mathrm{tr}(\mathbf{\Sigma})}{p}\max_i |\widetilde{\omega}_i/G_i - \frac{p}{\mathrm{tr}(\mathbf{\Sigma})}| \|\mathbf{L}_{11}\|_2$$

Then we know that

$$\|\mathbf{L}_1 - \frac{p_m}{\mathrm{tr}(\mathbf{\Sigma}_m)}\mathbf{\Sigma}_m\|_2 \leq \|\mathbf{L}_{11} - \frac{p_m}{\mathrm{tr}(\mathbf{\Sigma}_m)}\mathbf{\Sigma}_m\|_2 + \|\mathbf{L}_{12}\|_2 \leq \|\mathbf{L}_{11} - \frac{p_m}{\mathrm{tr}(\mathbf{\Sigma}_m)}\mathbf{\Sigma}_m\|_2$$
$$+ \frac{\mathrm{tr}(\mathbf{\Sigma})}{p}\max_i |\widetilde{\omega}_i/G_i - \frac{p}{\mathrm{tr}(\mathbf{\Sigma})}|(\|\mathbf{L}_{11} - \frac{p_m}{\mathrm{tr}(\mathbf{\Sigma}_m)}\mathbf{\Sigma}_m\|_2 + \|\frac{p_m}{\mathrm{tr}(\mathbf{\Sigma}_m)}\mathbf{\Sigma}_m\|_2)$$

By Lemma 2,

$$\|\mathbf{L}_1 - \frac{p_m}{\mathrm{tr}(\mathbf{\Sigma}_m)}\mathbf{\Sigma}_m\|_2 = O(\sqrt{p_m/(np_{-m})}) + O(\sqrt{\log(n)/p} + 1/n),$$

with probability at least $1 - 4n^{-C_1} - \exp(-p_m)$.

Next, we consider the term $\mathbf{L}_4$. Note that

$$\mathbf{L}_4 = \frac{1}{np_{-m}}\sum_{i=1}^n \widehat{\omega}_i\{(\mathbf{B} - \widehat{\mathbf{B}}^{\mathrm{OLS}})\bar{\times}_{(M+1)}\mathbf{X}_i\}_{(m)}\{(\mathbf{B} - \widehat{\mathbf{B}}^{\mathrm{OLS}})\bar{\times}_{(M+1)}\mathbf{X}_i\}_{(m)}^T$$
$$= \frac{p}{np_{-m}\mathrm{tr}(\mathbf{\Sigma})}\sum_{i=1}^n G_i\{(\mathbf{B} - \widehat{\mathbf{B}}^{\mathrm{OLS}})\bar{\times}_{(M+1)}\mathbf{X}_i\}_{(m)}\{(\mathbf{B} - \widehat{\mathbf{B}}^{\mathrm{OLS}})\bar{\times}_{(M+1)}\mathbf{X}_i\}_{(m)}^T$$
$$+ \frac{1}{np_{-m}}\sum_{i=1}^n (\widehat{\omega}_i/G_i - \frac{p}{\mathrm{tr}(\mathbf{\Sigma})})G_i\{(\mathbf{B} - \widehat{\mathbf{B}}^{\mathrm{OLS}})\bar{\times}_{(M+1)}\mathbf{X}_i\}_{(m)}\{(\mathbf{B} - \widehat{\mathbf{B}}^{\mathrm{OLS}})\bar{\times}_{(M+1)}\mathbf{X}_i\}_{(m)}^T$$
$$= \mathbf{L}_{41} + \mathbf{L}_{42}$$

Becasue

$$\sqrt{n}(\widehat{\mathbf{B}}^{\mathrm{OLS}} - \mathbf{B}) \mid (G_1, \cdots G_n) \sim \mathrm{TN}(0, \mathbf{\Sigma}_1, \cdots, \mathbf{\Sigma}_M, \mathbf{\Sigma}_{\mathbf{X},G}^{-1}).$$

By Lemma 4, we have

$$\mathbb{P}(\|\mathbf{L}_{41} - \frac{p_m\sum_{i=1}^n \mathbf{X}_i^T\mathbf{\Sigma}_{\mathbf{X},G}\mathbf{X}_iG_i/n}{n\mathrm{tr}(\mathbf{\Sigma}_m)}\mathbf{\Sigma}_m\|_2 \geq \epsilon \mid G_1, \cdots, G_n)$$
$$\leq \exp\big\{C_1p_m - C_2n^2p_{-m}\min\big(\frac{\epsilon^2}{\sum_{i=1}^n (G_i\mathbf{X}_i^T\mathbf{\Sigma}_{\mathbf{X},G}^{-1}\mathbf{X}_i)^2/n}, \frac{\epsilon}{\max_i(G_i\mathbf{X}_i^T\mathbf{\Sigma}_{\mathbf{X},G}^{-1}\mathbf{X}_i)}\big)\big\}.$$

Note that $\mathbf{X}_i$ are bounded, the eigenvalue of $\mathbf{\Sigma}_{\mathbf{X},G}$ are lower bounded by C_x, $\mathbf{G}_i \sim \chi^2_\nu/\nu$, which implies $\sum_{i=1}^n G_i^2/n$ are upper bounded by some constant and $\max_i G_i$ are upper bounded by $c\log(n)$ with high probablity. Here we let $\epsilon^2 = (C_1+1)p_m/(C_2n^2p_{-m})$. We have

$$\|\mathbf{L}_{41}\|_2 = 1/n \cdot O(1 + \sqrt{\frac{p_m}{np_{-m}}})$$

with probability at least $1-\exp(-p_m)$. For $\mathbf{L}_{42}$, we have

$$
\begin{aligned}
\|\mathbf{L}_{42}\|_2 &\leq \max_i |\widetilde{\omega}_i/G_i - \frac{p}{\mathrm{tr}(\boldsymbol{\Sigma})}| \|\sum_{i=1}^n G_i\{(\mathbf{B}-\widehat{\mathbf{B}}^{\mathrm{OLS}})\bar{\times}_{(M+1)}\mathbf{X}_i\}_{(m)}\{(\mathbf{B}-\widehat{\mathbf{B}}^{\mathrm{OLS}})\bar{\times}_{(M+1)}\mathbf{X}_i\}_{(m)}^T\| \\
&= \frac{\mathrm{tr}(\boldsymbol{\Sigma})}{p}\max_i |\widetilde{\omega}_i/G_i - \frac{p}{\mathrm{tr}(\boldsymbol{\Sigma})}| \|\mathbf{L}_{41}\|_2
\end{aligned}
$$

Thus,

$$
\|\mathbf{L}_4\|_2 = O(1/n + \sqrt{\frac{p_m}{np_{-m}}} + \sqrt{\log(n)/p})
$$

with probability at least $1-4n^{-C_1}-\exp(-p_m)$.

Next, we consider $\mathbf{L}_2$ and $\mathbf{L}_3$. Let $\mathbb{X}\in\mathbb{R}^{q\times n}$ be the stacked sample matrix of $\mathbf{X}_i$, $\mathbb{E}\in\mathbb{R}^{p_1\times\cdots\times p_M\times n}$ be the stacked sample tensor of $\mathbf{E}_i$, and $\mathbb{W}\in\mathbb{R}^{n\times n}$ be a diagonal matirx with the i-th diagonal element to be $\widehat{\omega}_i$. We have

$$
\begin{aligned}
\|\mathbf{L}_2\|_2 &= \|\frac{1}{np_{-m}}\{(\widehat{\mathbf{B}}^{\mathrm{OLS}}-\mathbf{B})\times_{M+1}\mathbb{X}\mathbb{W}^{1/2}\}_{(m)}\{\mathbb{E}\times_{M+1}\mathbb{W}^{1/2}\}_{(m)}^T\|_2 \\
&\leq \frac{1}{np_{-m}}\|\{(\widehat{\mathbf{B}}^{\mathrm{OLS}}-\mathbf{B})\times_{M+1}\mathbb{X}\mathbb{W}^{1/2}\}_{(m)}\{(\widehat{\mathbf{B}}^{\mathrm{OLS}}-\mathbf{B})\times_{M+1}\mathbb{X}\mathbb{W}^{1/2}\}_{(m)}^T\|_2^{1/2} \\
&\quad\cdot\|\{\mathbb{E}\times_{M+1}\mathbb{W}^{1/2}\}_{(m)}\{\mathbb{E}\times_{M+1}\mathbb{W}^{1/2}\}_{(m)}^T\|_2^{1/2} \\
&\leq \frac{1}{np_{-m}}\|\sum_{i=1}^n\widehat{\omega}_i\{(\widehat{\mathbf{B}}^{\mathrm{OLS}}-\mathbf{B})\bar{\times}_{M+1}\mathbf{X}_i\}_{(m)}\{(\widehat{\mathbf{B}}^{\mathrm{OLS}}-\mathbf{B})\bar{\times}_{M+1}\mathbf{X}_i\}_{(m)}^T\|_2^{1/2}\|\sum_{i=1}^n\widehat{\omega}_i(\mathbf{E}_i)_{(m)}(\mathbf{E}_i)_{(m)}^T\|_2^{1/2} \\
&= \sqrt{\|\mathbf{L}_1\|_2\|\mathbf{L}_4\|_2}.
\end{aligned}
$$

It follows that

$$
\|\mathbf{L}_2\| = O(\sqrt{1/n} + \sqrt{\frac{p_m}{np_{-m}}} + \sqrt{\log(n)/p})
$$

with probability at least $1-8n^{-C_1}-2\exp(-p_m)$. Note that $\mathbf{L}_3=\mathbf{L}_2^T$, we have the same conclusion for $\mathbf{L}_3$.

Lemma 6. *The estimated envelope score satisfies that* $|\widehat{\phi}_l-\phi_l| = O(\max_m\sqrt{p_m/(np_{-m})} + \sqrt{1/n}+\sqrt{\log(n)/p})$ *with probability at least* $1-4n^{-C_1}-C_2\exp(-C_3C_M)-C_4\sum_{m=1}^M\exp(-p_m)$, *for all* $l=1,\cdots,p$.

Proof. For any fixed l,

$$
|\widehat{\phi}_l-\phi_l| = |\|[\![\widehat{\mathbf{B}};\widehat{\mathbf{v}}_{l_1}^{(1)},\cdots,\widehat{\mathbf{v}}_{l_M}^{(M)}]\!]\|_2 - \|[\![\mathbf{B};\mathbf{v}_{l_1}^{(1)},\cdots,\mathbf{v}_{l_M}^{(M)}]\!]\|_2|,
$$

for some $(l_1, \cdots, l_M)$. Note that

$$
\begin{aligned}
&|\|[\![\widehat{\mathbf{B}}; \widehat{\mathbf{v}}_{l_1}^{(1)}, \cdots, \widehat{\mathbf{v}}_{l_M}^{(M)}]\!]\|_2 - \|[\![\mathbf{B}; \mathbf{v}_{l_1}^{(1)}, \cdots, \mathbf{v}_{l_M}^{(M)}]\!]\|_2| \\
&\leq \|[\![\widehat{\mathbf{B}}; \widehat{\mathbf{v}}_{l_1}^{(1)}, \cdots, \widehat{\mathbf{v}}_{l_M}^{(M)}]\!] - [\![\mathbf{B}; \widehat{\mathbf{v}}_{l_1}^{(1)}, \cdots, \widehat{\mathbf{v}}_{l_M}^{(M)}]\!]\|_2 + \|[\![\mathbf{B}; \widehat{\mathbf{v}}_{l_1}^{(1)}, \cdots, \widehat{\mathbf{v}}_{l_M}^{(M)}]\!] - [\![\mathbf{B}; \mathbf{v}_{l_1}^{(1)}, \cdots, \mathbf{v}_{l_M}^{(M)}]\!]\|_2 \\
&\leq \|[\![\widehat{\mathbf{B}} - \mathbf{B}; \widehat{\mathbf{v}}_{l_1}^{(1)}, \cdots, \widehat{\mathbf{v}}_{l_M}^{(M)}]\!]\|_2 + \|(\otimes_{m=M}^1 \widehat{\mathbf{v}}_{l_m}^{(m)} - \otimes_{m=M}^1 \mathbf{V}_{l_m}^{(m)})^T \mathbf{B}_{(M+1)}^T\|_2 \\
&= \|(\otimes_{m=M}^1 \widehat{\mathbf{v}}_{l_m}^{(m)})^T (\widehat{\mathbf{B}} - \mathbf{B})_{(M+1)}^T\|_2 + \|(\otimes_{m=M}^1 \widehat{\mathbf{v}}_{l_m}^{(m)} - \otimes_{m=M}^1 \mathbf{v}_{l_m}^{(m)})^T \mathbf{B}_{(M+1)}^T\|_2 \\
&= I + II.
\end{aligned}
$$

We first consider term II. By Theorem 2 in Yu et al. (2015), we know that

$$
\|\sin\Theta(\mathbf{v}_{l_m}, \widehat{\mathbf{v}}_{l_m})\|_F \leq \frac{2\|\widehat{\mathbf{\Sigma}}_m - \mathbf{\Sigma}_m\|_2}{\Delta},
$$

for a positive constant Δ. It follows that

$$
\|\mathbf{P}_{\widehat{\mathbf{v}}_{l_m}} - \mathbf{P}_{\mathbf{v}_{l_m}}\|_F \leq \frac{2\sqrt{2}\|\widehat{\mathbf{\Sigma}}_m - \mathbf{\Sigma}_m\|_2}{\Delta}. \tag{3}
$$

Then note that

$$
\begin{aligned}
&\|\mathbf{P}_{\otimes_{m=M}^1 \widehat{\mathbf{v}}_{l_m}^{(m)}} - \mathbf{P}_{\otimes_{m=M}^1 \mathbf{v}_{l_m}^{(m)}}\|_F \\
&\leq \|(\otimes_{m=M}^2 \mathbf{P}_{\widehat{\mathbf{v}}_{l_m}}) \otimes (\mathbf{P}_{\widehat{\mathbf{v}}_{l_1}} - \mathbf{P}_{\mathbf{v}_{l_1}})\|_F + \sum_{k=2}^{M-1} \|(\otimes_{m=M}^{k+1} (\mathbf{P}_{\widehat{\mathbf{v}}_{l_m}} - \mathbf{P}_{\mathbf{v}_{l_m}})(\otimes_{m=j-1}^1 \mathbf{P}_{\mathbf{v}_{l_m}})\|_F \\
&+ \|(\mathbf{P}_{\widehat{\mathbf{v}}_{l_M}} - \mathbf{P}_{\mathbf{v}_{l_M}}) \otimes (\otimes_{m=M-1}^1 \mathbf{P}_{\mathbf{v}_{l_m}})\|_F.
\end{aligned}
$$

By Lemma 5, we have

$$
\|\mathbf{P}_{\otimes_{m=M}^1 \widehat{\mathbf{v}}_{l_m}^{(m)}} - \mathbf{P}_{\otimes_{m=M}^1 \mathbf{v}_{l_m}^{(m)}}\|_F = O(\max_m \sqrt{\frac{p_m}{np_{-m}}} + \sqrt{\log(n)/p} + \sqrt{1/n})
$$

with probability at least $1 - MC_1 n^{-C_2} - C_3 \sum_{m=1}^M \exp(-p_m)$.

Hence

$$
II \leq \|\mathbf{P}_{\otimes_{m=M}^1 \widehat{\mathbf{v}}_{l_m}^{(m)}} - \mathbf{P}_{\otimes_{m=M}^1 \mathbf{v}_{l_m}^{(m)}}\|_F \|\mathbf{B}_{(M+1)}\|_2 = O(\max_m \sqrt{\frac{p_m}{np_{-m}}} + \sqrt{\log(n)/p} + \sqrt{1/n})
$$

Then we consider term I. Recall that

$$
\widehat{\mathbf{B}}_{(M+1)}^T = (\frac{1}{n}\sum_{i=1}^n \widehat{\omega}_i \text{vec}(\mathbf{Y}_i)\mathbf{X}_i^T)(\frac{1}{n}\sum_{i=1}^n \widehat{\omega}_i \mathbf{X}_i \mathbf{X}_i^T)^{-1}.
$$

Then we have

$$
\begin{aligned}
&\|(\otimes_{m=M}^1 \widehat{\mathbf{v}}_{l_m}^{(m)})^T (\widehat{\mathbf{B}} - \mathbf{B})_{(M+1)}^T\|_2 \\
&\leq \|\frac{1}{n}\sum_{i=1}^n \widehat{\omega}_i (\otimes_{m=M}^1 \widehat{\mathbf{v}}_{l_m}^{(m)})^T (\text{vec}(\mathbf{Y}_i) - \mathbf{B}_{(M+1)}^T \mathbf{X}_i)\mathbf{X}_i^T\|_2 \|(\frac{1}{n}\sum_{i=1}^n \widehat{\omega}_i \mathbf{X}_i \mathbf{X}_i^T)^{-1}\|_2^{-1}
\end{aligned}
$$

Firstly, note that

$$\begin{aligned}
\|\frac{1}{n}\sum_{i=1}^{n}\widehat{\omega}_i\mathbf{X}_i\mathbf{X}_i^T\|_2 &\leq \|\frac{1}{n}\sum_{i=1}^{n}(\widehat{\omega}_i/G_i - p/\mathrm{tr}(\mathbf{\Sigma}))G_i\mathbf{X}_i\mathbf{X}_i^T\|_2 + \|\frac{p}{n\mathrm{tr}(\mathbf{\Sigma})}\sum_{i=1}^{n}G_i\mathbf{X}_i\mathbf{X}_i^T\|_2 \\
&\leq (-\max_i|(\widehat{\omega}_i/G_i - p/\mathrm{tr}(\mathbf{\Sigma}))| + p/\mathrm{tr}(\mathbf{\Sigma}))\|\frac{1}{n}\sum_{i=1}^{n}G_i\mathbf{X}_i\mathbf{X}_i^T\|_2
\end{aligned}$$

By Lemma 2, the smallest eigenvalue of $\|\frac{1}{n}\sum_{i=1}^{n}\widehat{\omega}_i\mathbf{X}_i\mathbf{X}_i^T\|_2$ is lower bounded by a positive constant with probability at least $1-4n^{-C_2}$.

Then note that

$$\begin{aligned}
&\|\frac{1}{n}\sum_{i=1}^{n}\widehat{\omega}_i(\otimes_{m=M}^{1}\widehat{\mathbf{v}}_{l_m}^{(m)})^T(\mathrm{vec}(\mathbf{Y}_i) - \mathbf{B}_{(M+1)}^T\mathbf{X}_i)\mathbf{X}_i^T\|_2 \\
&\leq \|\frac{1}{n}\sum_{i=1}^{n}(\widehat{\omega}_i/G_i - p/\mathrm{tr}(\mathbf{\Sigma}))G_i(\otimes_{m=M}^{1}\widehat{\mathbf{v}}_{l_m}^{(m)})^T(\mathrm{vec}(\mathbf{Y}_i) - \mathbf{B}_{(M+1)}^T\mathbf{X}_i)\mathbf{X}_i^T\|_2 + \\
&\|\frac{p}{n\mathrm{tr}(\Sigma)}\sum_{i=1}^{n}G_i(\otimes_{m=M}^{1}\widehat{\mathbf{v}}_{l_m}^{(m)})^T(\mathrm{vec}(\mathbf{Y}_i) - \mathbf{B}_{(M+1)}^T\mathbf{X}_i)\mathbf{X}_i^T\|_2 \\
&= I_1 + I_2.
\end{aligned}$$

For I_2, we have

$$\mathbb{P}(I_2 \geq \epsilon) = \mathbb{P}(\|\frac{p}{n\mathrm{tr}(\Sigma)}\sum_{i=1}^{n}\sqrt{G_i}Q_i\mathbf{X}_i^T\|_2 \geq \epsilon),$$

where $Q_i \sim N(0, (\otimes_{m=M}^{1}\widehat{\mathbf{v}}_{l_m}^{(m)})^T\mathbf{\Sigma}(\otimes_{m=M}^{1}\widehat{\mathbf{v}}_{l_m}^{(m)}))$ (as a result of data splitting). Since all the elements of $\mathbf{X}_i$ are upper bounded by M_x, $\sqrt{G_i}$ is sub-Gaussian and independent of $\mathbf{Q}_i$, by Bernstein's inequality, $I_2 \leq C_M\sqrt{1/n}$ with probability at least $1-\exp(-C_1C_M)$.

For term I_1, we have

$$I_1 \leq \max_i|\widehat{\omega}_i/G_i - p/\mathrm{tr}(\mathbf{\Sigma})|I_2.$$

By Lemma 2, we have $I_1 = O(\sqrt{\log(n)/p}+\sqrt{1/n})\sqrt{1/n}$ with probability at least $1-\exp(-C_1C_M)-4n^{-C_2}$.

To sum up, we have $|\widehat{\phi}_l - \phi_l| = O(\max_m\sqrt{p_m/(np_{-m})}+\sqrt{1/n}+\sqrt{\log(n)/p})$ with probability at least $1-4n^{-C_1}-C_2\exp(-C_3C_M)-C_4\sum_{m=1}^{M}\exp(-p_m)$.

$\square$

The Lemma tells that we can correctly select the subspace dimension with high probability. Let $\widehat{\boldsymbol{\eta}}$ be an estimated basis matrix of $\mathcal{F}_{\mathbf{\Sigma}}(\mathbf{B})$ obtained by the proposed algorithm in sample and $\boldsymbol{\eta}$

is a basis matrix for $\mathcal{F}_{\boldsymbol{\Sigma}}(\mathbf{B})$. Combine this fact with Lemma 5 and Theorem 2 in Yu et al. (2015), we have the following result.

Lemma 7. *For positive integers C_1, C_2, and C_3,*

$$\|\mathbf{P}_{\widehat{\boldsymbol{\eta}}} - \mathbf{P}_{\boldsymbol{\eta}}\|_2 = O(\sqrt{1/n} + \sqrt{\frac{p_m}{np_{-m}}} + \sqrt{\log(n)/p})$$

with probability at least $1 - MC_1 n^{-C_2} - C_3 \sum_{m=1}^{M} \exp(-p_m)$.

We first consider the case eigenvalues of $\mathbf{P}_{\boldsymbol{\Gamma}_m}\boldsymbol{\Sigma}_m\mathbf{P}_{\boldsymbol{\Gamma}_m}$ are all different and are distinct from those of $\mathbf{Q}_{\boldsymbol{\Gamma}_m}\boldsymbol{\Sigma}_m\mathbf{Q}_{\boldsymbol{\Gamma}_m}$. In this case, the j-th column of $\boldsymbol{\beta}$ denoted as $\boldsymbol{\eta}_j = \otimes_{m=M}^{1}\mathbf{v}_{mj}$, where $\mathbf{v}_{mj}$ is an eigenvector of $\mathbf{P}_{\boldsymbol{\Gamma}_m}\boldsymbol{\Sigma}_m\mathbf{P}_{\boldsymbol{\Gamma}_m}$. By Theorem 2 in Yu et al. (2015), we know that

$$\|\sin\Theta(\mathbf{v}_{mj}, \widehat{\mathbf{v}}_{mj})\|_F \leq \frac{2\|\widehat{\boldsymbol{\Sigma}}_m - \boldsymbol{\Sigma}_m\|_2}{\Delta},$$

for a positive constant Δ. It follows that

$$\|\mathbf{P}_{\widehat{\mathbf{v}}_{mj}} - \mathbf{P}_{\mathbf{v}_{mj}}\|_F \leq \frac{2\sqrt{2}\|\widehat{\boldsymbol{\Sigma}}_m - \boldsymbol{\Sigma}_m\|_2}{\Delta}. \tag{4}$$

Then note that

$$\begin{aligned}\|\mathbf{P}_{\widehat{\boldsymbol{\eta}}_j} - \mathbf{P}_{\boldsymbol{\eta}_j}\|_F \leq{}& \|(\otimes_{m=M}^{2}\mathbf{P}_{\widehat{\mathbf{v}}_{mj}}) \otimes (\mathbf{P}_{\widehat{\mathbf{v}}_{1j}} - \mathbf{P}_{\mathbf{v}_{1j}})\|_F + \sum_{k=2}^{M-1} \|(\otimes_{m=M}^{k+1}(\mathbf{P}_{\widehat{\mathbf{v}}_{mj}} - \mathbf{P}_{\mathbf{v}_{mj}})(\otimes_{m=j-1}^{1}\mathbf{P}_{\mathbf{v}_{mj}})\|_F \\ &+ \|(\mathbf{P}_{\widehat{\mathbf{v}}_{Mj}} - \mathbf{P}_{\mathbf{v}_{Mj}}) \otimes (\otimes_{m=M-1}^{1}\mathbf{P}_{\mathbf{v}_{mj}})\|_F.\end{aligned}$$

By Lemma 5, we have

$$\|\mathbf{P}_{\widehat{\boldsymbol{\eta}}_j} - \mathbf{P}_{\boldsymbol{\eta}_j}\|_F = O(\max_m \sqrt{\frac{p_m}{np_{-m}}} + \sqrt{\log(n)/p} + \sqrt{1/n})$$

with probability at least $1 - MC_1 n^{-C_2} - C_3 \sum_{m=1}^{M} \exp(-p_m)$. It follows that

$$\|\mathbf{P}_{\widehat{\boldsymbol{\eta}}} - \mathbf{P}_{\boldsymbol{\eta}}\|_F = O(\max_m \sqrt{\frac{p_m}{np_{-m}}} + \sqrt{\log(n)/p} + \sqrt{1/n})$$

with probability at least $1 - MC_1 n^{-C_2} - C_3 \sum_{m=1}^{M} \exp(-p_m)$, since R is assume to be a constant.

□

2 Proof of Theorem 1

Proof. Recall that

$$\widehat{\mathbf{B}}^{\text{CATL}} = \sum_{j=1}^{R} [\![\widehat{\mathbf{B}}; \mathbf{v}^{(1)}_{l_1^{(j)}}(\mathbf{v}^{(1)}_{l_1^{(j)}})^T, \cdots, \mathbf{v}^{(M)}_{l_M^{(j)}}(\mathbf{v}^{(M)}_{l_M^{(j)}})^T]\!],$$

and $\widehat{\eta}$ is a basis matrix of $\mathcal{F}_{\mathbf{\Sigma}}(\mathbf{B})$, we have $(\widehat{\mathbf{B}}^{\text{CATL}}_{(M+1)})^T = \mathbf{P}_{\widehat{\boldsymbol{\eta}}}\widehat{\mathbf{B}}^T_{(M+1)}$, where

$$\widehat{\mathbf{B}}^T_{(M+1)} = (\frac{1}{n}\sum_{i=1}^{n}\widehat{\omega}_i \text{vec}(\mathbf{Y}_i)\mathbf{X}_i^T)(\frac{1}{n}\sum_{i=1}^{n}\widehat{\omega}_i\mathbf{X}_i\mathbf{X}_i^T)^{-1}.$$

By definition $\mathbf{B}^T_{(M+1)} = \mathbf{P}_{\boldsymbol{\eta}}\mathbf{B}^T_{(M+1)}$. We have

$$\begin{aligned}
\|\widehat{\mathbf{B}}^{\text{CATL}} - \mathbf{B}^T_{(M+1)}\|_2 &= \|\mathbf{P}_{\widehat{\boldsymbol{\eta}}}\widehat{\mathbf{B}}^T_{(M+1)} - \mathbf{P}_{\boldsymbol{\eta}}\mathbf{B}^T_{(M+1)}\|_2 \\
&\le \|\mathbf{P}_{\widehat{\boldsymbol{\eta}}}(\widehat{\mathbf{B}}^T_{(M+1)} - \mathbf{B}^T_{(M+1)})\|_2 + \|(\mathbf{P}_{\widehat{\boldsymbol{\eta}}} - \mathbf{P}_{\boldsymbol{\eta}})\mathbf{B}^T_{(M+1)}\|_2 \\
&\le \|\widehat{\boldsymbol{\eta}}^T(\widehat{\mathbf{B}}^T_{(M+1)} - \mathbf{B}^T_{(M+1)})\|_2 + \|\mathbf{P}_{\widehat{\boldsymbol{\eta}}} - \mathbf{P}_{\boldsymbol{\eta}}\|_2\|\mathbf{B}^T_{(M+1)}\|_2.
\end{aligned}$$

By Lemma 7, we have

$$\|\mathbf{P}_{\widehat{\boldsymbol{\eta}}} - \mathbf{P}_{\boldsymbol{\eta}}\|_F = O(\max_m \sqrt{\frac{p_m}{np_{-m}}} + \sqrt{\log(n)/p} + \sqrt{1/n})$$

with probability at least $1 - MC_1n^{-C_2} - C_3\sum_{m=1}^{M}\exp(-p_m)$. Also, since α_{rk} are all bounded, we have

$$\|\mathbf{P}_{\widehat{\boldsymbol{\eta}}} - \mathbf{P}_{\boldsymbol{\eta}}\|_2\|\mathbf{B}^T_{(M+1)}\|_2 = O(\max_m \sqrt{\frac{p_m}{np_{-m}}} + \sqrt{\log(n)/p} + \sqrt{1/n})$$

with probability at least $1 - MC_1n^{-C_2} - C_3\sum_{m=1}^{M}\exp(-p_m)$.

For the first term of right hand side, we have

$$\|\widehat{\boldsymbol{\eta}}^T(\widehat{\mathbf{B}}^T_{(M+1)} - \mathbf{B}^T_{(M+1)})\|_2 \le \|\frac{1}{n}\sum_{i=1}^{n}\widehat{\omega}_i\widehat{\boldsymbol{\eta}}^T(\text{vec}(\mathbf{Y}_i) - \mathbf{B}^T_{(M+1)}\mathbf{X}_i)\mathbf{X}_i^T\|_2\|\frac{1}{n}\sum_{i=1}^{n}\widehat{\omega}_i\mathbf{X}_i\mathbf{X}_i^T\|_2^{-1}.$$

Then note that

$$\begin{aligned}
&\|\frac{1}{n}\sum_{i=1}^{n}\widehat{\omega}_i\widehat{\boldsymbol{\eta}}^T(\text{vec}(\mathbf{Y}_i) - \mathbf{B}^T_{(M+1)}\mathbf{X}_i)\mathbf{X}_i^T\|_2 \\
&\le \|\frac{1}{n}\sum_{i=1}^{n}(\widehat{\omega}_i/G_i - p/\text{tr}(\mathbf{\Sigma}))\widehat{\boldsymbol{\eta}}^T(\text{vec}(\mathbf{Y}_i) - \mathbf{B}^T_{(M+1)}\mathbf{X}_i)\mathbf{X}_i^T\|_2 \\
&\quad + \|\frac{p}{n\text{tr}(\mathbf{\Sigma})}\sum_{i=1}^{n}G_i\widehat{\boldsymbol{\eta}}^T(\text{vec}(\mathbf{Y}_i) - \mathbf{B}^T_{(M+1)}\mathbf{X}_i)\mathbf{X}_i^T\|_2 \\
&= I_1 + I_2
\end{aligned}$$

For term I_2,

$$\mathbb{P}(I_2 \geq \epsilon) = \mathbb{P}(\|\frac{p}{n\mathrm{tr}(\Sigma)}\sum_{i=1}^{n}\sqrt{G_i}Q_i\mathbf{X}_i^T\|_2 \geq \epsilon),$$

where $Q_i \sim N(0, \widehat{\boldsymbol{\eta}}^T\boldsymbol{\Sigma}\widehat{\boldsymbol{\eta}})$ (as a result of data splitting). Since all the elements of $\mathbf{X}_i$ are upper bounded by M_x, $\sqrt{G_i}$ is sub-Gaussian and independent of $\mathbf{Q}_i$, by Bernstein's inequality, $I_2 \leq C_M\sqrt{1/n}$ with probability at least $1 - \exp(-C_1C_M)$.

For term I_1, we have

$$I_1 \leq \max_i |\widehat{\omega}_i/G_i - p/\mathrm{tr}(\boldsymbol{\Sigma})|I_2.$$

By Lemma 2, we have $I_1 = O(\sqrt{\log(n)/p}+\sqrt{1/n})\sqrt{1/n}$ with probability at least $1-\exp(-C_1C_M)-4n^{-C_2}$.

Finally, note that

$$\begin{aligned}\|\frac{1}{n}\sum_{i=1}^{n}\widehat{\omega}_i\mathbf{X}_i\mathbf{X}_i^T\|_2 &\leq \|\frac{1}{n}\sum_{i=1}^{n}(\widehat{\omega}_i/G_i - p/\mathrm{tr}(\boldsymbol{\Sigma}))G_i\mathbf{X}_i\mathbf{X}_i^T\|_2 + \|\frac{p}{n\mathrm{tr}(\boldsymbol{\Sigma})}\sum_{i=1}^{n}G_i\mathbf{X}_i\mathbf{X}_i^T\|_2 \\ &\leq (-\max_i |(\widehat{\omega}_i/G_i - p/\mathrm{tr}(\boldsymbol{\Sigma}))| + p/\mathrm{tr}(\boldsymbol{\Sigma}))\|\frac{1}{n}\sum_{i=1}^{n}G_i\mathbf{X}_i\mathbf{X}_i^T\|_2\end{aligned}$$

By Lemma 2, the smallest eigenvalue of $\|\frac{1}{n}\sum_{i=1}^{n}\widehat{\omega}_i\mathbf{X}_i\mathbf{X}_i^T\|_2$ is lower bounded by a positive constant with probability at least $1 - 4n^{-C_2}$.

To sum up

$$\|\widehat{\mathbf{B}}^{\mathrm{CATL}} - \mathbf{B}\|_2 = O(\sqrt{1/n} + \max_m\sqrt{\frac{p_m}{np_{-m}}} + \sqrt{\log(n)/p})$$

with probability at least $1 - 4n^{-C_1} - C_2\exp(-C_3C_M) - C_4\sum_{m=1}^{M}\exp(-p_m)$.

$\square$

References

Wang, N., Zhang, X. & Mai, Q. (2022), 'High-dimensional tensor response regression using the t-distribution', *Manuscript* .

Yu, Y., Wang, T. & Samworth, R. J. (2015), 'A useful variant of the davis–kahan theorem for statisticians', *Biometrika* **102**(2), 315–323.