Supplementary Material to "Efficient Inference for Low-Rank Regression via Covariance-weighted Projection"

Baofang Ke[1,2], Zihao Song[3], Weihua Zhao[3] and Lei Wang*[2]

[1] *Fujian Normal University,* [2] *Nankai University and* [3] *Nantong University*

S1. Matrix regression as the special case $K = 2$

Taking $K = 2$ yields low-rank matrix regression. Let $Z_i = \boldsymbol{X}_i \in \mathbb{R}^{p\times q}$, $\Theta^* = \mathbf{T}^* \in \mathbb{R}^{p\times q}$, and let $\mathcal{M}$ denote the manifold of rank-r matrices. Write $\mathbf{T}^* = \boldsymbol{U}^*\boldsymbol{V}^{*\top}$, where $\boldsymbol{U}^* \in \mathbb{R}^{p\times r}$ and $\boldsymbol{V}^* \in \mathbb{R}^{q\times r}$. The tangent space at $\mathbf{T}^*$ is

$$T_{\mathbf{T}^*}\mathcal{M} = \left\{\boldsymbol{U}^*\boldsymbol{B}^\top + \boldsymbol{C}\boldsymbol{V}^{*\top} \;\middle|\; \boldsymbol{B} \in \mathbb{R}^{q\times r},\ \boldsymbol{C} \in \mathbb{R}^{p\times r}\right\}.$$

Assume that the design covariance has the Kronecker form $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}^{[2]} \otimes \boldsymbol{\Sigma}^{[1]}$, where $\boldsymbol{\Sigma}^{[1]} \in \mathbb{R}^{p\times p}$ and $\boldsymbol{\Sigma}^{[2]} \in \mathbb{R}^{q\times q}$ are positive definite. The general operators introduced in Section 2.2 specialize to the following matrix forms:

$$P_{\boldsymbol{\Sigma},\mathbf{T}^*}(\mathbf{A}) = \mathbf{P}_{\mathbf{U}^*}^{\boldsymbol{\Sigma}^{[1]}}\,\boldsymbol{\Sigma}^{[1]}\mathbf{A}\boldsymbol{\Sigma}^{[2]}\,\mathbf{P}_{\mathbf{V}^*}^{\boldsymbol{\Sigma}^{[2]}} + \mathbf{P}_{\mathbf{U}^*}^{\boldsymbol{\Sigma}^{[1]}}\,\boldsymbol{\Sigma}^{[1]}\mathbf{A}\,(\mathbf{I}_q - \boldsymbol{\Sigma}^{[2]}\mathbf{P}_{\mathbf{V}^*}^{\boldsymbol{\Sigma}^{[2]}}) + (\mathbf{I}_p - \mathbf{P}_{\mathbf{U}^*}^{\boldsymbol{\Sigma}^{[1]}}\boldsymbol{\Sigma}^{[1]})\,\mathbf{A}\boldsymbol{\Sigma}^{[2]}\,\mathbf{P}_{\mathbf{V}^*}^{\boldsymbol{\Sigma}^{[2]}},$$

$$P^*_{\boldsymbol{\Sigma},\mathbf{T}^*}(\mathbf{A}) = \boldsymbol{\Sigma}^{[1]}\mathbf{P}_{\mathbf{U}^*}^{\boldsymbol{\Sigma}^{[1]}}\,\mathbf{A}\,\mathbf{P}_{\mathbf{V}^*}^{\boldsymbol{\Sigma}^{[2]}}\boldsymbol{\Sigma}^{[2]} + \boldsymbol{\Sigma}^{[1]}\mathbf{P}_{\mathbf{U}^*}^{\boldsymbol{\Sigma}^{[1]}}\,\mathbf{A}\,(\mathbf{I}_q - \mathbf{P}_{\mathbf{V}^*}^{\boldsymbol{\Sigma}^{[2]}}\boldsymbol{\Sigma}^{[2]}) + (\mathbf{I}_p - \boldsymbol{\Sigma}^{[1]}\mathbf{P}_{\mathbf{U}^*}^{\boldsymbol{\Sigma}^{[1]}})\,\mathbf{A}\,\mathbf{P}_{\mathbf{V}^*}^{\boldsymbol{\Sigma}^{[2]}}\boldsymbol{\Sigma}^{[2]}.$$

Corresponding author: Lei Wang, School of Statistics and Data Science & KLMDASR, LEBPS and LPMC, Nankai University, Tianjin, 300071, China. E-mail: lwangstat@nankai.edu.cn.

Here

$$\mathbf{P}_{\mathbf{U}^*}^{\mathbf{\Sigma}^{[1]}} = \mathbf{U}^*(\mathbf{U}^{*\top}\mathbf{\Sigma}^{[1]}\mathbf{U}^*)^{-1}\mathbf{U}^{*\top}, \qquad \mathbf{P}_{\mathbf{V}^*}^{\mathbf{\Sigma}^{[2]}} = \mathbf{V}^*(\mathbf{V}^{*\top}\mathbf{\Sigma}^{[2]}\mathbf{V}^*)^{-1}\mathbf{V}^{*\top}.$$

The corresponding projection-adjusted loading is

$$\Phi(\mathbf{A}) = P_{\mathbf{\Sigma},\mathbf{T}^*}\big((\mathbf{\Sigma}^{[1]})^{-1}\mathbf{A}(\mathbf{\Sigma}^{[2]})^{-1}\big).$$

The oracle estimator of $\Delta_{\mathbf{A}} = \langle \mathbf{A}, \mathbf{T}^* \rangle$ is obtained from (2.3) by substituting these matrix-specific expressions.

For the feasible construction, assume $N = 2n$ and split $\mathcal{D} = \{(y_i, \boldsymbol{X}_i)\}_{i=1}^{2n}$ into $\mathcal{D}_1 = \{(y_i, \boldsymbol{X}_i)\}_{i=1}^{n}$ and $\mathcal{D}_2 = \{(y_i, \boldsymbol{X}_i)\}_{i=n+1}^{2n}$. Define

$$\widetilde{\mathbf{\Sigma}}^{[1]} = \frac{1}{|\mathcal{D}_2|}\sum_{i\in\mathcal{D}_2} \boldsymbol{X}_i\boldsymbol{X}_i^\top, \qquad \widetilde{\mathbf{\Sigma}}^{[2]} = \frac{1}{|\mathcal{D}_2|}\sum_{i\in\mathcal{D}_2} \boldsymbol{X}_i^\top\boldsymbol{X}_i, \qquad b_n = \frac{1}{|\mathcal{D}_2|}\sum_{i\in\mathcal{D}_2} \|\boldsymbol{X}_i\|_F^2, \tag{S1}$$

and normalize

$$\widehat{\mathbf{\Sigma}}^{[1]} = \frac{1}{\sqrt{b_n}}\widetilde{\mathbf{\Sigma}}^{[1]}, \qquad \widehat{\mathbf{\Sigma}}^{[2]} = \frac{1}{\sqrt{b_n}}\widetilde{\mathbf{\Sigma}}^{[2]}. \tag{S2}$$

Set $\widehat{\mathbf{\Sigma}} = \widehat{\mathbf{\Sigma}}^{[2]} \otimes \widehat{\mathbf{\Sigma}}^{[1]}$. Let $\widehat{\mathbf{T}}$ be an initial estimator computed from $\mathcal{D}_2$, and let $\widehat{\mathbf{T}}_r = \widehat{\boldsymbol{U}}\widehat{\boldsymbol{V}}^\top$ denote its rank-r approximation, where $\widehat{\boldsymbol{U}} \in \mathbb{R}^{p\times r}$ and $\widehat{\boldsymbol{V}} \in \mathbb{R}^{q\times r}$ are the top-r singular vectors. Define

$$\widehat{\Phi}(\mathbf{A}) = P_{\widehat{\mathbf{\Sigma}},\widehat{\mathbf{T}}_r}\big((\widehat{\mathbf{\Sigma}}^{[1]})^{-1}\mathbf{A}(\widehat{\mathbf{\Sigma}}^{[2]})^{-1}\big).$$

The estimator based on $\mathcal{D}_1$ is

$$\widehat{\Delta}_{\mathbf{A}}^{(1)} = \langle \mathbf{A}, P_{\widehat{\mathbf{\Sigma}},\widehat{\mathbf{T}}_r}(\widehat{\mathbf{T}})\rangle + \frac{1}{n}\sum_{i=1}^{n}\langle \widehat{\Phi}(\mathbf{A}), \boldsymbol{X}_i\rangle\big(y_i - \langle \boldsymbol{X}_i, \widehat{\mathbf{T}}\rangle\big). \tag{S3}$$

Exchanging the roles of $\mathcal{D}_1$ and $\mathcal{D}_2$ yields $\widehat{\Delta}_{\mathbf{A}}^{(2)}$, and the final estimator is obtained by averaging. Algorithm S1 summarizes the matrix implementation.

Algorithm S1 Feasible Covariance-Weighted One-Step Debiased Estimate for Matrix Regression

Input: Independent datasets $\mathcal{D}_1$ and $\mathcal{D}_2$; initial estimators $\widehat{\mathbf{T}}^{(1)}$ and $\widehat{\mathbf{T}}^{(2)}$ computed from $\mathcal{D}_1$ and $\mathcal{D}_2$, respectively

Output: Debiased estimator $\widehat{\Delta}_{\mathbf{A}}$

1: Extract the top-r left and right singular vectors of $\widehat{\mathbf{T}}^{(2)}$, denoted by $\widehat{\boldsymbol{U}}$ and $\widehat{\boldsymbol{V}}$

2: Compute $\widehat{\boldsymbol{\Sigma}}^{[1]}$ and $\widehat{\boldsymbol{\Sigma}}^{[2]}$ from $\mathcal{D}_2$ according to (S1)–(S2), and set $\widehat{\boldsymbol{\Sigma}} = \widehat{\boldsymbol{\Sigma}}^{[2]} \otimes \widehat{\boldsymbol{\Sigma}}^{[1]}$

3: Compute $\widehat{\Phi}(\mathbf{A}) = P_{\widehat{\boldsymbol{\Sigma}},\widehat{\mathbf{T}}_r}\big((\widehat{\boldsymbol{\Sigma}}^{[1]})^{-1}\mathbf{A}(\widehat{\boldsymbol{\Sigma}}^{[2]})^{-1}\big)$

4: Compute $\widehat{\Delta}_{\mathbf{A}}^{(1)}$ via (S3)

5: Repeat Steps 1–4 with the roles of $\mathcal{D}_1$ and $\mathcal{D}_2$ exchanged, and obtain $\widehat{\Delta}_{\mathbf{A}}^{(2)}$

6: **return** $\widehat{\Delta}_{\mathbf{A}} = (\widehat{\Delta}_{\mathbf{A}}^{(1)} + \widehat{\Delta}_{\mathbf{A}}^{(2)})/2$

Define $\widehat{V} = \widehat{\sigma}^2\big\|(\widehat{\boldsymbol{\Sigma}}^{[1]})^{1/2}\widehat{\Phi}(\mathbf{A})(\widehat{\boldsymbol{\Sigma}}^{[2]})^{1/2}\big\|_F^2$ with $\widehat{\sigma}^2 = N^{-1}\sum_{i\in\mathcal{D}_1}\big(y_i - \langle \boldsymbol{X}_i, \widehat{\mathbf{T}}^{(2)}\rangle\big)^2 + N^{-1}\sum_{i\in\mathcal{D}_2}\big(y_i - \langle \boldsymbol{X}_i, \widehat{\mathbf{T}}^{(1)}\rangle\big)^2$. The usual Wald test and confidence interval are asymptotically valid by the following corollary.

Corollary 1 (Valid Wald inference in the matrix case)**.** Under the matrix-case specialization $K = 2$ in Corollary 2, $\sqrt{N}\,\widehat{V}^{-1/2}\big(\widehat{\Delta}_{\mathbf{A}} - \Delta_{\mathbf{A}}\big) \xrightarrow{d} N(0,1)$.

S2. Propositions

Proof of Proposition 1. Proof of (b). Note that $T_{\Theta^*}\mathcal{M}$ is a linear subspace. Let $U = P_{\mathbf{\Sigma},\Theta^*}(A)$ and $V = P_{\mathbf{\Sigma},\Theta^*}(B)$ such that U and $V \in T_{\Theta^*}\mathcal{M}$, and hence $U+V \in T_{\Theta^*}\mathcal{M}$. Also, by the characterization of orthogonal projections, for all $H \in T_{\Theta^*}\mathcal{M}$,

$$\langle A - U, H\rangle_{\mathbf{\Sigma}} = 0, \qquad \langle B - V, H\rangle_{\mathbf{\Sigma}} = 0.$$

Adding the two identities yields

$$\langle (A + B) - (U + V), H\rangle_{\mathbf{\Sigma}} = 0, \ \forall H \in T_{\Theta^*}\mathcal{M}.$$

By uniqueness of the orthogonal projection, $P_{\mathbf{\Sigma},\Theta^*}(A + B) = U + V = P_{\mathbf{\Sigma},\Theta^*}(A) + P_{\mathbf{\Sigma},\Theta^*}(B)$. *Proof of (c).* By the Theorem 7.3.3 of Horn and Johnson (2013),

$$\langle P_{\mathbf{\Sigma},\Theta^*}(A), B\rangle_{\mathbf{\Sigma}} = \langle A, P_{\mathbf{\Sigma},\Theta^*}(B)\rangle_{\mathbf{\Sigma}},$$

Let $\boldsymbol{a} = \mathrm{vec}(A)$ and $\boldsymbol{b} = \mathrm{vec}(B)$, according to the definition of $P^*_{\mathbf{\Sigma},\Theta^*}$,

$$\begin{aligned}
\langle A, P^*_{\mathbf{\Sigma},\Theta^*}(B)\rangle =& \boldsymbol{a}^\top \mathbf{\Sigma} \mathrm{vec}\big(P_{\mathbf{\Sigma},\Theta^*}(\mathrm{vec}^{-1}(\mathbf{\Sigma}^{-1}\boldsymbol{b}))\big) \\
=& \langle A, P_{\mathbf{\Sigma},\Theta^*}(\mathrm{vec}^{-1}(\mathbf{\Sigma}^{-1}\boldsymbol{b}))\rangle_{\mathbf{\Sigma}} \\
=& \langle P_{\mathbf{\Sigma},\Theta^*}(A), \mathrm{vec}^{-1}(\mathbf{\Sigma}^{-1}\boldsymbol{b})\rangle_{\mathbf{\Sigma}} \\
=& \langle P_{\mathbf{\Sigma},\Theta^*}(A), B\rangle.
\end{aligned}$$

□

Proof of (2.9) *and* (2.10). Recall the tangent space representation in (2.8). For each $\mathcal{A} \in T_{\mathcal{T}}\mathcal{M}$, define $\tilde{\mathcal{A}} = \mathcal{A} \times_1 (\mathbf{\Sigma}^{[1]})^{1/2} \times_2 \cdots \times_K (\mathbf{\Sigma}^{[K]})^{1/2} \in T_{\mathcal{T}}^{\mathbf{\Sigma}}\mathcal{M}$ with

$$\begin{aligned} T_{\mathcal{T}}^{\mathbf{\Sigma}}\mathcal{M} =& \Big\{\dot{\mathcal{G}} \times_{k=1}^{K} (\mathbf{\Sigma}^{[k]})^{1/2}\mathbf{U}_k + \sum_{k=1}^{K} \mathcal{G} \times_k (\mathbf{\Sigma}^{[k]})^{1/2}\dot{\mathbf{U}}_k \times_{j\neq k} (\mathbf{\Sigma}^{[j]})^{1/2}\mathbf{U}_j : \dot{\mathbf{U}}_j^\top \mathbf{U}_j = \mathbf{0}\Big\} \\ =& \Big\{\dot{\mathcal{G}} \times_{k=1}^{K} (\mathbf{\Sigma}^{[k]})^{1/2}\mathbf{U}_k + \sum_{k=1}^{K} \mathcal{G} \times_k \ddot{\mathbf{U}}_k \times_{j\neq k} (\mathbf{\Sigma}^{[j]})^{1/2}\mathbf{U}_j : \ddot{\mathbf{U}}_j^\top (\mathbf{\Sigma}^{[j]})^{1/2}\mathbf{U}_j = \mathbf{0}\Big\}. \end{aligned}$$

Thus, $T_{\mathcal{T}}^{\mathbf{\Sigma}}\mathcal{M} = T_{\tilde{\mathcal{T}}}\mathcal{M}$ for $\tilde{\mathcal{T}} = \mathcal{G} \times_{k=1}^{K} (\mathbf{\Sigma}^{[k]})^{1/2}\boldsymbol{U}_k$. Besides, Lemma 3.1 of Koch and Lubich (2010) yield the following projection operator under Frobenius-norm measure

$$P_{\boldsymbol{I},\mathcal{T}}(\mathcal{A}) = \mathcal{A} \times_{k=1}^{K} \mathbf{P}_{\boldsymbol{U}_k} + \sum_{k=1}^{K} \mathrm{Mat}_k^{-1}(\mathbf{P}_{\boldsymbol{U}_k}^{\perp}\mathbf{A}_k\mathbf{P}_{\check{\boldsymbol{W}}_k}),\ \forall \mathcal{A} \in \mathbb{R}^{p_1\times\cdots\times p_K},$$

where $\check{\boldsymbol{W}}_k = (\boldsymbol{U}_K \otimes \cdots \otimes \boldsymbol{U}_1)\mathrm{QR}(\boldsymbol{G}_k^\top)$. Note that under the Kronecker structure,

$$\begin{aligned} & P_{\mathbf{\Sigma},\mathcal{T}}(\mathcal{A}) \\ =& \arg\min_{\mathcal{B}\in T_{\mathcal{T}}\mathcal{M}} \|\mathcal{A} - \mathcal{B}\|_{\mathbf{\Sigma}}^2 \\ =& \arg\min_{\mathcal{B}\in T_{\mathcal{T}}\mathcal{M}} \|\mathcal{A} \times_1 (\mathbf{\Sigma}^{[1]})^{1/2} \times_2 \cdots \times_K (\mathbf{\Sigma}^{[K]})^{1/2} - \mathcal{B} \times_1 (\mathbf{\Sigma}^{[1]})^{1/2} \times_2 \cdots \times_K (\mathbf{\Sigma}^{[K]})^{1/2}\|^2 \\ =& \Big\{\arg\min_{\mathcal{B}\in T_{\tilde{\mathcal{T}}}\mathcal{M}} \|\mathcal{A} \times_1 (\mathbf{\Sigma}^{[1]})^{1/2} \times_2 \cdots \times_K (\mathbf{\Sigma}^{[K]})^{1/2} - \mathcal{B}\|^2\Big\} \times_1 (\mathbf{\Sigma}^{[1]})^{-1/2} \times_2 \cdots \times_K (\mathbf{\Sigma}^{[K]})^{-1/2} \\ =& P_{\boldsymbol{I},\tilde{\mathcal{T}}}(\tilde{\mathcal{A}}) \times_1 (\mathbf{\Sigma}^{[1]})^{-1/2} \times_2 \cdots \times_K (\mathbf{\Sigma}^{[K]})^{-1/2}, \end{aligned}$$

where $\tilde{\mathcal{A}} = \mathcal{A} \times_1 (\mathbf{\Sigma}^{[1]})^{1/2} \times_2 \cdots \times_K (\mathbf{\Sigma}^{[K]})^{1/2}$. Write $\tilde{\mathcal{T}} = \tilde{\mathcal{G}} \times_{k=1}^{K} \tilde{\boldsymbol{U}}_k$ with $\mathbf{P}_{\tilde{\boldsymbol{U}}_k} = \mathbf{P}_{(\mathbf{\Sigma}^{[k]})^{1/2}\boldsymbol{U}_k}$ and $\{\tilde{\boldsymbol{U}}_k\}_{k=1}^{K}$ are matrices with orthogonal columns. Therefore,

$$P_{\mathbf{\Sigma},\mathcal{T}}(\mathcal{A}) = \tilde{\mathcal{A}}\times_{k=1}^{K}(\mathbf{\Sigma}^{[k]})^{-1/2}\mathbf{P}_{\tilde{\boldsymbol{U}}_k} + \sum_{k=1}^{K} \mathrm{Mat}_k^{-1}(\mathbf{P}_{\tilde{\boldsymbol{U}}_k}^{\perp}\tilde{\mathbf{A}}_k\mathbf{P}_{\bar{\boldsymbol{W}}_k})\times_k(\mathbf{\Sigma}^{[k]})^{-1/2},\ \forall \mathcal{A} \in \mathbb{R}^{p_1\times\cdots\times p_K},$$

where $\bar{\boldsymbol{W}}_k = (\otimes_{j\neq k}\tilde{\boldsymbol{U}}_j)\mathrm{QR}(\tilde{\boldsymbol{G}}_k^\top)$. Note that

$$\begin{aligned}\mathbf{P}_{\bar{\boldsymbol{W}}_k} =& \bar{\boldsymbol{W}}_k(\bar{\boldsymbol{W}}_k^\top\bar{\boldsymbol{W}}_k)^{-1}\bar{\boldsymbol{W}}_k^\top\\ =&(\otimes_{j\neq k}\tilde{\boldsymbol{U}}_j)\mathrm{QR}(\tilde{\boldsymbol{G}}_k^\top)\big\{\mathrm{QR}(\tilde{\boldsymbol{G}}_k^\top)^\top\mathrm{QR}(\tilde{\boldsymbol{G}}_k^\top)\big\}^{-1}\mathrm{QR}(\tilde{\boldsymbol{G}}_k^\top)^\top(\otimes_{j\neq k}\tilde{\boldsymbol{U}}_j^\top)\\ =&(\otimes_{j\neq k}\tilde{\boldsymbol{U}}_j)\tilde{\boldsymbol{G}}_k^\top(\tilde{\boldsymbol{G}}_k\tilde{\boldsymbol{G}}_k^\top)^{-1}\tilde{\boldsymbol{G}}_k(\otimes_{j\neq k}\tilde{\boldsymbol{U}}_j^\top).\end{aligned}$$

Combined with the fact that $\tilde{\boldsymbol{G}}_k(\otimes_{j\neq k}\tilde{\boldsymbol{U}}_j^\top) = \tilde{\boldsymbol{U}}_k^\top\tilde{\mathbf{T}}_k$ and $\tilde{\boldsymbol{G}}_k = \tilde{\boldsymbol{U}}_k^\top\tilde{\mathbf{T}}_k(\otimes_{j\neq k}\tilde{\boldsymbol{U}}_j)$, we have

$$\begin{aligned}&\mathbf{P}_{\bar{\boldsymbol{W}}_k}\\ =&\tilde{\mathbf{T}}_k^\top\tilde{\boldsymbol{U}}_k\big\{\tilde{\boldsymbol{U}}_k^\top\tilde{\mathbf{T}}_k(\otimes_{j\neq k}\mathbf{P}_{\tilde{\boldsymbol{U}}_j})\tilde{\mathbf{T}}_k^\top\tilde{\boldsymbol{U}}_k\big\}^{-1}\tilde{\boldsymbol{U}}_k^\top\tilde{\mathbf{T}}_k\\ =&\tilde{\mathbf{T}}_k^\top(\boldsymbol{\Sigma}^{[k]})^{1/2}\boldsymbol{U}_k\big\{\boldsymbol{U}_k^\top(\boldsymbol{\Sigma}^{[k]})^{1/2}\tilde{\mathbf{T}}_k(\otimes_{j\neq k}\mathbf{P}_{\tilde{\boldsymbol{U}}_j})\tilde{\mathbf{T}}_k^\top(\boldsymbol{\Sigma}^{[k]})^{1/2}\boldsymbol{U}_k\big\}^{-1}(\boldsymbol{\Sigma}^{[k]})^{1/2}\boldsymbol{U}_k^\top\tilde{\mathbf{T}}_k\\ =&(\otimes_{j\neq k}(\boldsymbol{\Sigma}^{[j]})^{1/2}\boldsymbol{U}_j)\boldsymbol{G}_k^\top\big\{\boldsymbol{G}_k(\otimes_{j\neq k}\boldsymbol{U}_j^\top\boldsymbol{\Sigma}^{[j]}\boldsymbol{U}_j)\boldsymbol{G}_k^\top\big\}^{-1}\boldsymbol{G}_k(\otimes_{j\neq k}(\boldsymbol{\Sigma}^{[j]})^{1/2}\boldsymbol{U}_j)^\top,\end{aligned}$$

where the last equality holds since

$$\begin{aligned}&\boldsymbol{U}_k^\top(\boldsymbol{\Sigma}^{[k]})^{1/2}\tilde{\mathbf{T}}_k(\otimes_{j\neq k}\mathbf{P}_{\tilde{\boldsymbol{U}}_j})\tilde{\mathbf{T}}_k^\top(\boldsymbol{\Sigma}^{[k]})^{1/2}\boldsymbol{U}_k\\ =&\boldsymbol{U}_k^\top\boldsymbol{\Sigma}^{[k]}\boldsymbol{U}_k\boldsymbol{G}_k(\otimes_{j\neq k}(\boldsymbol{\Sigma}^{[j]})^{1/2}\boldsymbol{U}_j)^\top(\otimes_{j\neq k}\mathbf{P}_{\tilde{\boldsymbol{U}}_j})(\otimes_{j\neq k}(\boldsymbol{\Sigma}^{[j]})^{1/2}\boldsymbol{U}_j)\boldsymbol{G}_k^\top\boldsymbol{U}_k^\top\boldsymbol{\Sigma}^{[k]}\boldsymbol{U}_k\\ =&\boldsymbol{U}_k^\top\boldsymbol{\Sigma}^{[k]}\boldsymbol{U}_k\boldsymbol{G}_k(\otimes_{j\neq k}(\boldsymbol{\Sigma}^{[j]})^{1/2}\boldsymbol{U}_j)^\top(\otimes_{j\neq k}\mathbf{P}_{\tilde{\boldsymbol{U}}_j})(\otimes_{j\neq k}(\boldsymbol{\Sigma}^{[j]})^{1/2}\boldsymbol{U}_j)\boldsymbol{G}_k^\top\boldsymbol{U}_k^\top\boldsymbol{\Sigma}^{[k]}\boldsymbol{U}_k\\ =&\boldsymbol{U}_k^\top\boldsymbol{\Sigma}^{[k]}\boldsymbol{U}_k\boldsymbol{G}_k(\otimes_{j\neq k}\boldsymbol{U}_j^\top\boldsymbol{\Sigma}^{[j]}\boldsymbol{U}_j)\boldsymbol{G}_k^\top\boldsymbol{U}_k^\top\boldsymbol{\Sigma}^{[k]}\boldsymbol{U}_k\end{aligned}$$

Also, $\mathbf{P}_{\tilde{\boldsymbol{U}}_k}^\perp = \boldsymbol{I} - \mathbf{P}_{\tilde{\boldsymbol{U}}_k} = (\boldsymbol{\Sigma}^{[k]})^{1/2}(\boldsymbol{I}_{p_k} - \mathbf{P}_{\boldsymbol{U}_k}^{\boldsymbol{\Sigma}^{[k]}}\boldsymbol{\Sigma}^{[k]})(\boldsymbol{\Sigma}^{[k]})^{-1/2}$. Therefore,

$$\begin{aligned}P_{\boldsymbol{\Sigma},\mathcal{T}}(\mathcal{A}) =&\tilde{\mathcal{A}}\times_{k=1}^K(\boldsymbol{\Sigma}^{[k]})^{-1/2}\mathbf{P}_{\tilde{\boldsymbol{U}}_k} + \sum_{k=1}^K \mathrm{Mat}_k^{-1}(\mathbf{P}_{\tilde{\boldsymbol{U}}_k}^\perp\tilde{\mathbf{A}}_k\mathbf{P}_{\bar{\boldsymbol{W}}_k})\times_k(\boldsymbol{\Sigma}^{[k]})^{-1/2},\\ =&\mathcal{A}\times_{k=1}^K\mathbf{P}_{\boldsymbol{U}_k}^{\boldsymbol{\Sigma}^{[k]}}\boldsymbol{\Sigma}^{[k]} + \sum_{k=1}^K\mathcal{G}\times_k\boldsymbol{W}_{j1}\times_{j\neq k}\boldsymbol{U}_j,\end{aligned}$$

where

$$\boldsymbol{W}_{j1} = (\boldsymbol{I}_{p_k} - \mathbf{P}_{\boldsymbol{U}_k}^{\boldsymbol{\Sigma}^{[k]}} \boldsymbol{\Sigma}^{[k]}) \mathbf{A}_k (\otimes_{j\neq k} \boldsymbol{\Sigma}^{[j]} \boldsymbol{U}_j) \boldsymbol{G}_k^\top \{ \boldsymbol{G}_k (\otimes_{j\neq k} \boldsymbol{U}_j^\top \boldsymbol{\Sigma}^{[j]} \boldsymbol{U}_j) \boldsymbol{G}_k^\top \}^{-1} \boldsymbol{G}_k (\otimes_{j\neq k} (\boldsymbol{\Sigma}^{[j]})^{1/2} \boldsymbol{U}_j)^\top.$$

By the definition of the adjoint operator,

$$P^*_{\boldsymbol{\Sigma},\mathcal{T}}(\mathcal{A}) = P_{\boldsymbol{\Sigma},\mathcal{T}}(\mathcal{A} \times_k (\boldsymbol{\Sigma}^{[k]})^{-1}) \times_k \boldsymbol{\Sigma}^{[k]} = \mathcal{A} \times_{k=1}^K \boldsymbol{\Sigma}^{[k]} \mathbf{P}_{\boldsymbol{U}_k}^{\boldsymbol{\Sigma}^{[k]}} + \sum_{k=1}^K \mathcal{G} \times_k \boldsymbol{W}_{j2} \times_{j\neq k} \boldsymbol{\Sigma}^{[j]} \boldsymbol{U}_j.$$

$\square$

Proof of Proposition 2. The error decomposition of $\widehat{\Delta}_A^{\text{euc}}$ satisfying

$$\widehat{\Delta}_A^{\text{euc}} - \Delta_A = \frac{1}{N} \sum_{i=1}^N \langle \text{vec}(A), \boldsymbol{\Sigma}^{-1} \text{vec}(Z_i) \rangle \varepsilon_i + \big\langle A - \frac{1}{N} \sum_{i=1}^N \langle \text{vec}(A), \boldsymbol{\Sigma}^{-1} \text{vec}(Z_i) \rangle Z_i, \widehat{\Theta} - \Theta^* \big\rangle.$$

Following the proof of Theorem 1,

$$\widehat{\Delta}_A^{\text{euc}} - \Delta_A = \frac{1}{N} \sum_{i=1}^N \langle \text{vec}(A), \boldsymbol{\Sigma}^{-1} \text{vec}(Z_i) \rangle \varepsilon_i + O_p\big(\frac{1}{\sqrt{N}} \|A\|_F \|\widehat{\Theta} - \Theta^*\|_F \big).$$

The last term is negaligible by the consistency of $\widehat{\Theta}$. Thus, $\widehat{\Delta}_A^{\text{euc}}$ following asymptotic normality with asymptotic variance being

$$\text{Var}(\widehat{\Delta}_A^{\text{euc}}) = \frac{\sigma^2}{N} \|\boldsymbol{\Sigma}^{-1} \text{vec}(A)\|^2_{\boldsymbol{\Sigma}}.$$

We can also verify that $\mathbb{E}(\langle \bar{\Phi}(A), Z_i \rangle Z_i) = P^*_{\boldsymbol{I},\Theta^*}(A)$. Thus, following the proof of Theorem 1, $\widehat{\Delta}_A^{\text{or}}$ is unbiased with asymptotic variance

$$\text{Var}(\widetilde{\Delta}_A^{\text{or}}) = \frac{\sigma^2}{N} \|\bar{\Phi}(A)\|^2_{\boldsymbol{\Sigma}}.$$

Denote the dimensionality of $\text{vec}(A)$ as p for $A \in \mathcal{M}$, and the number of parameter of freedom being d. Since the tangent space $T_\Theta\mathcal{M}$ is a linear space, denote the basis matrix $\boldsymbol{M} \in \mathbb{R}^{p\times d}$ with orthogonal columns. Then for each $B \in T_\Theta\mathcal{M}$, $\text{vec}(B) = \boldsymbol{M}\boldsymbol{u}$ for some $\boldsymbol{u} \in \mathbb{R}^d$. Therefore,

$$P_{\boldsymbol{\Sigma},\Theta^*}(A) = \boldsymbol{M}(\boldsymbol{M}^\top\boldsymbol{\Sigma}\boldsymbol{M})^{-1}\boldsymbol{M}^\top\boldsymbol{\Sigma}\text{vec}(A).$$

This implies that

$$\begin{aligned}
\|\bar{\Phi}(A)\|_{\boldsymbol{\Sigma}}^2 =& \|P_{\boldsymbol{I},\Theta^*}(\mathcal{A})\|_{\boldsymbol{\Sigma}^{-1}}^2 = \text{vec}(A)^\top\boldsymbol{M}\boldsymbol{M}^\top\boldsymbol{\Sigma}^{-1}\boldsymbol{M}\boldsymbol{M}^\top\text{vec}(A) \\
\|\Phi(A)\|_{\boldsymbol{\Sigma}}^2 =& \|P_{\boldsymbol{\Sigma},\Theta^*}(\text{vec}^{-1}(\boldsymbol{\Sigma}^{-1}\text{vec}(A)))\|_{\boldsymbol{\Sigma}}^2 = \text{vec}(A)^\top\boldsymbol{M}(\boldsymbol{M}^\top\boldsymbol{\Sigma}\boldsymbol{M})^{-1}\boldsymbol{M}^\top\text{vec}(A).
\end{aligned}$$

By the inequality that $(\mathbf{P}^\top\boldsymbol{\Sigma}\mathbf{P})^{-1} \lesssim \mathbf{P}^\top\boldsymbol{\Sigma}^{-1}\mathbf{P}$, we have $\text{Var}(\widehat{\Delta}_A^{\text{or}}) \leq \text{Var}(\widetilde{\Delta}_A^{\text{or}}) \leq \text{Var}(\widehat{\Delta}_A^{\text{euc}})$.

$\square$

S3. Proofs of Section 3.1

In the main text, the estimator-specific consistency requirements are incorporated directly into the statements of Theorems 1 and 2 in order to streamline the presentation. For ease of reference in the Supplementary Material , we record these conditions here as Assumptions 1–3 and verify them separately in the matrix and tensor settings.

Assumption 1 (Consistency of $\widehat{\Theta}$)**.** Assume $\widehat{\Theta}$ satisfies $\|\widehat{\Theta} - \Theta^*\|_{\boldsymbol{\Sigma}} = o_p(1)$.

Assumption 2 (Consistency of $\widehat{\mathbf{\Sigma}}$)**.** Assume $\widehat{\mathbf{\Sigma}}$ satisfies $\|\widehat{\mathbf{\Sigma}} - \mathbf{\Sigma}\|_2 = O_p(n^{-1/2})$.

Assumption 3 (Consistency of $\widehat{\Phi}$)**.** Assume $\widehat{\Phi}$ satisfies $\|\widehat{\Phi}(A) - \Phi(A)\|_{\mathbf{\Sigma}} = o_p\big(\|\Phi(A)\|_{\mathbf{\Sigma}}\big)$.

Proof of Theorem 1. According to (2.4),

$$\widehat{\Delta}_A^{\mathrm{or}} - \Delta_A = \frac{1}{N}\sum_{i=1}^{N}\langle \Phi(A), Z_i\rangle \varepsilon_i + \Big\langle P^*_{\mathbf{\Sigma},\Theta^*}(A) - \frac{1}{N}\sum_{i=1}^{N}\langle \Phi(A), Z_i\rangle Z_i, \widehat{\Theta} - \Theta^* \Big\rangle.$$

According to Assumptions 1 and 3, $\langle \Phi(A), Z_i\rangle \varepsilon_i$'s are i.i.d. with mean-zero and $\|\langle \Phi(A), Z_i\rangle \varepsilon_i\|_{\psi_1} \le \|\langle \Phi(A), Z_i\rangle\|_{\psi_2}\|\varepsilon_i\|_{\psi_2} \le C\sigma\kappa\|\Phi(A)\|_{\mathbf{\Sigma}}$. Then we have $\mathbb{E}|\langle \Phi(A), Z_i\rangle \varepsilon_i|^3 \le C(\sigma\kappa\|\Phi(A)\|_{\mathbf{\Sigma}})^3$. Apply the Berry-Esseen Theorem

$$\frac{1}{\sqrt{N}}V^{-1/2}\sum_{i=1}^{N}\langle \Phi(A), Z_i\rangle \varepsilon_i \xrightarrow{d} N(0,1),$$

where $V = \mathrm{Var}(\langle \Phi(A), Z_i\rangle \varepsilon_i) = \sigma^2\|\Phi(A)\|^2_{\mathbf{\Sigma}}$. Note that $\widehat{\Theta}$ is independent with the data samples and $\langle \Phi(A), Z_i\rangle\langle Z_i, \widehat{\Theta} - \Theta^*\rangle$'s are i.i.d. and $\mathbb{E}\big[\langle \Phi(A), Z_i\rangle\langle Z_i, \widehat{\Theta} - \Theta^*\rangle\big] = \langle P^*_{\mathbf{\Sigma},\Theta^*}(A), \widehat{\Theta} - \Theta^*\rangle$, such that $\|\langle P^*_{\mathbf{\Sigma},\Theta^*}(A) - \langle \Phi(A), Z_i\rangle Z_i, \widehat{\Theta} - \Theta^*\rangle\|_{\psi_1} \le \|\langle \Phi(A), Z_i\rangle\langle Z_i, \widehat{\Theta} - \Theta^*\rangle\|_{\psi_1} \le C\|\Phi(A)\|_{\mathbf{\Sigma}}\|\widehat{\Theta} - \Theta^*\|_{\mathbf{\Sigma}}$. Apply Bernstein's inequality,

$$P\Big(\Big|\Big\langle P^*_{\mathbf{\Sigma},\Theta^*}(A) - \frac{1}{N}\sum_{i=1}^{N}\langle \Phi(A), Z_i\rangle Z_i, \widehat{\Theta} - \Theta^*\Big\rangle\Big| \ge C\kappa^2\|\Phi(A)\|_{\mathbf{\Sigma}}\|\widehat{\Theta} - \Theta^*\|_{\mathbf{\Sigma}}\Big(\sqrt{\frac{t}{N}} + \frac{t}{N}\Big)\Big) \le C\exp(-t).$$

This implies that

$$\Big|\Big\langle P^*_{\mathbf{\Sigma},\Theta^*}(A) - \frac{1}{N}\sum_{i=1}^{N}\langle \Phi(A), Z_i\rangle Z_i, \widehat{\Theta} - \Theta^*\Big\rangle\Big| = O_p(N^{-1/2}\kappa^2\|\Phi(A)\|_{\mathbf{\Sigma}}\|\widehat{\Theta} - \Theta^*\|_{\mathbf{\Sigma}}).$$

Thus, when $\|\widehat{\Theta} - \Theta^*\|_{\mathbf{\Sigma}} = o_p(1)$, $\widehat{\Delta}_A^{\mathrm{or}} - \Delta_A$ is dominated by the first term and

$$\sqrt{N}V^{-1/2}(\widehat{\Delta}_A^{\mathrm{or}} - \Delta_A) \xrightarrow{d} N(0,1).$$

□

Proof of Theorem 2. Recall the error decomposition in (2.6),

$$\widehat{\Delta}_A^{\rm fe}-\Delta_A=\frac{1}{N}\sum_{i=1}^N\langle\widehat{\Phi}(A),Z_i\rangle\varepsilon_i+\Big\langle P^*_{\widehat{\mathbf{\Sigma}},\widehat{\Theta}}(A)-\frac{1}{N}\sum_{i=1}^N\langle\widehat{\Phi}(A),Z_i\rangle Z_i,\widehat{\Theta}-\Theta^*\Big\rangle+\langle P^*_{\widehat{\mathbf{\Sigma}},\widehat{\Theta}}(A)-A,\Theta^*\rangle.$$

The first term of (2.6) can be decomposed into

$$\frac{1}{N}\sum_{i=1}^N\langle\Phi(A),Z_i\rangle\varepsilon_i+\frac{1}{N}\sum_{i=1}^N\langle\widehat{\Phi}(A)-\Phi(A),Z_i\rangle\varepsilon_i.$$

Note that $\langle\widehat{\Phi}(A)-\Phi(A),Z_i\rangle\varepsilon_i$'s are i.i.d. with mean-zero, and $\|\langle\widehat{\Phi}(A)-\Phi(A),Z_i\rangle\varepsilon_i\|_{\psi_1}\le\|\varepsilon_i\|_{\psi_2}\|\langle\widehat{\Phi}(A)-\Phi(A),Z_i\rangle\|_{\psi_2}\le C\sigma\kappa\|\widehat{\Phi}(A)-\Phi(A)\|_{\mathbf{\Sigma}}$. Apply Bernstein's inequality,

$$P\Big(\Big|\frac{1}{N}\sum_{i=1}^N\langle\widehat{\Phi}(A)-\Phi(A),Z_i\rangle\varepsilon_i\Big|\ge C\sigma\kappa\|\widehat{\Phi}(A)-\Phi(A)\|_{\mathbf{\Sigma}}\Big(\sqrt{\frac{t}{N}}+\frac{t}{N}\Big)\Big)\le C\exp(-t).$$

Thus,

$$N^{-1}\sum_{i=1}^N\langle\widehat{\Phi}(A)-\Phi(A),Z_i\rangle\varepsilon_i=O_p(N^{-1/2}\sigma\kappa\|\widehat{\Phi}(A)-\Phi(A)\|_{\mathbf{\Sigma}}).$$

Next we analysis the second term, which can be rewritten as follows:

$$\langle P^*_{\widehat{\mathbf{\Sigma}},\widehat{\Theta}}(A)-\widetilde{P}^*_{\widehat{\mathbf{\Sigma}},\widehat{\Theta}}(A),\widehat{\Theta}-\Theta^*\rangle+\Big\langle\widetilde{P}^*_{\widehat{\mathbf{\Sigma}},\widehat{\Theta}}(A)-\frac{1}{N}\sum_{i=1}^N\langle\widehat{\Phi}(A),Z_i\rangle Z_i,\widehat{\Theta}-\Theta^*\Big\rangle:=\xi_{N1}+\xi_{N2}.$$

Note that $\widetilde{P}^*_{\widehat{\mathbf{\Sigma}},\widehat{\Theta}}(A)={\rm vec}^{-1}\big(\mathbf{\Sigma}\,{\rm vec}(\widehat{\Phi}(A))\big)$ and $P^*_{\widehat{\mathbf{\Sigma}},\widehat{\Theta}}(A)={\rm vec}^{-1}\big(\widehat{\mathbf{\Sigma}}\,{\rm vec}(\widehat{\Phi}(A))\big)$. Then

$$|\xi_{N1}|=|\langle(\widehat{\mathbf{\Sigma}}-\mathbf{\Sigma}){\rm vec}(\widehat{\Phi}(A)),{\rm vec}(\widehat{\Theta}-\Theta^*)\rangle|\le\|\widehat{\Phi}(A)\|_{\mathbf{\Sigma}}\|\widehat{\Theta}-\Theta^*\|_{\mathbf{\Sigma}}\|\mathbf{\Sigma}^{-1/2}(\widehat{\mathbf{\Sigma}}-\mathbf{\Sigma})\mathbf{\Sigma}^{-1/2}\|_2.$$

Besides, similar with the proof of Theorem 1, $\xi_{N2}=O_p(N^{-1/2}\kappa^2\|\Phi(A)\|_{\mathbf{\Sigma}}\|\widehat{\Theta}-\Theta^*\|_{\mathbf{\Sigma}})$.

Under Assumptions 1-3,

$$\widehat{\Delta}_A^{\rm fe}-\Delta_A=\frac{1}{N}\sum_{i=1}^N\langle\Phi(A),Z_i\rangle\varepsilon_i+\langle P^*_{\widehat{\mathbf{\Sigma}},\widehat{\Theta}}(A)-\widetilde{P}^*_{\widehat{\mathbf{\Sigma}},\widehat{\Theta}}(A),\widehat{\Theta}-\Theta^*\rangle+\langle P^*_{\widehat{\mathbf{\Sigma}},\widehat{\Theta}}(A)-A,\Theta^*\rangle.$$

Following the proof of Theorem 1, when $\langle P^*_{\widehat{\mathbf{\Sigma}},\widehat{\Theta}}(A) - A, \Theta^* \rangle = o_p(N^{-1/2}\sigma\|\Phi(A)\|_{\mathbf{\Sigma}})$,

$$N^{1/2}V^{-1/2}(\widehat{\Delta}^{\text{fe}}_A - \Delta_A) \xrightarrow{d} N(0,1).$$

$\square$

Proof of Theorem 3. Let

$$H := \Phi(A) = P_{\mathbf{\Sigma},\Theta^*}\big(\text{vec}^{-1}(\mathbf{\Sigma}^{-1}\text{vec}(A))\big).$$

If $H = 0$, then the claimed lower bound is immediate. We therefore consider only the case $H \neq 0$. Since $\mathcal{M}$ is a C^1 embedded submanifold and $H \in T_{\Theta^*}\mathcal{M}$, there exists a C^1 curve $\Theta(t) \in \mathcal{M}$ for $|t| < t_0$ such that $\Theta(0) = \Theta^*$ and $\Theta'(0) = H$. Because $\mathcal{U}$ contains a relative neighborhood of Θ^* in $\mathcal{M}$, there exists $t_1 > 0$ such that $\Theta(t) \in \mathcal{U}$ for $|t| < t_1$. By differentiability,

$$\Theta(t) - \Theta^* = tH + r(t), \qquad \|r(t)\|_F = o(|t|).$$

Since all norms are equivalent in finite dimensions and $\mathbf{\Sigma}$ is positive definite, this also implies $\|r(t)\|_{\mathbf{\Sigma}} = o(|t|)$. Hence

$$\|\Theta(t) - \Theta^*\|_{\mathbf{\Sigma}} = |t|\,\|H\|_{\mathbf{\Sigma}} + o(|t|\,\|H\|_{\mathbf{\Sigma}}). \tag{S4}$$

By definition, H is the $\mathbf{\Sigma}$-orthogonal projection of $\text{vec}^{-1}(\mathbf{\Sigma}^{-1}\text{vec}(A))$ onto $T_{\Theta^*}\mathcal{M}$. Thus, for any $G \in T_{\Theta^*}\mathcal{M}$,

$$\Big\langle \text{vec}^{-1}(\mathbf{\Sigma}^{-1}\text{vec}(A)) - H,\ G \Big\rangle_{\mathbf{\Sigma}} = 0.$$

Taking $G = H$ yields

$$\langle A, H\rangle = \left\langle \mathrm{vec}^{-1}(\mathbf{\Sigma}^{-1}\mathrm{vec}(A)), H \right\rangle_{\mathbf{\Sigma}} = \|H\|_{\mathbf{\Sigma}}^2.$$

Therefore

$$\Delta_A(\Theta(t)) - \Delta_A(\Theta^*) = \langle A, \Theta(t) - \Theta^*\rangle = t\langle A, H\rangle + \langle A, r(t)\rangle = t\|H\|_{\mathbf{\Sigma}}^2 + o(|t|\,\|H\|_{\mathbf{\Sigma}}).$$

Let P_Θ denote the distribution of $\mathcal{D}$ under parameter Θ. For any $\Theta_1 \in \mathcal{U}$ and any $CI \in \mathcal{I}_\alpha(\mathcal{U}, A)$, the standard two-point testing argument implies

$$\max\Big\{\mathbb{E}_{\Theta^*} L(CI),\ \mathbb{E}_{\Theta_1} L(CI)\Big\} \geq |\Delta_A(\Theta_1) - \Delta_A(\Theta^*)|\Big(1 - 2\alpha - \mathrm{TV}(P_{\Theta^*}, P_{\Theta_1})\Big). \quad \text{(S5)}$$

Under (1.1) with Gaussian noise, conditional on $\{Z_i\}_{i=1}^N$, the responses are Gaussian with means shifted by $\langle Z_i, \Theta_1 - \Theta^*\rangle$ and common variance σ^2. Using the definition of $\mathbf{\Sigma}$,

$$\mathbb{E}\langle \mathcal{X}_i, \Theta_1 - \Theta^*\rangle^2 = \mathrm{vec}(\Theta_1 - \Theta^*)^\top \mathbf{\Sigma}\,\mathrm{vec}(\Theta_1 - \Theta^*) = \|\Theta_1 - \Theta^*\|_{\mathbf{\Sigma}}^2.$$

Hence

$$D_{\mathrm{KL}}(P_{\Theta^*}\|P_{\Theta_1}) = \frac{N}{2\sigma^2}\|\Theta_1 - \Theta^*\|_{\mathbf{\Sigma}}^2.$$

By Pinsker's inequality,

$$\mathrm{TV}(P_{\Theta^*}, P_{\Theta_1}) \leq \sqrt{\frac{1}{2} D_{\mathrm{KL}}(P_{\Theta^*}\|P_{\Theta_1})}. \quad \text{(S6)}$$

Set

$$t_* = a\,\frac{\sigma}{\sqrt{N}\,\|H\|_{\mathbf{\Sigma}}},$$

where $a > 0$ is a sufficiently small constant. For all large N, we have $|t_*| < t_1$ and $\Theta(t_*) \in \mathcal{U}$. Using (S4),

$$D_{\mathrm{KL}}(P_{\Theta^*} \| P_{\Theta(t_*)}) = \frac{N}{2\sigma^2} \|\Theta(t_*) - \Theta^*\|_{\boldsymbol{\Sigma}}^2 = \frac{a^2}{2} + o(1).$$

Therefore, by (S6),

$$\mathrm{TV}(P_{\Theta^*}, P_{\Theta(t_*)}) \leq \frac{a}{2} + o(1).$$

Choosing a small enough ensures $1 - 2\alpha - \mathrm{TV}(P_{\Theta^*}, P_{\Theta(t_*)}) \geq c_1 > 0$ for some constant c_1 depending only on α. On the other hand, (S5) gives

$$|\Delta_A(\Theta(t_*)) - \Delta_A(\Theta^*)| = a \, \frac{\sigma}{\sqrt{N}} \, \|H\|_{\boldsymbol{\Sigma}} + o\Big(\frac{\sigma}{\sqrt{N}} \, \|H\|_{\boldsymbol{\Sigma}} \Big).$$

Substituting $\Theta_1 = \Theta(t_*)$ into (S5), we obtain

$$\max \Big\{ \mathbb{E}_{\Theta^*} L(CI), \ \mathbb{E}_{\Theta(t_*)} L(CI) \Big\} \geq c_2 \, \frac{\sigma}{\sqrt{N}} \, \|H\|_{\boldsymbol{\Sigma}},$$

for some constant $c_2 > 0$ depending only on α. Since both Θ^* and $\Theta(t_*)$ belong to $\mathcal{U}$,

$$\sup_{\Theta \in \mathcal{U}} \mathbb{E}_{\Theta} L(CI) \geq c_2 \, \frac{\sigma}{\sqrt{N}} \, \|\Phi(A)\|_{\boldsymbol{\Sigma}}.$$

Taking the infimum over $CI \in \mathcal{I}_\alpha(\mathcal{U}, A)$ completes the proof. $\square$

S4. Additional lemmas

Lemma 1 (Residual-based variance estimator)**.** Assume the regression model 1.1 hold. Assume $\widehat{\Theta}$ is independent of $\{(Z_i, y_i)\}_{i=1}^N$, and Assumptions 1 and 3 hold. Let

$$\widehat{\sigma}^2 := \frac{1}{N}\sum_{i=1}^N \big(y_i - \langle Z_i, \widehat{\Theta}\rangle\big)^2.$$

Then

$$\widehat{\sigma}^2 = \sigma^2 + O_p\left(\frac{\sigma^2}{\sqrt{N}} + \frac{\kappa\sigma\|\widehat{\Theta} - \Theta\|_{\mathbf{\Sigma}}}{\sqrt{N}} + \frac{\kappa^2\|\widehat{\Theta} - \Theta\|_{\mathbf{\Sigma}}^2}{\sqrt{N}}\right).$$

Proof of Lemma 1. Let $\Delta := \widehat{\Theta} - \Theta$. Using $y_i = \langle Z_i, \Theta\rangle + \varepsilon_i$,

$$y_i - \langle Z_i, \widehat{\Theta}\rangle = \varepsilon_i - \langle Z_i, \Delta\rangle.$$

Hence, expanding the square,

$$\widehat{\sigma}^2 = \frac{1}{N}\sum_{i=1}^N \left(\varepsilon_i^2 - 2\varepsilon_i\langle Z_i, \Delta\rangle + \langle Z_i, \Delta\rangle^2\right) =: T_1 - 2T_2 + T_3.$$

We bound $T_1 - \sigma^2$, T_2, and $T_3 - \|\Delta\|_{\mathbf{\Sigma}}^2$, separately.

By Assumption 1, ε_i are mean-zero sub-Gaussian with variance σ^2. It is standard that $\varepsilon_i^2 - \mathbb{E}\varepsilon_i^2$ is sub-exponential and $\|\varepsilon_i^2 - \sigma^2\|_{\psi_1} \le C\sigma^2$ for a universal C. Applying Bernstein's inequality for sub-exponential variables, for any $t > 0$, with probability at least $1 - 2\exp(-ct^2)$,

$$|T_1 - \sigma^2| \le C\sigma^2\sqrt{\frac{1}{N}}\,t. \tag{S7}$$

Next we consider T_2. Because $\widehat{\Theta}$ is independent of the sample, Δ is independent of $\{(Z_i, \varepsilon_i)\}_{i=1}^N$. In particular, conditioning on Δ is valid and fixed. Moreover, $\mathbb{E}[\varepsilon_i \mid$

$Z_i] = 0$ implies $\mathbb{E}[\varepsilon_i \langle Z_i, \Delta\rangle \mid \Delta] = 0$. Let $U_i := \langle Z_i, \Delta\rangle$. Assumption 3 gives, for all $s \in \mathbb{R}$,

$$\mathbb{E}\exp(sU_i) = \mathbb{E}\exp(s\langle Z_i, \Delta\rangle) \le \exp\Big(\frac{\kappa^2 s^2}{2}\|\Delta\|_{\boldsymbol{\Sigma}}^2\Big).$$

Thus U_i is sub-Gaussian with $\|U_i\|_{\psi_2} \le C\kappa\|\Delta\|_{\boldsymbol{\Sigma}}$. Also Assumption 1 implies $\|\varepsilon_i\|_{\psi_2} \le C\sigma$. Define $W_i := \varepsilon_i U_i$. A standard product rule gives sub-exponentiality:

$$\|W_i\|_{\psi_1} \le C\|\varepsilon_i\|_{\psi_2}\|U_i\|_{\psi_2} \le C(\sigma)(\kappa\|\Delta\|_{\boldsymbol{\Sigma}}) = C\kappa\sigma\|\Delta\|_{\boldsymbol{\Sigma}}.$$

Applying Bernstein to $\{W_i\}$ (conditionally on Δ), for any $t > 0$, with probability at least $1 - 2\exp(-ct^2)$,

$$|T_2| = \left|\frac{1}{N}\sum_{i=1}^{N} W_i\right| \le C\kappa\sigma\|\Delta\|_{\boldsymbol{\Sigma}}\sqrt{\frac{1}{N}}\,t. \tag{S8}$$

Still conditioning on Δ, note $\mathbb{E}[U_i^2 \mid \Delta] = \mathrm{Var}(U_i \mid \Delta) = \|\Delta\|_{\boldsymbol{\Sigma}}^2$. Moreover, since U_i is sub-Gaussian with $\|U_i\|_{\psi_2} \le C\kappa\|\Delta\|_{\boldsymbol{\Sigma}}$, the centered square is sub-exponential with

$$\|U_i^2 - \mathbb{E}U_i^2\|_{\psi_1} \le C\|U_i\|_{\psi_2}^2 \le C\kappa^2\|\Delta\|_{\boldsymbol{\Sigma}}^2.$$

Apply Bernstein to $U_i^2 - \mathbb{E}U_i^2$ to get: for any $t > 0$, with probability at least $1 - 2\exp(-ct^2)$,

$$\big|T_3 - \|\Delta\|_{\boldsymbol{\Sigma}}^2\big| = \left|\frac{1}{N}\sum_{i=1}^{N}(U_i^2 - \mathbb{E}U_i^2)\right| \le C\kappa^2\|\Delta\|_{\boldsymbol{\Sigma}}^2\sqrt{\frac{1}{N}}\,t. \tag{S9}$$

By the decomposition $\widehat{\sigma}^2 - \sigma^2 = (T_1 - \sigma^2) - 2T_2 + (T_3 - \|\Delta\|_{\boldsymbol{\Sigma}}^2)$, combine (S7), (S8), (S9) and take a union bound to conclude that, with probability at least $1 - 6\exp(-ct^2)$,

$$|\widehat{\sigma}^2 - \sigma^2| \le C\Big(\sigma^2\sqrt{\frac{1}{N}}\,t + \kappa\sigma\|\Delta\|_{\boldsymbol{\Sigma}}\sqrt{\frac{1}{N}}\,t + \kappa^2\|\Delta\|_{\boldsymbol{\Sigma}}^2\sqrt{\frac{1}{N}}\,t\Big).$$

This implies the stated $O_p(\cdot)$ rate, and the proof is finished. $\square$

Lemma 2. Assume Assumptions 1–3 hold, and a variance estimate by $\widehat{V} = \widehat{\sigma}^2 \|\widehat{\Phi}(\mathcal{A})\|^2_{\widehat{\boldsymbol{\Sigma}}}$, then $\widehat{V} \xrightarrow{p} V$.

Proof. The proof is straightforward by Lemma 1 and the Slutsky's theorem. $\square$

Lemma 3. Assume Assumption 4 holds with $K = 2$. The normalized estimator defined in (2.12) satisfies

$$\left\|\widehat{\boldsymbol{\Sigma}}^{[j]} - \sqrt{\frac{\operatorname{tr}(\boldsymbol{\Sigma}^{[k]})}{\operatorname{tr}(\boldsymbol{\Sigma}^{[j]})}}\,\boldsymbol{\Sigma}^{[j]}\right\|_2 = O_p\Big(\frac{1}{\sqrt{n}}\Big),\ j,k \in \{1,2\}, k \neq j.$$

Proof of Lemma 3. By the definition of (2.12), it equivalents with proving

$$\left\|\frac{\widetilde{\boldsymbol{\Sigma}}^{[j]}}{\sqrt{\operatorname{tr}(\boldsymbol{\Sigma}^{[j]})\operatorname{tr}(\boldsymbol{\Sigma}^{[k]})}} - \sqrt{\frac{\operatorname{tr}(\boldsymbol{\Sigma}^{[k]})}{\operatorname{tr}(\boldsymbol{\Sigma}^{[j]})}}\,\boldsymbol{\Sigma}^{[j]}\right\|_2 = O_p\Big(\frac{1}{\sqrt{n}}\Big).$$

Take $j = 1, k = 2$ without loss of generality. We first control $\|\widetilde{\boldsymbol{\Sigma}}^{[1]} - \operatorname{tr}(\boldsymbol{\Sigma}^{[2]})\boldsymbol{\Sigma}^{[1]}\|_2$. Note that

$$\|\widetilde{\boldsymbol{\Sigma}}^{[1]} - \operatorname{tr}(\boldsymbol{\Sigma}^{[2]})\boldsymbol{\Sigma}^{[1]}\|_2 = \sup_{\boldsymbol{a}\in\mathbb{S}^{p-1}} \left\{\frac{1}{n}\sum_{i=1}^{n} \boldsymbol{a}^\top \boldsymbol{X}_i \boldsymbol{X}_i^\top \boldsymbol{a} - \operatorname{tr}(\boldsymbol{\Sigma}^{[2]})\,\boldsymbol{a}^\top \boldsymbol{\Sigma}^{[1]}\boldsymbol{a}\right\}.$$

where $\boldsymbol{X}_i = \mathbf{L}_1 \boldsymbol{S}_i \mathbf{L}_2^\top \in \mathbb{R}^{p\times q}$. Hence,

$$\boldsymbol{a}^\top \boldsymbol{X}_i \boldsymbol{X}_i^\top \boldsymbol{a} = \boldsymbol{a}^\top \mathbf{L}_1 \boldsymbol{S}_i \mathbf{L}_2^\top \mathbf{L}_2 \boldsymbol{S}_i^\top \mathbf{L}_1^\top \boldsymbol{a} = \widetilde{\boldsymbol{a}}^\top \boldsymbol{S}_i \boldsymbol{\Sigma}_2 \boldsymbol{S}_i^\top \widetilde{\boldsymbol{a}} = \widetilde{\boldsymbol{s}}_i^\top \boldsymbol{\Sigma}_2\, \widetilde{\boldsymbol{s}}_i,$$

where $\boldsymbol{\Sigma}_2 = \mathbf{L}_2^\top \mathbf{L}_2 \in \mathbb{R}^{q\times q}$, $\widetilde{\boldsymbol{a}} = \mathbf{L}_1^\top \boldsymbol{a}$, and $\widetilde{\boldsymbol{s}}_i = \boldsymbol{S}_i^\top \widetilde{\boldsymbol{a}} \in \mathbb{R}^q$. For each $j \in [q]$, $(\widetilde{\boldsymbol{s}}_i)_j = \sum_{k=1}^{p} (\boldsymbol{S}_i)_{kj} \widetilde{\boldsymbol{a}}_k$ is sub-Gaussian with $\|(\widetilde{\boldsymbol{s}}_i)_j\|_{\psi_2} \leq C\kappa \|\widetilde{\boldsymbol{a}}\|_2$, and the components of $\widetilde{\boldsymbol{s}}_i$

are independent across j. Let $\boldsymbol{s} = (\widetilde{\boldsymbol{s}}_1^\top, \ldots, \widetilde{\boldsymbol{s}}_n^\top)^\top \in \mathbb{R}^{nq}$. Then $\sum_{i=1}^n \boldsymbol{a}^\top \boldsymbol{X}_i \boldsymbol{X}_i^\top \boldsymbol{a} = \boldsymbol{s}^\top(\boldsymbol{\Sigma}_2 \otimes \boldsymbol{I}_n)\boldsymbol{s}$. By Hanson–Wright inequality (Theorem 6.2.1 of Vershynin (2018)), for all $t > 0$,

$$P\Big(\big|\boldsymbol{s}^\top(\boldsymbol{\Sigma}_1 \otimes \boldsymbol{I}_n)\boldsymbol{s} - n\mathrm{tr}(\boldsymbol{\Sigma}^{[2]})\boldsymbol{a}^\top \boldsymbol{\Sigma}^{[1]}\boldsymbol{a}\big| \geq t\Big) \leq 2\exp\left(-c\min\left\{\frac{t^2}{C^4\|\boldsymbol{\Sigma}_1 \otimes \boldsymbol{I}_n\|_F^2}, \frac{t}{C^2\|\boldsymbol{\Sigma}_1 \otimes \boldsymbol{I}_n\|_2}\right\}\right), \tag{S10}$$

where $\|\boldsymbol{\Sigma}_1 \otimes \boldsymbol{I}_n\|_F = \|\boldsymbol{\Sigma}^{[2]}\|_F\|\boldsymbol{I}_n\|_F = O(\sqrt{nq})$ and $\|\boldsymbol{\Sigma}_1 \otimes \boldsymbol{I}_n\|_2 = \|\boldsymbol{\Sigma}^{[2]}\|_2 = O(1)$. Let $\{\boldsymbol{u}_1, \ldots, \boldsymbol{u}_{N_1}\}$ be a 1/8-net of $\mathbb{S}^{p-1}$ with $N_1 \leq 17^p$. By a standard net argument (cf. Lemma 2 in Lu et al. (2023)),

$$\|\widetilde{\boldsymbol{\Sigma}}^{[1]} - \mathrm{tr}(\boldsymbol{\Sigma}^{[2]})\boldsymbol{\Sigma}^{[1]}\|_2 \leq 2\max_{1\leq k\leq N_1}\left|\frac{1}{n}\sum_{i=1}^n \boldsymbol{u}_k^\top \boldsymbol{X}_i \boldsymbol{X}_i^\top \boldsymbol{u}_k - \mathrm{tr}(\boldsymbol{\Sigma}^{[2]})\,\boldsymbol{u}_k^\top \boldsymbol{\Sigma}^{[1]}\boldsymbol{u}_k\right|.$$

Together with (S10) and a union bound, we obtain that for a sufficiently large constant C,

$$P\Big(\|\widetilde{\boldsymbol{\Sigma}}^{[1]} - \mathrm{tr}(\boldsymbol{\Sigma}^{[2]})\boldsymbol{\Sigma}^{[1]}\|_2 \geq C\Big(\sqrt{\frac{pq}{n}} + \frac{p}{n}\Big)\Big) \leq 2\exp(-cp).$$

Since $\sqrt{\mathrm{tr}(\boldsymbol{\Sigma}^{[1]})\mathrm{tr}(\boldsymbol{\Sigma}^{[2]})} = O(\sqrt{pq})$, the claimed $O_p(n^{-1/2})$ bound for $\widetilde{\boldsymbol{\Sigma}}^{[1]}/\sqrt{\mathrm{tr}(\boldsymbol{\Sigma}^{[1]})\mathrm{tr}(\boldsymbol{\Sigma}^{[2]})}$ follows. The bound for $\widetilde{\boldsymbol{\Sigma}}^{[2]}$ is analogous.

Furthermore, by Lemma 11 in Ke et al. (2025), there exists some constant $C > 0$ such that

$$P\Big(\Big|\frac{1}{npq}\sum_{i=1}^n \|\boldsymbol{X}_i\|_F^2 - \frac{1}{pq}\mathrm{tr}(\boldsymbol{\Sigma}^{[1]})\mathrm{tr}(\boldsymbol{\Sigma}^{[2]})\Big| \geq C\Big(\sqrt{\frac{t}{n}} + \frac{t}{n}\Big)\Big) \leq C\exp(-t).$$

This implies

$$\frac{\sum_{i=1}^n \|\boldsymbol{X}_i\|_F^2/(npq)}{\mathrm{tr}(\boldsymbol{\Sigma}^{[1]})\mathrm{tr}(\boldsymbol{\Sigma}^{[2]})/(pq)} = 1 + O_p\Big(\frac{1}{\sqrt{n}}\Big).$$

Combined with the definition of the normalized estimator in (2.12), it yields the desired bounds. $\square$

Lemma 4. Assume Assumption 4 holds. The normalized estimators defined in (2.12) satisfy

$$\Big\|\widehat{\boldsymbol{\Sigma}}^{[j]} - \frac{\mathrm{tr}(\boldsymbol{\Sigma}^{[j+2]})\mathrm{tr}(\boldsymbol{\Sigma}^{[j+1]})}{\{\prod_{j=1}^{3}\mathrm{tr}(\boldsymbol{\Sigma}^{[j]})\}^{2/3}} \cdot \boldsymbol{\Sigma}^{[j]}\Big\|_2 = O_p\Big(\frac{1}{\sqrt{n}}\Big),\ j \in [K].$$

Proof of Lemma 4. Under Assumption 4,

$$\mathrm{Mat}_j(\mathcal{X}_i) = \mathbf{L}_j\,\mathrm{Mat}_j(\mathcal{S}_i)\,\big(\otimes_{k\neq j}\mathbf{L}_k\big)^\top,$$

Define $\boldsymbol{\Sigma}^{[j]} := \mathbf{L}_j\mathbf{L}_j^\top$ and $\boldsymbol{\Sigma}_{-j} := \otimes_{k\neq j}\boldsymbol{\Sigma}^{[k]}$ for $j \in [K]$. Since the entries of $\mathcal{S}_i$ are i.i.d. mean 0, variance 1, and sub-Gaussian with parameter κ (Assumption 4), the entries of $\mathrm{Mat}_j(\mathcal{S}_i)$ inherit the same property. Thus $\mathrm{Mat}_j(\mathcal{X}_i)$ fits the matrix structured sub-Gaussian design with "row covariance" $\check{\boldsymbol{\Sigma}}^{[1]} = \boldsymbol{\Sigma}^{[j]}$ and "column covariance" $\check{\boldsymbol{\Sigma}}^{[2]} = \boldsymbol{\Sigma}_{-j}$. Let $\widetilde{\boldsymbol{\Sigma}}^{[j]}$ be the unnormalized estimator in (2.11). Then

$$\mathbb{E}\big(\widetilde{\boldsymbol{\Sigma}}^{[j]}\big) = \mathrm{tr}(\boldsymbol{\Sigma}_{-j})\,\boldsymbol{\Sigma}^{[j]} = \prod_{k\neq j}\mathrm{tr}(\boldsymbol{\Sigma}^{[k]})\cdot\boldsymbol{\Sigma}^{[j]}.$$

Applying Lemma 3 to the matrix representation yields

$$\Big\|\frac{\widetilde{\boldsymbol{\Sigma}}^{[j]}}{\{\prod_{k=1}^{K}\mathrm{tr}(\boldsymbol{\Sigma}^{[k]})\}^{1/2}} - \frac{\prod_{k\neq j}\mathrm{tr}(\boldsymbol{\Sigma}^{[k]})}{\{\prod_{k=1}^{K}\mathrm{tr}(\boldsymbol{\Sigma}^{[k]})\}^{1/2}}\Big\|_2 = O_p\Big(\frac{1}{\sqrt{n}}\Big).$$

Then

$$\Big\|\frac{\widetilde{\boldsymbol{\Sigma}}^{[j]}}{\{\prod_{k=1}^{K}\mathrm{tr}(\boldsymbol{\Sigma}^{[k]})\}^{(K-1)/K}} - \frac{\prod_{k\neq j}\mathrm{tr}(\boldsymbol{\Sigma}^{[k]})}{\{\prod_{k=1}^{K}\mathrm{tr}(\boldsymbol{\Sigma}^{[k]})\}^{(K-1)/K}}\Big\|_2 = O_p\Big(\frac{1}{\sqrt{n}}(\prod_{k=1}^{K}p_k)^{-(\frac{1}{2}-\frac{1}{K})}\Big) = O_p\Big(\frac{1}{\sqrt{n}}\Big).$$

Finally, by Lemma 11 in Ke et al. (2025),

$$\frac{1}{n}\sum_{i=1}^{n}\|\mathcal{X}_i\|_F^2 \xrightarrow{p} \prod_{k=1}^{K}\mathrm{tr}(\mathbf{\Sigma}^{[k]}) = O_p(\prod_{k=1}^{K}p_k),$$

Combining the above displays with Slutsky's theorem gives the claim. $\square$

To apply Theorem 2, it remains to control the projection-mismatch term $\langle P^*_{\widehat{\mathbf{\Sigma}},\widehat{\Theta}}(A)-A,\Theta^*\rangle$. The next lemma provide sufficient bounds in the tensor linear regression setting.

Lemma 5 (Projection mismatch control1)**.** Assume the low-Tucker-rank tensor model holds. Assume Assumptions 1–3, 5 and 7 hold. If, in addition,

$$\lambda^{-2}\|\widehat{\mathcal{T}}-\mathcal{T}^*\|_F^2\cdot K\eta_{\boldsymbol{p},\boldsymbol{r}}\frac{\bar{r}}{\sqrt{\underline{p}}} = o_p(n^{-1/2}\sigma),$$

then the projection-mismatch remainder (3.16) satisfies

$$\left\langle P^*_{\widehat{\mathbf{\Sigma}},\widehat{\mathcal{T}}}(\mathcal{A})-\mathcal{A},\ \mathcal{T}^*\right\rangle = o_p\big(n^{-1/2}\sigma\|\Phi(\mathcal{A})\|_{\mathbf{\Sigma}}\big).$$

Proof of Lemma 5. According to the analysis of the proof of (2.9),

$$\begin{aligned}
P^*_{\widehat{\mathbf{\Sigma}},\widehat{\mathcal{T}}}(\mathcal{A}) =& P_{\widehat{\mathbf{\Sigma}},\widehat{\mathcal{T}}}(\mathcal{A}\times_k(\widehat{\mathbf{\Sigma}}^{[k]})^{-1})\times_k\widehat{\mathbf{\Sigma}}^{[k]}\\
=& P^*_{\boldsymbol{I},\widetilde{\mathcal{T}}}(\mathcal{A}\times_k(\widehat{\mathbf{\Sigma}}^{[k]})^{1/2}\times_k(\widehat{\mathbf{\Sigma}}^{[k]})^{-1})\times_k(\widehat{\mathbf{\Sigma}}^{[k]})^{-1/2}\times_k(\widehat{\mathbf{\Sigma}}^{[k]})\\
=& P^*_{\boldsymbol{I},\widetilde{\mathcal{T}}}(\mathcal{A}\times_k(\widehat{\mathbf{\Sigma}}^{[k]})^{-1/2})\times_k(\widehat{\mathbf{\Sigma}}^{[k]})^{1/2},
\end{aligned}$$

where $\widetilde{\mathcal{T}}=\widehat{\mathcal{T}}\times_k(\widehat{\mathbf{\Sigma}}^{[k]})^{1/2}$. Define $\widetilde{\mathcal{A}}=\mathcal{A}\times_k(\widehat{\mathbf{\Sigma}}^{[k]})^{-1/2}$, then

$$P^*_{\widehat{\mathbf{\Sigma}},\widehat{\mathcal{T}}}(\mathcal{A})-\mathcal{A} = P^*_{\boldsymbol{I},\widetilde{\mathcal{T}}}(\widetilde{\mathcal{A}})\times_k(\widehat{\mathbf{\Sigma}}^{[k]})^{1/2}-\widetilde{\mathcal{A}}\times_k(\widehat{\mathbf{\Sigma}}^{[k]})^{1/2}.$$

Define $\widetilde{\mathcal{T}}^* = \mathcal{T}^* \times_k (\widehat{\mathbf{\Sigma}}^{[k]})^{1/2}$ and $\bar{\mathcal{T}}^* = \mathcal{T}^* \times_k (\mathbf{\Sigma}^{[k]})^{1/2}$, then

$$\langle P^*_{\widehat{\mathbf{\Sigma}},\widehat{\mathcal{T}}}(\mathcal{A}) - \mathcal{A},\ \mathcal{T}^*\rangle = \langle P^*_{\boldsymbol{I},\widetilde{\mathcal{T}}}(\widetilde{\mathcal{A}}) - \widetilde{\mathcal{A}}, \widetilde{\mathcal{T}}^*\rangle.$$

Rewirte $\widetilde{\mathcal{T}}^* = \widetilde{\mathcal{G}}^* \times_k \widetilde{\boldsymbol{U}}_k^*$ and $\bar{\mathcal{T}}^* = \bar{\mathcal{G}}^* \times_k \bar{\boldsymbol{U}}_k^*$ with $\widetilde{\boldsymbol{U}}_k^*, \bar{\boldsymbol{U}}_k^* \in \mathbb{R}^{p_k \times r_k}$ having orthogonal columns. Then by Lemma 3.1 of Koch and Lubich (2010),

$$\begin{aligned}
&\widetilde{\mathcal{A}} - P^*_{\boldsymbol{I},\widetilde{\mathcal{T}}}(\widetilde{\mathcal{A}}) \\
=&\widetilde{\mathcal{A}} - \widetilde{\mathcal{A}} \times_{k=1}^K \mathbf{P}_{\widetilde{\boldsymbol{U}}_k} - \sum_{k=1}^K \mathrm{Mat}_k^{-1}(\mathbf{P}^\perp_{\widetilde{\boldsymbol{U}}_k} \widetilde{\mathbf{A}}_k \mathbf{P}_{\check{\boldsymbol{W}}_k}) \\
=&\sum_{\emptyset \neq S \subseteq [K]} \widetilde{\mathcal{A}} \times_{k\in S} \mathbf{P}^\perp_{\widetilde{\boldsymbol{U}}_k} \times_{k\notin S} \mathbf{P}_{\widetilde{\boldsymbol{U}}_k} - \sum_{k=1}^K \mathrm{Mat}_k^{-1}(\mathbf{P}^\perp_{\widetilde{\boldsymbol{U}}_k} \widetilde{\mathbf{A}}_k \mathbf{P}_{\check{\boldsymbol{W}}_k}) \\
=&\sum_{k=1}^K \mathrm{Mat}_k^{-1}\big(\mathbf{P}^\perp_{\widetilde{\boldsymbol{U}}_k} \widetilde{\mathbf{A}}_k (\otimes_{j\neq k} \widetilde{\boldsymbol{U}}_j) \mathbf{P}^\perp_{\widetilde{\boldsymbol{G}}_k^\top} (\otimes_{j\neq k} \widetilde{\boldsymbol{U}}_j)^\top\big) + \sum_{S\subseteq[K], |S|\geq 2} \widetilde{\mathcal{A}} \times_{k\in S} \mathbf{P}^\perp_{\widetilde{\boldsymbol{U}}_k} \times_{k\notin S} \mathbf{P}_{\widetilde{\boldsymbol{U}}_k},
\end{aligned}$$

where $\check{\boldsymbol{W}}_k = (\otimes_{j\neq k} \widetilde{\boldsymbol{U}}_j)\mathrm{QR}(\widetilde{\boldsymbol{G}}_k^\top)$. The last equality holds since $\mathrm{Mat}_k\big(\widetilde{\mathcal{A}} \times_k \mathbf{P}^\perp_{\widetilde{\boldsymbol{U}}_k} \times_{j\neq k} \mathbf{P}_{\widetilde{\boldsymbol{U}}_j}\big) = \mathbf{P}^\perp_{\widetilde{\boldsymbol{U}}_k} \widetilde{\mathbf{A}}_k (\otimes_{j\neq k} \mathbf{P}_{\widetilde{\boldsymbol{U}}_j})$ and $\mathbf{P}_{\check{\boldsymbol{W}}_k} = (\otimes_{j\neq k} \widetilde{\boldsymbol{U}}_j) \widetilde{\boldsymbol{G}}_k^\top (\widetilde{\boldsymbol{G}}_k \widetilde{\boldsymbol{G}}_k^\top)^{-1} \widetilde{\boldsymbol{G}}_k^\top (\otimes_{j\neq k} \widetilde{\boldsymbol{U}}_j)^\top$, and then

$$\mathrm{Mat}_k\big(\widetilde{\mathcal{A}} \times_k \mathbf{P}^\perp_{\widetilde{\boldsymbol{U}}_k} \times_{j\neq k} \mathbf{P}_{\widetilde{\boldsymbol{U}}_j}\big) - \mathbf{P}^\perp_{\widetilde{\boldsymbol{U}}_k} \widetilde{\mathbf{A}}_k \mathbf{P}_{\check{\boldsymbol{W}}_k} = \mathbf{P}^\perp_{\widetilde{\boldsymbol{U}}_k} \widetilde{\mathbf{A}}_k (\otimes_{j\neq k} \widetilde{\boldsymbol{U}}_j) \mathbf{P}^\perp_{\widetilde{\boldsymbol{G}}_k^\top} (\otimes_{j\neq k} \widetilde{\boldsymbol{U}}_j)^\top.$$

Therefore,

$$
\begin{aligned}
&\langle \widetilde{\mathcal{A}} - P_{\boldsymbol{I},\widetilde{\mathcal{T}}}(\widetilde{\mathcal{A}}), \widetilde{\mathcal{T}}^* \rangle \\
=&\sum_{k=1}^{K} \langle \mathbf{P}_{\widetilde{\boldsymbol{U}}_k}^{\perp} \widetilde{\mathbf{A}}_k (\otimes_{j\neq k} \widetilde{\boldsymbol{U}}_j) \mathbf{P}_{\widetilde{\boldsymbol{G}}_k^\top}^{\perp} (\otimes_{j\neq k} \widetilde{\boldsymbol{U}}_j)^\top, \widetilde{\boldsymbol{U}}_k^* \widetilde{\boldsymbol{G}}_k (\otimes_{j\neq k} \widetilde{\boldsymbol{U}}_j^*)^\top \rangle \\
&+ \sum_{S\subseteq[K],|S|\geq 2} \langle \widetilde{\mathcal{A}} \times_{k\in S} \mathbf{P}_{\widetilde{\boldsymbol{U}}_k}^{\perp} \times_{k\notin S} \mathbf{P}_{\widetilde{\boldsymbol{U}}_k}, \widetilde{\mathcal{G}}^* \times_k \widetilde{\boldsymbol{U}}_k^* \rangle \\
=&\sum_{k=1}^{K} \operatorname{tr}(\widetilde{\boldsymbol{U}}_k^{*\top} \mathbf{P}_{\widetilde{\boldsymbol{U}}_k}^{\perp} \widetilde{\mathbf{A}}_k (\otimes_{j\neq k} \widetilde{\boldsymbol{U}}_j) \mathbf{P}_{(\widetilde{\boldsymbol{G}}_k)^\top}^{\perp} (\otimes_{j\neq k} \widetilde{\boldsymbol{U}}_j)^\top (\otimes_{j\neq k} \widetilde{\boldsymbol{U}}_j^*) \widetilde{\boldsymbol{G}}_k^{*\top}) + \\
&+ \sum_{S\subseteq[K],|S|\geq 2} \langle \widetilde{\mathcal{A}} \times_{k\in S} \mathbf{P}_{\widetilde{\boldsymbol{U}}_k}^{\perp} \times_{k\notin S} \mathbf{P}_{\widetilde{\boldsymbol{U}}_k}, \widetilde{\mathcal{G}}^* \times_k \widetilde{\boldsymbol{U}}_k^* \rangle \\
=&\sum_{k=1}^{K} \operatorname{tr}(\widetilde{\boldsymbol{U}}_k^{*\top} \mathbf{P}_{\widetilde{\boldsymbol{U}}_k}^{\perp} \widetilde{\mathbf{A}}_k \mathbf{P}_{\widetilde{\boldsymbol{W}}_k}^{\perp} \widetilde{\boldsymbol{W}}_k^*) - \sum_{k=1}^{K} \operatorname{tr}(\widetilde{\boldsymbol{U}}_k^{*\top} \mathbf{P}_{\widetilde{\boldsymbol{U}}_k}^{\perp} \widetilde{\mathbf{A}}_k \mathbf{P}_{(\otimes_{j\neq k} \widetilde{\boldsymbol{U}}_j)}^{\perp} (\otimes_{j\neq k} \widetilde{\boldsymbol{U}}_j^*) \widetilde{\boldsymbol{G}}_k^{*\top}) + \\
&+ \sum_{S\subseteq[K],|S|\geq 2} \langle \widetilde{\mathcal{A}} \times_{k\in S} \mathbf{P}_{\widetilde{\boldsymbol{U}}_k}^{\perp} \times_{k\notin S} \mathbf{P}_{\widetilde{\boldsymbol{U}}_k}, \widetilde{\mathcal{G}}^* \times_k \widetilde{\boldsymbol{U}}_k^* \rangle \\
=&\sum_{k=1}^{K} \operatorname{tr}(\widetilde{\boldsymbol{U}}_k^{*\top} \mathbf{P}_{\widetilde{\boldsymbol{U}}_k}^{\perp} \widetilde{\mathbf{A}}_k \mathbf{P}_{\widetilde{\boldsymbol{W}}_k}^{\perp} \widetilde{\boldsymbol{W}}_k^*) - \sum_{S\subseteq[K],|S|\geq 2} (|S|-1) \langle \widetilde{\mathcal{A}} \times_{k\in S} \mathbf{P}_{\widetilde{\boldsymbol{U}}_k}^{\perp} \times_{k\notin S} \mathbf{P}_{\widetilde{\boldsymbol{U}}_k}, \widetilde{\mathcal{G}}^* \times_k \widetilde{\boldsymbol{U}}_k^* \rangle
\end{aligned}
$$

Denote $\widetilde{\boldsymbol{W}}_k = (\otimes_{j\neq k} \widetilde{\boldsymbol{U}}_j) \widetilde{\boldsymbol{G}}_k^\top$ and $\widetilde{\boldsymbol{W}}_k^* = (\otimes_{j\neq k} \widetilde{\boldsymbol{U}}_j^*) \widetilde{\boldsymbol{G}}_k^{*\top}$. Define the SVDs of $\mathbf{P}_{\widetilde{\boldsymbol{U}}_k}^{\perp} \widetilde{\boldsymbol{U}}_k^*$ as $\mathbf{P}_{\widetilde{\boldsymbol{U}}_k}^{\perp} \widetilde{\boldsymbol{U}}_k^* = \boldsymbol{Q}_k \boldsymbol{\Lambda}_k \mathbf{R}_k^\top$ and $\mathbf{P}_{\widetilde{\boldsymbol{W}}_k}^{\perp} \widetilde{\boldsymbol{W}}_k^* = \boldsymbol{Q}_k^\circ \boldsymbol{\Lambda}_k^\circ \mathbf{R}_k^{\circ\top}$ with $\boldsymbol{\Lambda}_k, \boldsymbol{\Lambda}_k^\circ \in \mathbb{R}^{r_k\times r_k}$ for $k\in[K]$. Note that by the consistency of $\widehat{\boldsymbol{\Sigma}}$ and the Wedin's $\sin\Theta$ theorem (Theorem 2.9 of Chen et al. (2021)),

$$
\|\boldsymbol{\Lambda}_k\|_2 \lesssim \lambda^{-1} \|\widehat{\mathcal{T}} - \mathcal{T}^*\|_{\boldsymbol{\Sigma}},\ \|\boldsymbol{\Lambda}_k^\circ\|_2 \lesssim \lambda^{-1} \|\widehat{\mathcal{T}} - \mathcal{T}^*\|_{\boldsymbol{\Sigma}},
$$

for $k\in[K]$. Using $|\operatorname{tr}(XY)| \leq \|X\|_* \|Y\|_2$, we obtain

$$
\left|\operatorname{tr}(\widetilde{\boldsymbol{U}}_k^{*\top} \mathbf{P}_{\widetilde{\boldsymbol{U}}_k}^{\perp} \widetilde{\mathbf{A}}_k \mathbf{P}_{\widetilde{\boldsymbol{W}}_k}^{\perp} \widetilde{\boldsymbol{W}}_k^*)\right| \lesssim r_k \|\boldsymbol{\Lambda}_k\|_2 \|\boldsymbol{\Lambda}_k^\circ\|_2 \|\boldsymbol{Q}_k^\top \bar{\mathbf{A}}_k \boldsymbol{Q}_k^\circ\|_F,
$$

---

where $\bar{\mathcal{A}} = \mathcal{A} \times_k (\mathbf{\Sigma}^{[k]})^{1/2}$ and $\widetilde{\mathcal{A}} \xrightarrow{p} \bar{\mathcal{A}}$. Here $\boldsymbol{Q}_k \in \mathbb{R}^{p_k \times r_k}, \boldsymbol{Q}_k^\circ \in \mathbb{R}^{p_{-k} \times r_k}$ have orthonormal columns extracted from mismatch subspaces. Then Assumption 7 implies

$$\|\boldsymbol{Q}_k^\top \bar{\mathbf{A}}_k \boldsymbol{Q}_k^\circ\|_F \lesssim \eta \cdot \max\Big\{\sqrt{\frac{r_k}{p_k}}\ \|\mathbf{P}^\perp_{\widetilde{\boldsymbol{U}}_k} \bar{\mathbf{A}}_k \mathbf{P}_{\widetilde{\boldsymbol{W}}_k}\|_F,\ \sqrt{\frac{r_k}{p_{-k}}}\ \|\mathbf{P}_{\widetilde{\boldsymbol{U}}_k} \widetilde{\mathbf{A}}_k \mathbf{P}^\perp_{\widetilde{\boldsymbol{W}}_k}\|_F\Big\}.$$

Thus,

$$\Big|\sum_{k=1}^K \mathrm{tr}(\widetilde{\boldsymbol{U}}_k^{*\top} \mathbf{P}^\perp_{\widetilde{\boldsymbol{U}}_k} \widetilde{\mathbf{A}}_k \mathbf{P}^\perp_{\widetilde{\boldsymbol{W}}_k} \widetilde{\boldsymbol{W}}_k^*)\Big| \le \lambda^{-2} \|\widehat{\mathcal{T}} - \mathcal{T}\|^2_{\mathbf{\Sigma}} \cdot \eta \frac{\bar{r}}{\sqrt{\underline{p}}} \times \sum_{k=1}^K \|\mathbf{P}^\perp_{\widetilde{\boldsymbol{U}}_k} \bar{\mathbf{A}}_k \mathbf{P}_{\widetilde{\boldsymbol{W}}_k}\|_F,$$

and

$$\begin{aligned}&\sum_{S\subseteq[K],|S|\ge 2} (|S|-1)\Big|\langle \widetilde{\mathcal{A}} \times_{k\in S} \mathbf{P}^\perp_{\widetilde{\boldsymbol{U}}_k} \times_{k\notin S} \mathbf{P}_{\widetilde{\boldsymbol{U}}_k}, \widetilde{\mathcal{G}}^* \times_k \widetilde{\boldsymbol{U}}_k^*\rangle\Big| \\ \lesssim_{\underline{r}} &\lambda^{-2} \|\widehat{\mathcal{T}} - \mathcal{T}\|^2_{\mathbf{\Sigma}} \|\widetilde{\mathcal{A}} \times_k \widetilde{\boldsymbol{U}}_k\|_F^2 = O_p\Big(\lambda^{-2} \|\widehat{\mathcal{T}} - \mathcal{T}\|^2_{\mathbf{\Sigma}} \cdot \eta K \frac{\bar{r}}{\sqrt{\underline{p}}} \sum_{k=1}^K \|\mathbf{P}^\perp_{\widetilde{\boldsymbol{U}}_k} \bar{\mathbf{A}}_k \mathbf{P}_{\widetilde{\boldsymbol{W}}_k}\|_F\Big).\end{aligned}$$

Therefore,

$$\big|\langle \widetilde{\mathcal{A}} - P_{\boldsymbol{I},\widetilde{\mathcal{T}}}(\widetilde{\mathcal{A}}), \widetilde{\mathcal{T}}^*\rangle\big| \lesssim \lambda^{-2} \|\widehat{\mathcal{T}} - \mathcal{T}\|^2_{\mathbf{\Sigma}} \cdot \eta K \frac{\bar{r}}{\sqrt{\underline{p}}} \sum_{k=1}^K \|\mathbf{P}^\perp_{\widetilde{\boldsymbol{U}}_k} \bar{\mathbf{A}}_k \mathbf{P}_{\widetilde{\boldsymbol{W}}_k}\|_F.$$

Recall that

$$\begin{aligned}\|\Phi(\mathcal{A})\|^2_{\mathbf{\Sigma}} =& \|P_{\mathbf{\Sigma},\mathcal{T}^*}(\mathcal{A} \times_k (\mathbf{\Sigma}^{[k]})^{-1})\|_{\mathbf{\Sigma}} \\ =& \|P_{\boldsymbol{I},\widetilde{\mathcal{T}}^*}(\mathcal{A} \times_k (\mathbf{\Sigma}^{[k]})^{-1/2}) \times_k (\mathbf{\Sigma}^{[k]})^{-1/2}\|^2_{\mathbf{\Sigma}} \\ =& \|P_{\boldsymbol{I},\widetilde{\mathcal{T}}^*}(\mathcal{A} \times_k (\mathbf{\Sigma}^{[k]})^{-1/2})\|^2_F \\ =& \|\bar{\mathcal{A}} \times_k \mathbf{P}_{\bar{\boldsymbol{U}}_k^*}\|_F^2 + \sum_k \|\mathbf{P}^\perp_{\bar{\boldsymbol{U}}_k^*} \bar{\mathbf{A}}_k \mathbf{P}_{\bar{\boldsymbol{W}}_k^*}\|_F^2,\end{aligned}$$

where $\bar{\boldsymbol{W}}_k^* = (\otimes_{j\ne k} \bar{\boldsymbol{U}}_j^*)\mathrm{QR}(\bar{\boldsymbol{G}}_k^*)^\top$. Then we can conclude that

$$\big|\langle \widetilde{\mathcal{A}} - P_{\boldsymbol{I},\widetilde{\mathcal{T}}}(\widetilde{\mathcal{A}}), \widetilde{\mathcal{T}}^*\rangle\big| \lesssim \lambda^{-2} \|\widehat{\mathcal{T}} - \mathcal{T}\|^2_{\mathbf{\Sigma}} \cdot \eta K \frac{\bar{r}}{\sqrt{\underline{p}}} \|\Phi(\mathcal{A})\|_{\mathbf{\Sigma}}.$$

Under the scaling condition in that theorem implies

$$\lambda^{-2}\|\widehat{\mathcal{T}} - \mathcal{T}^*\|_F^2 \cdot \eta K \frac{\bar{r}}{\sqrt{\underline{p}}} = o_p(n^{-1/2}\sigma).$$

Therefore, $\big|\langle P^*_{\widehat{\mathbf{\Sigma}},\widehat{\mathcal{T}}}(\mathcal{A}) - \mathcal{A},\ \mathcal{T}^*\rangle\big| = o_p\big(n^{-1/2}\sigma\|\Phi(\mathcal{A})\|_{\mathbf{\Sigma}}\big)$, which completes the proof.

$\square$

S5. Proofs of Section 3.2

Next we verify the assumptions in Theorem 2 and prove Corollary 1.

Proof of Corollary 1. Under Assumption 4, $\mathrm{vec}(\mathcal{X}_i)$ is sub-Gaussian and the covariances admit a Kronecker structure; hence the basic design conditions required by Theorem 2 hold.

Verification of Assumption 2. In the feasible procedure, $\widehat{\mathbf{\Sigma}}$ is constructed from the mode-wise estimators in (2.13). Lemma 4 proves that $\|\widehat{\mathbf{\Sigma}} - \mathbf{\Sigma}\|_2 = O_p(n^{-1/2})$, verifying the covariance consistency condition. *Verification of Assumption 3.* The feasible loading $\widehat{\Phi}(\mathcal{A})$ is defined via the plug-in operator that $\widehat{\Phi}(\mathcal{A}) = P_{\widehat{\mathbf{\Sigma}},\widehat{\mathcal{T}}_{\boldsymbol{r}}}(\mathrm{vec}^{-1}(\widehat{\mathbf{\Sigma}}^{-1}\mathrm{vec}(\mathcal{A})))$. Using Assumption 2 and the mode-wise subspace consistency in Assumption 6, standard perturbation arguments yield

$$\|\widehat{\Phi}(\mathcal{A}) - \Phi(\mathcal{A})\|_{\mathbf{\Sigma}} = o_p(1)\,\|\Phi(\mathcal{A})\|_{\mathbf{\Sigma}}.$$

Verification of the projection-mismatch term in (3.16)*.* Together with Assumption 7, Lemma 5 implies condition (3.16) holds.

With the sample-splitting independence between $(\widehat{\Phi}(\mathcal{A}), \widehat{\boldsymbol{\Sigma}}, \widehat{\mathcal{T}})$ and $\{(\mathcal{X}_i, \varepsilon_i)\}$'s are ensured. Theorem 2 applies to each split estimator $\widehat{\Delta}_{\mathcal{A}}^{(1)}$ and $\widehat{\Delta}_{\mathcal{A}}^{(2)}$, respectively. In particular, for each split $j = 1, 2$,

$$\sqrt{|\mathcal{D}_j|}\, V^{-1/2}\big(\widehat{\Delta}_{\mathcal{A}}^{(j)} - \Delta_{\mathcal{A}}\big) = \frac{1}{\sqrt{|\mathcal{D}_j|}} \sum_{i \in \mathcal{D}_j} \langle \Phi(\mathcal{A}), \mathcal{X}_i \rangle \varepsilon_i + o_p(1).$$

Let $\widehat{\Delta}_{\mathcal{A}} := \frac{1}{2}\big(\widehat{\Delta}_{\mathcal{A}}^{(1)} + \widehat{\Delta}_{\mathcal{A}}^{(2)}\big)$. If $|\mathcal{D}_1|/N \to 1/2$ and $|\mathcal{D}_2|/N \to 1/2$, then

$$\sqrt{N}\, V^{-1/2}\big(\widehat{\Delta}_{\mathcal{A}} - \Delta_{\mathcal{A}}\big) = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \langle \Phi(\mathcal{A}), \mathcal{X}_i \rangle \varepsilon_i + o_p(1) \xrightarrow{d} N(0,1).$$

□

Proof of Corollary 2. Following the proof of Corollary 1, we can verify the assumptions in Lemma 2 and prove $\widehat{V} \xrightarrow{p} V$. Then Slutsky's theorem yields $\sqrt{N}\, \widehat{V}^{-1/2}\big(\widehat{\Delta}_A - \Delta_A\big) \xrightarrow{d} N(0,1)$. □

S6. Robustness to rank misspecification and heteroscedastic heavy-tailed errors

We further examine the robustness of the proposed inference procedure to rank misspecification under heteroscedastic heavy-tailed errors. For both the matrix and tensor regression settings, the errors are generated as

$$\varepsilon_i = \sqrt{\frac{1 + g_i^2}{2}}\, z_i, \qquad g_i = \frac{\langle \mathcal{X}_i, \mathcal{T}^\star \rangle}{\big\|\mathcal{T}^\star \times_{k=1}^{K} (\boldsymbol{\Sigma}^{[k]})^{1/2}\big\|_F},$$

where $z_i = t_{3,i}/\sqrt{3}$ and $t_{3,i} \overset{\text{i.i.d.}}{\sim} t_3$. For matrix regression, the true rank is $r^\star = 2$, whereas the projection operator is constructed using the working ranks $\widetilde{r} \in \{1, \ldots, 5\}$. For tensor regression, the true Tucker rank is $\boldsymbol{r}^\star = (2, 2, 2)$, and we consider both under-specified and over-specified working Tucker ranks. All remaining data-generating settings, loading constructions, and sample-size regimes follow those in Sections 4.1 and 4.2.

Tables S1 and S2 report the empirical bias (BIAS), empirical standard deviation (SD), average estimated standard error (SE), and coverage probability (CP) of the nominal 95% Wald confidence intervals for the matrix and tensor settings, respectively. The results reveal a consistent distinction between rank under-specification and over-specification. When the working rank is smaller than the true rank, genuine signal directions are omitted from the working local geometry, leading to persistent bias and substantial undercoverage. In several settings, the coverage deteriorates further as the sample size increases because the bias does not vanish at the same rate as the confidence intervals shrink. By contrast, the correctly specified and over-specified working ranks yield negligible bias, close agreement between SE and SD, and coverage probabilities near the nominal 95% level. The primary empirical consequence of rank over-specification is an increase in variability. Overall, these results suggest that the proposed procedure remains empirically stable under heteroscedastic heavy-tailed errors, provided that the working rank is not under-specified.

References

Chen, Y., Y. Chi, J. Fan, and C. Ma (2021). Spectral methods for data science: A statistical perspective. *Foundations and Trends® in Machine Learning 14*(5), 566–806.

Horn, R. A. and C. R. Johnson (2013). *Matrix Analysis* (2 ed.). Cambridge, UK: Cambridge University Press.

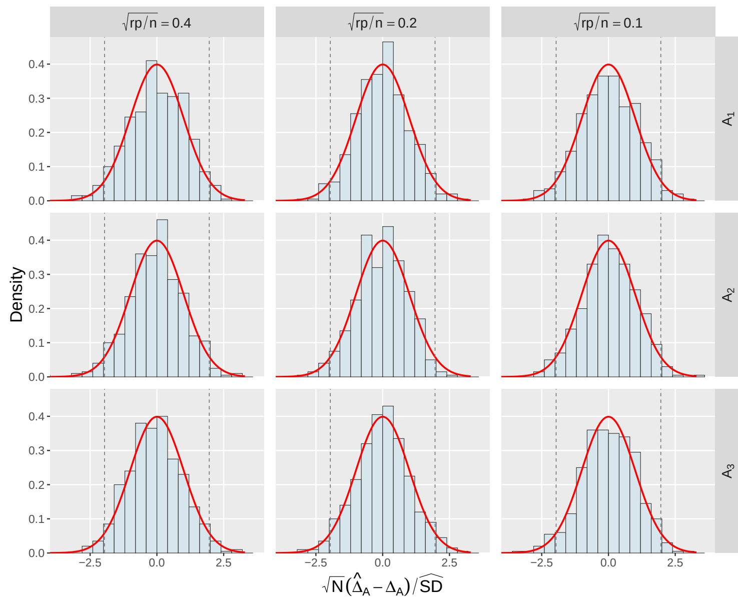
Ke, B., W. Zhao, and L. Wang (2025). Statistical inference for matrix-vector linear regression without debiasing under kronecker covariance structure. *Statistics and Computing 35*(6), 201.

Koch, O. and C. Lubich (2010). Dynamical tensor approximation. *SIAM Journal on Matrix Analysis and Applications 31*(5), 2360–2375.

Lu, W., Z. Zhu, and H. Lian (2023). Sparse and low-rank matrix quantile estimation with application to quadratic regression. *Statistica Sinica 33*(2), 945–959.

Vershynin, R. (2018). *High-Dimensional Probability: An Introduction with Applications in Data Science.* Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press.

(a) Identity covariance



(b) General Kronecker covariance

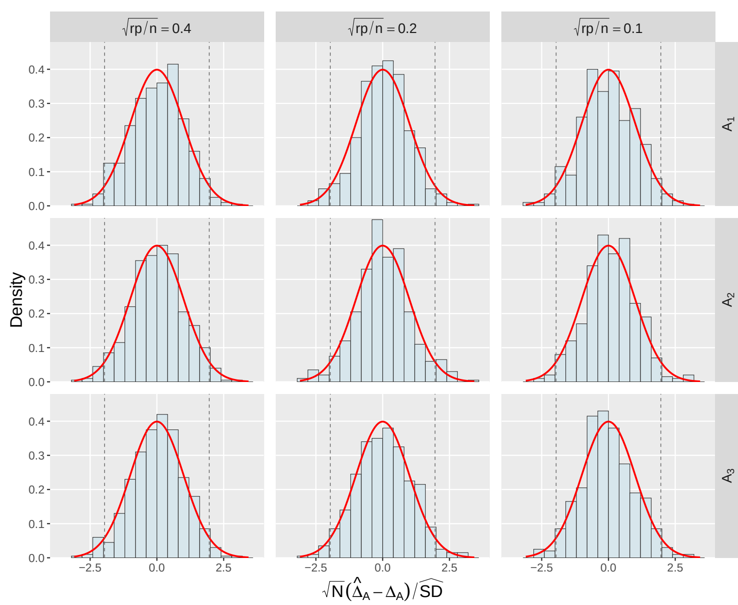
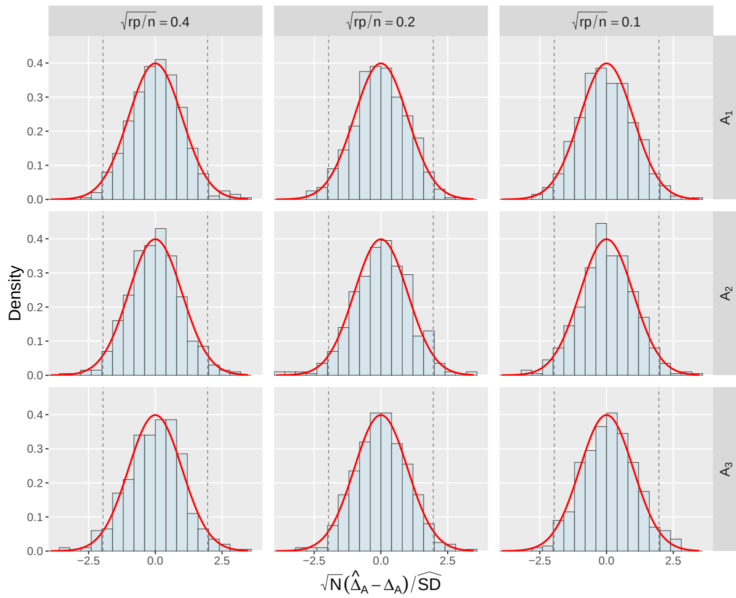


Figure S1: Empirical densities of the studentized statistic $\sqrt{N}\,\widehat{V}^{-1/2}(\widehat{\Delta}_{\mathbf{A}} - \Delta_{\mathbf{A}})$ in matrix trace regression under identity and general Kronecker covariances.

(a) Identity covariance



(b) General Kronecker covariance

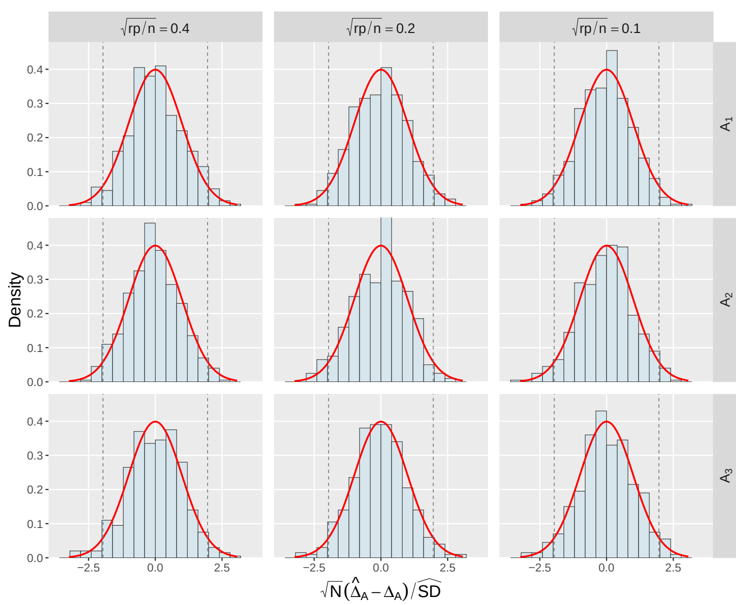


Figure S2: Empirical densities of the studentized statistic $\sqrt{N}\,\widehat{V}^{-1/2}(\widehat{\Delta}_{\mathcal{A}} - \Delta_{\mathcal{A}})$ in Tucker tensor regression under identity and general Kronecker covariances with $K = 3$.

Table S1: Sensitivity of matrix regression inference to the working rank under Kronecker covariance and heteroscedastic t_3 errors. The true rank is $r^\star = 2$. BIAS, SD, and SE are multiplied by 100.

	$\sqrt{r^\star p/n} = 0.4$				$\sqrt{r^\star p/n} = 0.2$				$\sqrt{r^\star p/n} = 0.1$			
$\widetilde{r}$	BIAS	SD	SE	CP	BIAS	SD	SE	CP	BIAS	SD	SE	CP
					Case (i): Loading A_1							
1	4.918	0.911	0.865	0.006	5.080	0.467	0.436	0.000	5.092	0.233	0.220	0.000
2	-0.037	1.636	1.612	0.944	0.046	0.753	0.808	0.956	0.036	0.424	0.401	0.946
3	-0.052	1.830	1.819	0.958	0.057	0.857	0.904	0.968	0.047	0.463	0.450	0.954
4	-0.040	2.049	2.056	0.950	0.030	0.951	1.021	0.968	0.049	0.514	0.512	0.958
5	-0.016	2.320	2.313	0.950	0.041	1.109	1.150	0.950	0.016	0.597	0.574	0.946
					Case (ii): Loading A_2							
1	-6.367	0.629	0.550	0.000	-6.445	0.313	0.270	0.000	-6.458	0.168	0.136	0.000
2	0.106	1.510	1.493	0.964	0.031	0.830	0.752	0.938	0.002	0.353	0.373	0.958
3	0.123	2.009	1.930	0.966	0.037	1.036	0.962	0.948	0.003	0.444	0.473	0.964
4	0.079	2.336	2.201	0.956	0.000	1.195	1.167	0.952	0.013	0.587	0.598	0.954
5	0.076	2.535	2.329	0.958	0.013	1.246	1.227	0.950	0.008	0.629	0.630	0.950
					Case (iii): Loading A_3							
1	-2.064	0.813	0.712	0.194	-2.193	0.417	0.357	0.002	-2.166	0.212	0.179	0.000
2	0.061	1.313	1.389	0.972	-0.013	0.697	0.699	0.960	0.019	0.354	0.353	0.948
3	0.093	1.640	1.732	0.968	-0.050	0.834	0.832	0.954	0.017	0.406	0.406	0.950
4	0.029	1.858	1.925	0.962	-0.101	1.000	0.995	0.946	0.038	0.507	0.506	0.948
5	0.056	2.129	2.215	0.962	-0.105	1.135	1.137	0.954	0.037	0.579	0.582	0.942

Table S2: Sensitivity of tensor regression inference to the working Tucker rank under Kronecker covariance and heteroscedastic t_3 errors. The true Tucker rank is $\boldsymbol{r}^\star = (2, 2, 2)$. BIAS, SD, and SE are multiplied by 100.

	$\sqrt{r^\star p/n} = 0.4$				$\sqrt{r^\star p/n} = 0.2$				$\sqrt{r^\star p/n} = 0.1$			
$\widetilde{\boldsymbol{r}}$	BIAS	SD	SE	CP	BIAS	SD	SE	CP	BIAS	SD	SE	CP
						Case (i): Loading A_1						
$(1,1,1)$	-0.141	0.168	0.113	0.606	-0.160	0.124	0.055	0.348	-0.159	0.114	0.027	0.222
$(1,2,2)$	-0.150	0.158	0.130	0.676	-0.158	0.106	0.064	0.394	-0.156	0.091	0.032	0.190
(2,2,2)	-0.005	0.177	0.170	0.948	0.008	0.089	0.086	0.944	0.005	0.043	0.043	0.956
$(2,2,3)$	-0.005	0.184	0.176	0.946	0.007	0.092	0.089	0.948	0.005	0.045	0.044	0.954
$(3,3,3)$	-0.009	0.213	0.204	0.948	0.008	0.106	0.102	0.936	0.004	0.051	0.051	0.954
$(4,4,4)$	-0.001	0.239	0.230	0.960	0.011	0.116	0.114	0.946	0.004	0.057	0.057	0.952
$(5,5,5)$	-0.014	0.270	0.255	0.968	0.011	0.132	0.128	0.942	0.005	0.064	0.064	0.950
						Case (ii): Loading A_2						
$(1,1,1)$	-0.200	0.192	0.111	0.486	-0.212	0.144	0.054	0.254	-0.215	0.144	0.027	0.186
$(1,2,2)$	-0.199	0.183	0.136	0.584	-0.210	0.129	0.067	0.288	-0.214	0.110	0.033	0.138
(2,2,2)	-0.015	0.196	0.183	0.936	-0.003	0.095	0.092	0.936	0.000	0.047	0.046	0.946
$(2,2,3)$	-0.015	0.195	0.183	0.940	-0.003	0.096	0.092	0.934	0.000	0.047	0.046	0.942
$(3,3,3)$	-0.017	0.214	0.207	0.956	-0.004	0.111	0.105	0.938	0.002	0.055	0.052	0.950
$(4,4,4)$	-0.018	0.244	0.232	0.944	-0.010	0.124	0.117	0.936	0.002	0.059	0.058	0.960
$(5,5,5)$	-0.018	0.264	0.259	0.956	-0.011	0.136	0.130	0.944	0.002	0.066	0.065	0.964
						Case (iii): Loading A_3						
$(1,1,1)$	-0.569	0.381	0.224	0.368	-0.519	0.313	0.114	0.174	-0.524	0.295	0.057	0.050
$(1,2,2)$	-0.559	0.426	0.285	0.504	-0.519	0.347	0.146	0.232	-0.525	0.306	0.072	0.132
(2,2,2)	0.002	0.351	0.329	0.952	0.013	0.168	0.166	0.950	0.001	0.083	0.082	0.958
$(2,2,3)$	0.001	0.372	0.349	0.960	0.017	0.180	0.177	0.948	0.001	0.088	0.088	0.950
$(3,3,3)$	0.003	0.381	0.357	0.954	0.021	0.184	0.182	0.958	0.001	0.091	0.090	0.948
$(4,4,4)$	-0.001	0.414	0.383	0.964	0.018	0.197	0.194	0.952	0.002	0.096	0.096	0.958
$(5,5,5)$	-0.019	0.425	0.408	0.946	0.015	0.209	0.206	0.952	0.005	0.104	0.102	0.958