Supplementary Materials for "Bias–variance Tradeoff in Tensor Estimation"

Without loss of generality, if $W \sim \text{subGaussian}(0, \kappa^2)$, we assume throughout, unless otherwise specified in the appendix, that $\mathbb{E}[W^2] = \kappa^2$.

A Additional numerical results

Spectral gap assumption. Theorem 1 requires a mode-wise spectral gap condition on the singular values of $\mathcal{M}_k(X^*)$ as in (3). We empirically examine this assumption on the same 3D brain MRI dataset as in Section 4.1. Specifically, we randomly draw 50 T1-weighted brain volumes from the IXI dataset (IXI, 2002). For each volume and each mode $k \in \{1, 2, 3\}$, we compute the singular values of $\mathcal{M}_k(X^*)$ and record the gaps

$$\sigma_{r_k}(\mathcal{M}_k(X^*)) - \sigma_{r_k+1}(\mathcal{M}_k(X^*)), \qquad r_k = 1, \ldots, 145.$$

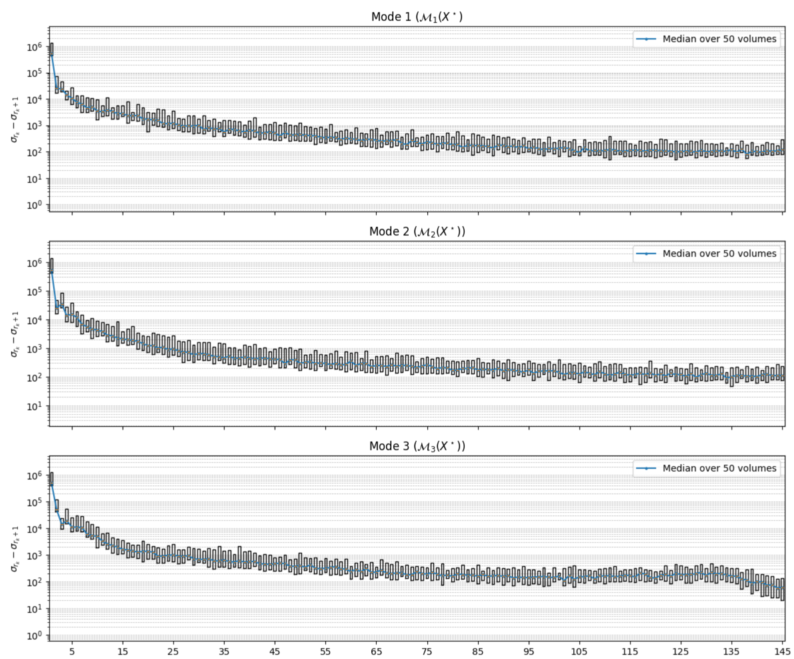

Figure 2: Empirical median and interquartile range (Q1–Q3) of the mode-wise singular-value gaps (log scale) over 50 T1-weighted IXI brain MRI volumes.

Figure 2 summarizes the distribution of these gaps across subjects. The empirical medians remain strictly positive over all ranks and modes, and the inter-quartile bands indicate that this behavior is stable across the 50 volumes. This provides concrete evidence that the spectral gap condition in (3) is not merely a idealized technical assumption of the analysis, but is indeed observed in practice.

Bias–variance curve. To empirically quantify the contributions of bias and variance to the error $\|Y - X^*\|_{\mathrm{F}}$, we consider the synthetic tensor model of Section 4.2 with $p = 100$. For each candidate Tucker rank $r \in \{12, 15, 18, \ldots, 84, 87\}$, we run 50 independent repetitions. The bias component is obtained as the error $\|\widetilde{X} - X^*\|_{\mathrm{F}}$, where $\widetilde{X}$ denotes the output of Algorithm 1 when applied directly to the noiseless tensor X^*. The variance component is obtained as

$$\max\left\{\sqrt{r}\left\|\mathcal{M}_1(X^*)\cdot\left\{U_2^{(1)}\otimes U_3^{(1)}\right\}\right\|, \sqrt{r}\left\|\mathcal{M}_2(X^*)\cdot\left\{U_1^{(1)}\otimes U_3^{(1)}\right\}\right\|, \sqrt{r}\left\|\mathcal{M}_3(X^*)\cdot\left\{U_1^{(1)}\otimes U_2^{(1)}\right\}\right\|, \left\|X^*\times_1 U_1^{(1)}\times_2 U_2^{(1)}\times_3 U_3^{(1)}\right\|_{\mathrm{F}}\right\}.$$

The Figure 3 reports the resulting mean bias and variance curves as functions of the target rank.

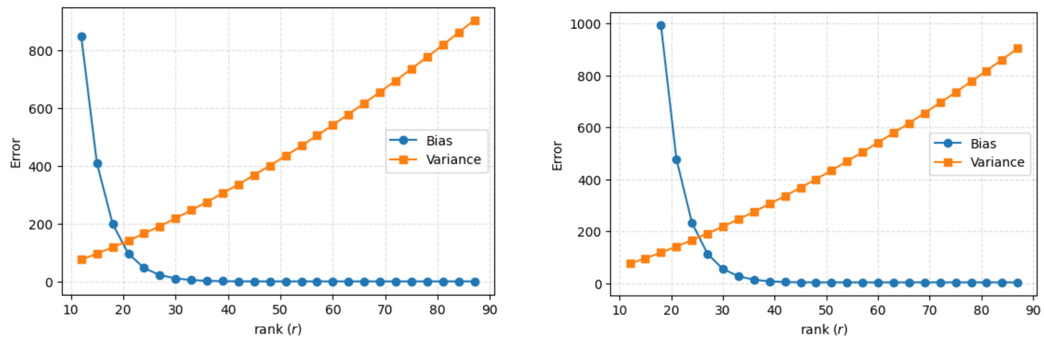

Figure 3: Mean bias and variance components of the error as functions of the target rank in the synthetic experiment with $p = 100$. The left panel reporesent the curve with scaling paramter $\lambda = 10$ and the right panel $\lambda = 50$.

Figure 3 exhibits the expected bias–variance trade-off: as the target rank increases, the bias term decreases while the variance term grows, leading to an intermediate rank that minimizes the overall error.

Practical choice of target ranks. There are several well-known ways to select the target Tucker ranks. To select the rank parameters in a data-driven way, we follow the discussion in (Ozdemir et al., 2017) and adopt an adaptive thresholding scheme. More precisely, we choose r_k so that $\sigma_{r_k}^2(\mathcal{M}_k(Y)) \geq \tau$, where τ is a threshold. A simple choice is $\tau = \alpha\|Y\|_{\mathrm{F}}^2$ with $\alpha = 0.01$; we retain all singular values of $\mathcal{M}_k(Y)$ up to the largest index j satisfying $\sigma_j^2(\mathcal{M}_k(Y)) \geq \tau$ and discard the rest.

B Lower bounds

In this section we prove the complementary minimax lower bounds Theorem 7 and Theorem 2 that match the estimation rates in Theorem 3 and Theorem 1. These results show that our upper bounds are sharp (up to universal constants), and therefore establish the minimax optimality of our estimation procedure.

B.1 Tensor lower bound

Proof of Theorem 2.
Step 1. We follow a similar strategy as in the proof of Theorem 7 to show

$$\inf_{\widehat{X}} \sup_{X^* \in \mathcal{T}^{\xi}_{(r_1,r_2,r_3)}} \mathbb{E}\big\|\widehat{X} - X^*\big\|_{\mathrm{F}}^2 \geq c(\kappa^2 r_1 r_2 r_3 + \xi^2). \tag{9}$$

Let

$$r_1' = \lfloor \xi^2/(r_2 r_3 \kappa^2) \rfloor.$$

Consider the set of tensors

$$\mathcal{B} = \{B \in \mathbb{R}^{p_1 \times p_2 \times p_3} : \|B\|_\infty \leq \kappa^2, B_{i,j,k} = 0 \text{ if } i > r_1 + r_1' \text{ or } j > r_2 \text{ or } k > r_3\}.$$

In this step, we show that

$$\mathcal{B} \subset \mathcal{T}^{\xi}_{(r_1,r_2,r_3)}. \tag{10}$$

To see this, let $B \in \mathcal{B}$, and denote B' to be the matrix such that

$$B'_{i,j,k} = \begin{cases} B_{i,j,k} & \text{if } i \leq r_1 \text{ and } j \leq r_2 \text{ and } k \leq r_3; \\ 0 & \text{otherwise.} \end{cases}$$

Then $B' \in \mathcal{T}_{(r_1,r_2,r_3)}$ and

$$\|B - B'\|_{\mathrm{F}}^2 \leq r_1' r_2 r_3 \|B\|_\infty = r_1' r_2 r_3 \kappa^2 \leq \xi^2.$$

Therefore $\mathcal{B} \subset \mathcal{T}^{\xi}_{(r_1,r_2,r_3)}$. So

$$\inf_{\widehat{X}} \sup_{X^* \in \mathcal{T}^{\xi}_{(r_1,r_2,r_3)}} \mathbb{E}\left\|\widehat{X} - X^*\right\|_{\mathrm{F}}^2 \geq \inf_{\widehat{X}} \sup_{X^* \in \mathcal{B}} \mathbb{E}\left\|\widehat{X} - X^*\right\|_{\mathrm{F}}^2. \tag{11}$$

Step 1.1. Note that the tensors in $\mathcal{B}$ are having at most $(r_1 + r_1')r_2r_3$ nonzero entries, the set of matrices $\mathcal{B}$ can be viewed as vectors in $\mathbb{R}^d$, with $d = (r_1 + r_1')r_2r_3$. Denote

$$\mathcal{V} = \{V \in \mathbb{R}^d, \|V\|_\infty \leq \kappa\}.$$

Then

$$\inf_{\widehat{X}} \sup_{X^* \in \mathcal{B}} \mathbb{E}\left\|\widehat{X} - X^*\right\|_{\mathrm{F}}^2 = \inf_{\widehat{V}} \sup_{V \in \mathcal{V}} \mathbb{E}\|\widehat{V} - V\|^2.$$

Since by Lemma 8, $\inf_{\widehat{V}} \sup_{V \in \mathcal{V}} \mathbb{E}\|\widehat{V} - V\|^2 \geq c_2\kappa^2 d$, it follows that

$$\inf_{\widehat{X}} \sup_{X^* \in \mathcal{B}} \mathbb{E}\left\|\widehat{X} - X^*\right\|_{\mathrm{F}}^2 = \inf_{\widehat{V}} \sup_{V \in \mathcal{V}} \mathbb{E}\|\widehat{V} - V\|^2 \geq c_2\kappa^2 d = c_2\kappa^2(r_1 + r_1')r_2r_3.$$

Step 1.2. To show (9), there are two cases. **(a)** Suppose $\xi^2 \geq r_2r_3\kappa^2$, then $r_1' = \lfloor \xi^2/(r_2r_3\kappa^2) \rfloor \geq 1$ and therefore

$$\inf_{\widehat{X}} \sup_{X^* \in \mathcal{T}^{\xi}_{(r_1,r_2,r_3)}} \mathbb{E}\left\|\widehat{X} - X^*\right\|_{\mathrm{F}}^2 \geq c_2\kappa^2(r_1 + r_1')r_2r_3 \geq c_2r_1r_2r_3 + c_3\xi^2.$$

(b) Suppose $\xi^2 < r_2r_3\kappa^2$. Then $r_1' = 0$

$$\inf_{\widehat{X}} \sup_{X^* \in \mathcal{T}^{\xi}_{(r_1,r_2,r_3)}} \mathbb{E}\left\|\widehat{X} - X^*\right\|_{\mathrm{F}}^2 \geq c_2\kappa^2 r_1r_2r_3 \geq \frac{c_2}{2}(\kappa^2 r_1r_2r_3 + \xi^2),$$

where the first inequality follows from **Step 1.1**, the second inequality follows from $\xi^2 < r_2r_3\kappa^2$.
Step 2. Note that

$$\inf_{\widehat{X}} \sup_{X^* \in \mathcal{T}^{\xi}_{(r_1,r_2,r_3)}} \mathbb{E}\left\|\widehat{X} - X^*\right\|_{\mathrm{F}}^2 \geq \inf_{\widehat{X}} \sup_{X^* \in \mathcal{T}_{(r_1,r_2,r_3)}} \mathbb{E}\left\|\widehat{X} - X^*\right\|_{\mathrm{F}}^2 \geq c_3\kappa^2 \sum_{k=1}^{3} p_kr_k, \tag{12}$$

where the second inequality follows from Zhang & Xia (2018). The desired result directly follows from (9) and (12). □

B.2 Matrix lower bound

Proof. Without loss of generality, suppose that $m \leq n$ and $r \leq m/2$. Therefore it suffices to show that

$$\inf_{\widehat{X}} \sup_{X^* \in \mathcal{M}^{\xi}_r} \mathbb{E}\left\|\widehat{X} - X^*\right\|_{\mathrm{F}}^2 \geq c_1\left\{\kappa^2 nr + \xi^2\right\}. \tag{13}$$

Let $r' = \lfloor \xi^2/(n\kappa^2) \rfloor$. Note that since $\xi^2 \leq (m - r)n\kappa^2$, it follows that

$$r' + r \leq m.$$

Step 1. Consider the set of matrices

$$\mathcal{B} = \{B \in \mathbb{R}^{m \times n} : \|B\|_\infty \leq \kappa^2, B_{i,j} = 0 \text{ if } i > r + r'\}.$$

In this step, we show that

$$\mathcal{B} \subset \mathcal{M}^{\xi}_r. \tag{14}$$

To see this, let $B \in \mathcal{B}$, and denote B' to be the matrix such that the first r rows of B' equals the first r rows of B, and the rest of $(m - r)$ rows of B' are all 0. Then since B' has at most r rows that are nonzero, $\mathrm{rank}(B') \leq r$ and

$$\|B - B'\|_{\mathrm{F}}^2 \leq r'n\|B\|_\infty = r'n\kappa^2 \leq \xi^2.$$

Therefore $\mathcal{B} \subset \mathcal{M}_r^{\xi}$. So

$$\inf_{\widehat{X}} \sup_{X^* \in \mathcal{M}_r^{\xi}} \mathbb{E}\left\|\widehat{X} - X^*\right\|_{\mathrm{F}}^2 \geq \inf_{\widehat{X}} \sup_{X^* \in \mathcal{B}} \mathbb{E}\left\|\widehat{X} - X^*\right\|_{\mathrm{F}}^2. \tag{15}$$

Step 2. Note that matrices in $\mathcal{B}$ are having at most $(r + r')n$ nonzero entries, the set of matrices $\mathcal{B}$ can be viewed as vectors in $\mathbb{R}^d$, with $d = (r + r')n$. Denote

$$\mathcal{V} = \{V \in \mathbb{R}^d, \|V\|_\infty \leq \kappa\}.$$

Then

$$\inf_{\widehat{X}} \sup_{X^* \in \mathcal{B}} \mathbb{E}\left\|\widehat{X} - X^*\right\|_{\mathrm{F}}^2 = \inf_{\widehat{V}} \sup_{V \in \mathcal{V}} \mathbb{E}\|\widehat{V} - V\|^2.$$

Since by Lemma 8, $\inf_{\widehat{V}} \sup_{V \in \mathcal{V}} \mathbb{E}\|\widehat{V} - V\|^2 \geq c_2\kappa^2 d$, it follows that

$$\inf_{\widehat{X}} \sup_{X^* \in \mathcal{M}_r^{\xi}} \mathbb{E}\left\|\widehat{X} - X^*\right\|_{\mathrm{F}}^2 \geq \inf_{\widehat{X}} \sup_{X^* \in \mathcal{B}} \mathbb{E}\left\|\widehat{X} - X^*\right\|_{\mathrm{F}}^2 = \inf_{\widehat{V}} \sup_{V \in \mathcal{V}} \mathbb{E}\|\widehat{V} - V\|^2 \geq c_2\kappa^2 d = c_2\kappa^2(r + r')n.$$

Step 3. There are two cases. **(a)** Suppose $\xi^2 \geq n\kappa^2$, then $r' = \lfloor \xi^2/(n\kappa^2) \rfloor \geq 1$ and therefore

$$\inf_{\widehat{X}} \sup_{X^* \in \mathcal{M}_r^{\xi}} \mathbb{E}\left\|\widehat{X} - X^*\right\|_{\mathrm{F}}^2 \geq c_2\kappa^2(r + r')n \geq c_2 rn + c_3\xi^2.$$

(b) Suppose $\xi^2 < n\kappa^2$. Then $r' = 0$

$$\inf_{\widehat{X}} \sup_{X^* \in \mathcal{M}_r^{\xi}} \mathbb{E}\left\|\widehat{X} - X^*\right\|_{\mathrm{F}}^2 \geq c_2\kappa^2 rn \geq \frac{c_2}{2}(\kappa^2 rn + \xi^2), \tag{16}$$

where the first inequality follows from **Step 1** and **Step 2**, the second inequality follows from $\xi^2 < n\kappa^2$. □

B.3 Technical results

Lemma 8. *Let $\kappa > 0$ and $d \in \mathbb{Z}^+$, and denote*

$$\mathcal{V} = \{V \in \mathbb{R}^d, \|V\|_\infty \leq \kappa\}.$$

Consider the model $Y \sim \mathcal{N}(V, \kappa^2 I_d)$ where $V \in \mathcal{V}$. Then there exists a universal constant c such that

$$\inf_{\widehat{V}} \sup_{V \in \mathcal{V}} \mathbb{E}\|\widehat{V} - V\|^2 \geq c\kappa^2 d,$$

where infimum is taken over all the estimators $\widehat{V}$ based on Y.

Proof. This is a standard minimax lower bound for bounded normal means; see for example Berry (1990) and Donoho et al. (1990). □

C Proof of Theorem 1

Many of the proof techniques employed in this work build on ideas developed in the existing tensor literature, including Vannieuwenhoven et al. (2012); Zhang & Xia (2018); Han et al. (2022).

Proof of Theorem 1. For any orthogonal matrix $U \in \mathbb{O}^{p \times r}$, we denote $U_\perp \in \mathbb{O}^{p \times (p-r)}$ to be the orthogonal complement of U, and

$$\mathcal{P}_U = UU^\top \quad \text{and} \quad \mathcal{P}_{U_\perp} = U_\perp U_\perp^\top = I_p - \mathcal{P}_U.$$

For $k \in \{1, 2, 3\}$, let $U_k^* \in \mathbb{O}^{p_k \times r_k}$ be the matrix whose columns corresponds to the top r_k singular vectors of $\mathcal{M}_k(X^*)$.
Throughout this proof, we assume the following good events hold:

$$\sup_{\substack{A \in \mathbb{R}^{p_1 \times p_2 \times p_3}, \\ \|A\|_{\mathrm{F}} \leq 1, A \in \mathcal{T}_{(r_1, r_2, r_3)}}} \langle Z, A \rangle \leq C\kappa\sqrt{r_1 r_2 r_3 + p_1 r_1 + p_2 r_2 + p_3 r_3}; \tag{17}$$

$$\|\mathcal{M}_1(Z)\cdot W_2\otimes W_3\| \le C\kappa\left(\sqrt{p_1+s_2s_3}\right) \text{ for non-random } W_2\in\mathbb{O}^{p_2\times s_2},\ W_3\in\mathbb{O}^{p_3\times s_3}; \quad (18)$$

$$\left\|\sin\Theta(U_k^{(0)},U_k^*)\right\| = \|U_k^{*\top}U_k^{(0)}\| \le \frac{1}{2\sqrt{r_{\max}}} \text{ for } k\in\{1,2,3\}. \quad (19)$$

Indeed in Lemma 10, Lemma 11 and Corollary 17, we show that (17), (18), and (19) hold with probability at least $1-C\exp(-cp_{\min})$.

Note that

$$\begin{aligned}
&\left\|\widetilde{X}-X^*\right\|_{\mathrm{F}} = \left\|Y\times_1\mathcal{P}_{U_1^{(1)}}\times_2\mathcal{P}_{U_2^{(1)}}\times_3\mathcal{P}_{U_3^{(1)}}-X^*\right\|_{\mathrm{F}}\\
&\le \left\|X^*\times_1\mathcal{P}_{U_1^{(1)}}\times_2\mathcal{P}_{U_2^{(1)}}\times_3\mathcal{P}_{U_3^{(1)}}-X^*\right\|_{\mathrm{F}} + \left\|Z\times_1\mathcal{P}_{U_1^{(1)}}\times_2\mathcal{P}_{U_2^{(1)}}\times_3\mathcal{P}_{U_3^{(1)}}\right\|_{\mathrm{F}}\\
&= \mathrm{I}_1+\mathrm{I}_2.
\end{aligned}$$

Step 1. For the term I_2, observe that

$$\begin{aligned}
\mathrm{I}_2 &= \left\|Z\times_1\mathcal{P}_{U_1^{(1)}}\times_2\mathcal{P}_{U_2^{(1)}}\times_3\mathcal{P}_{U_3^{(1)}}\right\|_{\mathrm{F}}\\
&= \sup_{W\in\mathbb{R}^{p_1\times p_2\times p_3},\|W\|_{\mathrm{F}}\le 1}\left\langle Z\times_1\mathcal{P}_{U_1^{(1)}}\times_2\mathcal{P}_{U_2^{(1)}}\times_3\mathcal{P}_{U_3^{(1)}},W\right\rangle\\
&= \sup_{W\in\mathbb{R}^{p_1\times p_2\times p_3},\|W\|_{\mathrm{F}}\le 1}\left\langle Z, W\times_1\mathcal{P}_{U_1^{(1)}}\times_2\mathcal{P}_{U_2^{(1)}}\times_3\mathcal{P}_{U_3^{(1)}}\right\rangle\\
&\le \sup_{\substack{A\in\mathbb{R}^{p_1\times p_2\times p_3},\\ \|A\|_{\mathrm{F}}\le 1, A\in\mathcal{T}_{(r_1,r_2,r_3)}}}\langle Z,A\rangle \le C\kappa\left(r_1r_2r_3+\sum_{k=1}^3 p_kr_k\right)^{1/2},
\end{aligned}$$

where the second equality follows from the duality of the Frobenius norm, and the last inequality follows from (17).

Step 2. For the term I_1, we have that

$$\begin{aligned}
\mathrm{I}_1 &= \left\|X^*\times_1\mathcal{P}_{U_1^{(1)}}\times_2\mathcal{P}_{U_2^{(1)}}\times_3\mathcal{P}_{U_3^{(1)}}-X^*\right\|_{\mathrm{F}}\\
&\le \left\|X^*\times_1(I_{p_1}-\mathcal{P}_{U_1^{(1)}})\right\|_{\mathrm{F}} + \left\|X^*\times_1\mathcal{P}_{U_1^{(1)}}\times_2(I_{p_2}-\mathcal{P}_{U_2^{(1)}})\right\|_{\mathrm{F}}\\
&\quad + \left\|X^*\times_1\mathcal{P}_{U_1^{(1)}}\times_2\mathcal{P}_{U_2^{(1)}}\times_3(I_{p_3}-\mathcal{P}_{U_3^{(1)}})\right\|_{\mathrm{F}}\\
&\le \sum_{k=1}^3\left\|X^*\times_k(I_{p_k}-\mathcal{P}_{U_k^{(1)}})\right\|_{\mathrm{F}},
\end{aligned}$$

where the last inequality follows from the observation that

$$\left\|X^*\times_1\mathcal{P}_{U_1^{(1)}}\times_2(I_{p_2}-\mathcal{P}_{U_2^{(1)}})\right\|_{\mathrm{F}} \le \left\|X^*\times_2(I_{p_2}-\mathcal{P}_{U_2^{(1)}})\right\|_{\mathrm{F}}\|\mathcal{P}_{U_1^{(1)}}\| \le \left\|X^*\times_2(I_{p_2}-\mathcal{P}_{U_2^{(1)}})\right\|_{\mathrm{F}}.$$

We only consider the case when $k=1$, since the same arguments apply for $k=2,3$. We have that

$$\begin{aligned}
&\left\|X^*\times_1(I_{p_1}-\mathcal{P}_{U_1^{(1)}})\right\|_{\mathrm{F}}\\
&\le \left\|X^*\times_1(I_{p_1}-\mathcal{P}_{U_1^{(1)}})\times_2\mathcal{P}_{U_2^*}\right\|_{\mathrm{F}} + \left\|X^*\times_1(I_{p_1}-\mathcal{P}_{U_1^{(1)}})\times_2(I_{p_2}-\mathcal{P}_{U_2^*})\right\|_{\mathrm{F}}\\
&\le \left\|X^*\times_1(I_{p_1}-\mathcal{P}_{U_1^{(1)}})\times_2\mathcal{P}_{U_2^*}\right\|_{\mathrm{F}} + \left\|X^*\times_2(I_{p_2}-\mathcal{P}_{U_2^*})\right\|_{\mathrm{F}}\\
&\le \left\|X^*\times_1(I_{p_1}-\mathcal{P}_{U_1^{(1)}})\times_2\mathcal{P}_{U_2^*}\times_3\mathcal{P}_{U_3^*}\right\|_{\mathrm{F}} + \left\|X^*\times_1(I_{p_1}-\mathcal{P}_{U_1^{(1)}})\times_2\mathcal{P}_{U_2^*}\times_3(I_{p_3}-\mathcal{P}_{U_3^*})\right\|_{\mathrm{F}}\\
&+ \left\|X^*\times_2(I_{p_2}-\mathcal{P}_{U_2^*})\right\|_{\mathrm{F}}\\
&\le \left\|X^*\times_1(I_{p_1}-\mathcal{P}_{U_1^{(1)}})\times_2\mathcal{P}_{U_2^*}\times_3\mathcal{P}_{U_3^*}\right\|_{\mathrm{F}} + \left\|X^*\times_3(I_{p_3}-\mathcal{P}_{U_3^*})\right\|_{\mathrm{F}} + \left\|X^*\times_2(I_{p_2}-\mathcal{P}_{U_2^*})\right\|_{\mathrm{F}}
\end{aligned}$$

$$= \left\| X^* \times_1 (I_{p_1} - \mathcal{P}_{U_1^{(1)}}) \times_2 \mathcal{P}_{U_2^*} \times_3 \mathcal{P}_{U_3^*} \right\|_{\mathrm{F}} + 2\xi_{(r_1,r_2,r_3)},$$

where the second follows from that $\|I_{p_1} - \mathcal{P}_{U_1^{(1)}}\| \le 1$, and the equality follows from Lemma 29.

Step 3. We bound $\|X^* \times_1 (I_{p_1} - \mathcal{P}_{U_1^{(1)}}) \times_2 \mathcal{P}_{U_2^*} \times_3 \mathcal{P}_{U_3^*}\|_{\mathrm{F}}$ in this step. Consider the different but relevant quantity

$$\begin{aligned}
&\left\| X^* \times_1 (I_{p_1} - \mathcal{P}_{U_1^{(1)}}) \times_2 \mathcal{P}_{U_2^{(0)}} \times_3 \mathcal{P}_{U_3^{(0)}} \right\|_{\mathrm{F}} \\
&= \left\| (I_{p_1} - \mathcal{P}_{U_1^{(1)}}) \cdot \mathcal{M}_1(X^*) \cdot (\mathcal{P}_{U_2^{(0)}} \otimes \mathcal{P}_{U_3^{(0)}}) \right\|_{\mathrm{F}} \\
&= \left\| (I_{p_1} - \mathcal{P}_{U_1^{(1)}}) \cdot \mathcal{M}_1(X^*) \cdot (U_2^{(0)} \otimes U_3^{(0)})(U_2^{(0)} \otimes U_3^{(0)})^\top \right\|_{\mathrm{F}} \\
&= \left\| (I_{p_1} - \mathcal{P}_{U_1^{(1)}}) \cdot \mathcal{M}_1(X^*) \cdot (U_2^{(0)} \otimes U_3^{(0)}) \right\|_{\mathrm{F}} \\
&\ge \left\| (I_{p_1} - \mathcal{P}_{U_1^{(1)}}) \cdot \mathcal{M}_1(X^*) \cdot \mathcal{P}_{U_2^* \otimes U_3^*} \cdot (U_2^{(0)} \otimes U_3^{(0)}) \right\|_{\mathrm{F}} \\
&\quad - \left\| (I_{p_1} - \mathcal{P}_{U_1^{(1)}}) \cdot \mathcal{M}_1(X^*) \cdot (I_{p_2p_3} - \mathcal{P}_{U_2^* \otimes U_3^*}) \cdot (U_2^{(0)} \otimes U_3^{(0)}) \right\|_{\mathrm{F}} \\
&= \mathrm{II}_1 - \mathrm{II}_2,
\end{aligned}$$

where the first inequality follows from that $(U_2^{(0)} \otimes U_3^{(0)}) \in \mathbb{O}^{(p_2p_3)\times(r_2r_3)}$, and Lemma 22.
Before analyzing the terms II_1 and II_2, we firstly note that

$$\begin{aligned}
(U_2^* \otimes U_3^*)^\top \cdot (U_2^{(0)} \otimes U_3^{(0)}) &= (U_2^{*\top} U_2^{(0)}) \otimes (U_3^{*\top} U_3^{(0)}), \\
\sigma_{\min}\left((U_2^{*\top} U_2^{(0)}) \otimes (U_3^{*\top} U_3^{(0)})\right) &= \sigma_{\min}\left(U_2^{*\top} U_2^{(0)}\right) \sigma_{\min}\left(U_3^{*\top} U_3^{(0)}\right), \\
\sigma_{\min}^2\left(U_k^{*\top} U_k^{(0)}\right) = 1 - \left\| U_{k\perp}^{*\top} U_k^{(0)} \right\|^2 &= 1 - \left\| \sin\Theta(U_k^*, U_k^{(0)}) \right\|^2,
\end{aligned} \tag{20}$$

which hold due to the properties of the Kronecker product and Lemma 25.
For the term II_1, we have that

$$\begin{aligned}
\mathrm{II}_1 &= \left\| (I_{p_1} - \mathcal{P}_{U_1^{(1)}}) \cdot \mathcal{M}_1(X^*) \cdot (U_2^* \otimes U_3^*) \cdot (U_2^* \otimes U_3^*)^\top \cdot (U_2^{(0)} \otimes U_3^{(0)}) \right\|_{\mathrm{F}} \\
&\ge \left\| (I_{p_1} - \mathcal{P}_{U_1^{(1)}}) \cdot \mathcal{M}_1(X^*) \cdot (U_2^* \otimes U_3^*) \right\|_{\mathrm{F}} \sigma_{\min}(U_2^{*\top} U_2^{(0)}) \sigma_{\min}(U_3^{*\top} U_3^{(0)}) \\
&= \left\| X^* \times_1 (I_{p_1} - \mathcal{P}_{U_1^{(1)}}) \times_2 \mathcal{P}_{U_2^*} \times_3 \mathcal{P}_{U_3^*} \right\|_{\mathrm{F}} \sqrt{\left(1 - \left\| \sin\Theta(U_2^*, U_2^{(0)}) \right\|^2\right)\left(1 - \left\| \sin\Theta(U_3^*, U_3^{(0)}) \right\|^2\right)} \\
&\ge \frac{3}{4} \left\| X^* \times_1 (I_{p_1} - \mathcal{P}_{U_1^{(1)}}) \times_2 \mathcal{P}_{U_2^*} \times_3 \mathcal{P}_{U_3^*} \right\|_{\mathrm{F}}
\end{aligned}$$

where the first inequality follows from Lemma 21, and the second equality follows from (20), and the last inequality follows from (19).

For the term II_2, we have that

$$\begin{aligned}
\mathrm{II}_2 &= \left\| (I_{p_1} - \mathcal{P}_{U_1^{(1)}}) \cdot \mathcal{M}_1(X^*) \cdot (I_{p_2p_3} - \mathcal{P}_{U_2^* \otimes U_3^*}) \cdot (U_2^{(0)} \otimes U_3^{(0)}) \right\|_{\mathrm{F}} \\
&= \left\| (I_{p_1} - \mathcal{P}_{U_1^{(1)}}) \cdot \mathcal{M}_1(X^*) \cdot (U_2^* \otimes U_3^*)_\perp \cdot (U_2^* \otimes U_3^*)_\perp^\top \cdot (U_2^{(0)} \otimes U_3^{(0)}) \right\|_{\mathrm{F}} \\
&\le \left\| I_{p_1} - \mathcal{P}_{U_1^{(1)}} \right\| \|\mathcal{M}_1(X^*) \cdot (U_2^* \otimes U_3^*)_\perp\|_{\mathrm{F}} \left\| (U_2^* \otimes U_3^*)_\perp^\top \cdot (U_2^{(0)} \otimes U_3^{(0)}) \right\| \\
&= \|\mathcal{M}_1(X^*) \cdot (U_2^* \otimes U_3^*)_\perp\|_{\mathrm{F}}.
\end{aligned}$$

Note that

$$\begin{aligned}
&\|\mathcal{M}_1(X^*) \cdot (U_2^* \otimes U_3^*)_\perp\|_{\mathrm{F}} \\
&= \|\mathcal{M}_1(X^*) \cdot [U_{2\perp}^* \otimes U_3^* \quad U_2^* \otimes U_{3\perp}^* \quad U_{2\perp}^* \otimes U_{3\perp}^*]\|_{\mathrm{F}} \\
&= \sqrt{\|\mathcal{M}_1(X^*) \cdot (U_{2\perp}^* \otimes U_3^*)\|_{\mathrm{F}}^2 + \|\mathcal{M}_1(X^*) \cdot (U_2^* \otimes U_{3\perp}^*)\|_{\mathrm{F}}^2 + \|\mathcal{M}_1(X^*) \cdot (U_{2\perp}^* \otimes U_{3\perp}^*)\|_{\mathrm{F}}^2}
\end{aligned}$$

$$
\begin{aligned}
&\leq \|\mathcal{M}_1(X^*)\cdot(U_{2\perp}^*\otimes U_3^*)\|_{\mathrm{F}} + \|\mathcal{M}_1(X^*)\cdot(U_2^*\otimes U_{3\perp}^*)\|_{\mathrm{F}} + \|\mathcal{M}_1(X^*)\cdot(U_{2\perp}^*\otimes U_{3\perp}^*)\|_{\mathrm{F}} \\
&= \|X^*\times_2 U_{2\perp}^*\times_3 U_3^*\|_{\mathrm{F}} + \|X^*\times_2 U_2^*\times_3 U_{3\perp}^*\|_{\mathrm{F}} + \|X^*\times_2 U_{2\perp}^*\times_3 U_{3\perp}^*\|_{\mathrm{F}} \\
&\leq \|X^*\times_2 U_{2\perp}^*\|_{\mathrm{F}} + \|X^*\times_3 U_{3\perp}^*\|_{\mathrm{F}} + \|X^*\times_2 U_{2\perp}^*\|_{\mathrm{F}} \\
&\leq 3\xi_{(r_1,r_2,r_3)}
\end{aligned}
$$

where the last inequality follows from Lemma 29. Therefore,

$$
\mathrm{II}_2 \leq 3\xi_{(r_1,r_2,r_3)}.
$$

Combining II_1 and II_2, we have

$$
\left\|X^*\times_1(I_{p_1}-\mathcal{P}_{U_1^{(1)}})\times_2\mathcal{P}_{U_2^*}\times_3\mathcal{P}_{U_3^*}\right\|_{\mathrm{F}} \leq \frac{4}{3}\left\|X^*\times_1(I_{p_1}-\mathcal{P}_{U_1^{(1)}})\times_2\mathcal{P}_{U_2^{(0)}}\times_3\mathcal{P}_{U_3^{(0)}}\right\|_{\mathrm{F}} + 4\xi_{(r_1,r_2,r_3)}.
$$

Step 4. We bound $\left\|X^*\times_1(I_{p_1}-\mathcal{P}_{U_1^{(1)}})\times_2\mathcal{P}_{U_2^{(0)}}\times_3\mathcal{P}_{U_3^{(0)}}\right\|_{\mathrm{F}}$ in this step. Note that

$$
\begin{aligned}
&\|X^*\times_1(I_{p_1}-\mathcal{P}_{U_1^{(1)}})\times_2\mathcal{P}_{U_2^{(0)}}\times_3\mathcal{P}_{U_3^{(0)}}\|_{\mathrm{F}} \\
=&\|\mathcal{P}_{U_{1\perp}^{(1)}}\cdot\mathcal{M}_1(X^*)\cdot(\mathcal{P}_{U_2^{(0)}}\otimes\mathcal{P}_{U_3^{(0)}})\|_{\mathrm{F}} = \|\mathcal{P}_{U_{1\perp}^{(1)}}\cdot\mathcal{M}_1(X^*)\cdot(U_2^{(0)}\otimes U_3^{(0)})\|_{\mathrm{F}}
\end{aligned}
$$

It suffices to apply Lemma 23 to bound

$$
\|\mathcal{P}_{U_{1\perp}^{(1)}}\cdot\mathcal{M}_1(X^*)\cdot(U_2^{(0)}\otimes U_3^{(0)})\|_{\mathrm{F}}.
$$

Since $U_{1\perp}^{(1)}$ corresponds to the SVD of $\mathcal{M}_1(Y)\cdot(U_2^{(0)}\otimes U_3^{(0)})$, let

$$
A=\mathcal{M}_1(Y)\cdot(U_2^{(0)}\otimes U_3^{(0)}) \quad\text{and}\quad B=\mathcal{M}_1(X^*)\cdot(U_2^{(0)}\otimes U_3^{(0)}).
$$

It follows from Lemma 23 that

$$
\begin{aligned}
\left\|X^*\times_1(I_{p_1}-\mathcal{P}_{U_1^{(1)}})\times_2\mathcal{P}_{U_2^{(0)}}\times_3\mathcal{P}_{U_3^{(0)}}\right\|_{\mathrm{F}} =&\|\mathcal{P}_{U_{1\perp}^{(1)}}\cdot\mathcal{M}_1(X^*)\cdot(U_2^{(0)}\otimes U_3^{(0)})\|_{\mathrm{F}} \\
\leq& C_1\|B-B_{r_1}\|_F + C_2\sqrt{r_1}\|A-B\|. \qquad (21)
\end{aligned}
$$

Here

$$
\begin{aligned}
\|A-B\| &= \left\|\mathcal{M}_1(Z)\cdot U_2^{(0)}\otimes U_3^{(0)}\right\| \\
&= \left\|\mathcal{M}_1(Z)\cdot(U_2^*U_2^{*\top}+U_{2\perp}^*U_{2\perp}^{\top *})U_2^{(0)}\otimes U_3^{(0)}\right\| \\
&\leq \left\|\mathcal{M}_1(Z)\cdot(U_2^*U_2^{*\top}U_2^{(0)})\otimes U_3^{(0)}\right\| + \left\|\mathcal{M}_1(Z)\cdot(U_{2\perp}^*U_{2\perp}^{\top *}U_2^{(0)})\otimes U_3^{(0)}\right\| \\
&\leq \left\|\mathcal{M}_1(Z)\cdot U_2^*\otimes U_3^{(0)}\right\|\|U_2^{*\top}U_2^{(0)}\| + \left\|\mathcal{M}_1(Z)\cdot U_{2\perp}^*\otimes U_3^{(0)}\right\|\|U_{2\perp}^{\top *}U_2^{(0)}\|. \qquad (22)
\end{aligned}
$$

Note that

$$
\begin{aligned}
&\left\|\mathcal{M}_1(Z)\cdot U_2^*\otimes U_3^{(0)}\right\|\|U_2^{*\top}U_2^{(0)}\| = \left\|\mathcal{M}_1(Z)\cdot U_2^*\otimes(U_3^*U_3^{*\top}+U_{3\perp}^*U_{3\perp}^{\top *})U_3^{(0)}\right\| \\
\leq& \left\|\mathcal{M}_1(Z)\cdot U_2^*\otimes(U_3^*U_3^{*\top}U_3^{(0)})\right\| + \left\|\mathcal{M}_1(Z)\cdot U_2^*\otimes(U_{3\perp}^*U_{3\perp}^{\top *})U_3^{(0)}\right\| \\
\leq& \|\mathcal{M}_1(Z)\cdot U_2^*\otimes U_3^*\|\,\|U_3^{*\top}U_3^{(0)}\| + \|\mathcal{M}_1(Z)\cdot U_2^*\otimes U_{3\perp}^*\|\,\|U_{3\perp}^{\top *}U_3^{(0)}\| \\
\leq& C_4\kappa(\sqrt{p_1+r_2r_3}) + C_5\kappa(\sqrt{p_1+p_3r_2})r_{\max}^{-1/2}, \qquad (23)
\end{aligned}
$$

where the last inequality follows from (18) and the fact that $\|U_2^{*\top}U_2^{(0)}\|\leq 1$. In addition

$$
\begin{aligned}
&\left\|\mathcal{M}_1(Z)\cdot U_{2\perp}^*\otimes U_3^{(0)}\right\|\|U_{2\perp}^{\top *}U_2^{(0)}\| \leq \left\|\mathcal{M}_1(Z)\cdot U_{2\perp}^*\otimes(U_3^*U_3^{\top *}+U_{3\perp}^*U_{3\perp}^{\top *})U_3^{(0)}\right\|\|U_{2\perp}^{\top *}U_2^{(0)}\| \\
\leq& \left\|\mathcal{M}_1(Z)\cdot U_{2\perp}^*\otimes(U_3^*U_3^{*\top}U_3^{(0)})\right\|\|U_{2\perp}^{\top *}U_2^{(0)}\| + \left\|\mathcal{M}_1(Z)\cdot U_2^*\otimes(U_{3\perp}^*U_{3\perp}^{\top *}U_3^{(0)})\right\|\|U_{2\perp}^{\top *}U_2^{(0)}\| \\
\leq& \|\mathcal{M}_1(Z)\cdot U_{2\perp}^*\otimes U_3^*\|\,\|U_3^{*\top}U_3^{(0)}\|\|U_{2\perp}^{\top *}U_2^{(0)}\| + \|\mathcal{M}_1(Z)\cdot U_2^*\otimes U_{3\perp}^*\|\,\|U_{3\perp}^{\top *}U_3^{(0)}\|\|U_{2\perp}^{\top *}U_2^{(0)}\| \\
\leq& C\kappa(\sqrt{p_1+p_2r_3})r_{\max}^{-1/2} + C\kappa(\sqrt{p_1+p_3r_2})r_{\max}^{-1}. \qquad (24)
\end{aligned}
$$

Therefore (22), (23) and (24) leads to

$$\|A - B\| \le C_4\kappa(\sqrt{p_1} + \sqrt{r_2 r_3} + \sqrt{p_2 r_3} r_{\max}^{-1/2} + \sqrt{p_3 r_2} r_{\max}^{-1/2}). \tag{25}$$

In addition,

$$\begin{aligned}
\|B - B_{r_1}\|_F =& \|\mathcal{M}_1(X^*) \cdot U_2^{(0)} \otimes U_3^{(0)} - \{\mathcal{M}_1(X^*) \cdot U_2^{(0)} \otimes U_3^{(0)}\}_{r_1}\|_{\mathrm{F}} \\
\le& \|\mathcal{M}_1(X^*) \cdot U_2^{(0)} \otimes U_3^{(0)} - \{\mathcal{M}_1(X^*)\}_{r_1} \cdot U_2^{(0)} \otimes U_3^{(0)}\|_{\mathrm{F}} \\
\le& \|(\mathcal{M}_1(X^*) - \{\mathcal{M}_1(X^*)\}_{r_1}) U_2^{(0)} \otimes U_3^{(0)}\|_{\mathrm{F}} \\
\le& \|\mathcal{M}_1(X^*) - \{\mathcal{M}_1(X^*)\}_{r_1}\|_{\mathrm{F}} = \sqrt{\sum_{j=r_1+1}^{\mathrm{rank}(\mathcal{M}_1(X^*))} \sigma_j^2(\mathcal{M}_1(X^*))} \le \xi_{(r_1,r_2,r_3)}.
\end{aligned} \tag{26}$$

Here $\{\mathcal{M}_1(X^*) \cdot U_2^{(0)} \otimes U_3^{(0)}\}_{r_1}$ indicate the best rank r_1 estimate of the matrix $\mathcal{M}_1(X^*) \cdot U_2^{(0)} \otimes U_3^{(0)}$ in the first equality, and so for any rank r_1 matrix Φ,

$$\|\mathcal{M}_1(X^*) \cdot U_2^{(0)} \otimes U_3^{(0)} - \{\mathcal{M}_1(X^*) \cdot U_2^{(0)} \otimes U_3^{(0)}\}_{r_1}\|_{\mathrm{F}} \le \|\mathcal{M}_1(X^*) \cdot U_2^{(0)} \otimes U_3^{(0)} - \Phi\|_{\mathrm{F}};$$

the second inequality holds because $\{\mathcal{M}_1(X^*)\}_{r_1} \cdot U_2^{(0)} \otimes U_3^{(0)}$ is at most rank r_1, and the last inequality follows from Lemma 29.

It follows from (21), (25) and (26) that

$$\left\| X^* \times_1 (I_{p_1} - \mathcal{P}_{U_1^{(1)}}) \times_2 \mathcal{P}_{U_2^{(0)}} \times_3 \mathcal{P}_{U_3^{(0)}} \right\|_{\mathrm{F}} \le C_5\kappa(\sqrt{p_1 r_1} + \sqrt{r_1 r_2 r_3} + \sqrt{p_2 r_2} + \sqrt{p_3 r_3}) + C_5 \xi_{(r_1,r_2,r_3)}.$$

Step 5. The conclusions of **Step 3** and **Step 4** lead to

$$\left\| X^* \times_1 (I_{p_1} - \mathcal{P}_{U_1^{(1)}}) \times_2 \mathcal{P}_{U_2^*} \times_3 \mathcal{P}_{U_3^*} \right\|_{\mathrm{F}} \le C_6\kappa(\sqrt{p_1 r_1} + \sqrt{r_1 r_2 r_3} + \sqrt{p_2 r_2} + \sqrt{p_3 r_3}) + C_6 \xi_{(r_1,r_2,r_3)}.$$

This bound together with **Step 1** and **Step 2** leads to

$$\left\| \widetilde{X} - X^* \right\|_{\mathrm{F}} \le C_7\kappa(\sqrt{p_1 r_1} + \sqrt{p_2 r_2} + \sqrt{p_3 r_3} + \sqrt{r_1 r_2 r_3}) + C_7 \xi_{(r_1,r_2,r_3)}.$$

□

D PROOF OF THEOREM 3

Proof of Theorem 3. By the triangle inequality,

$$\|Y_{(r)} - X^*\|_{\mathrm{F}} \le \|X^*_{(r)} - X^*\|_{\mathrm{F}} + \|Y_{(r)} - X^*_{(r)}\|_{\mathrm{F}} = \xi_{(r)} + \|Y_{(r)} - X^*_{(r)}\|_{\mathrm{F}}. \tag{27}$$

Since $\mathrm{rank}(Y_{(r)} - X^*_{(r)}) \le 2r$, we have

$$\|Y_{(r)} - X^*_{(r)}\|_{\mathrm{F}}^2 = \sum_{k=1}^{2r} \sigma_k^2(Y_{(r)} - X^*_{(r)}). \tag{28}$$

Write

$$Y_{(r)} - X^*_{(r)} = (X^* - X^*_{(r)}) + Z + (Y_{(r)} - Y).$$

It follows from Ky Fan's Theorem (Lemma 28) that,

$$\sqrt{\sum_{i=1}^{2r} \sigma_i^2(Y_{(r)} - X^*_{(r)})} \le \sqrt{\sum_{i=1}^{2r} \sigma_i^2(X^* - X^*_{(r)})} + \sqrt{\sum_{i=1}^{2r} \sigma_i^2(Z)} + \sqrt{\sum_{i=1}^{2r} \sigma_i^2(Y_{(r)} - Y)}.$$

Using the identity $\sigma_i(X^* - X^*_{(r)}) = \sigma_{r+i}(X^*)$ and likewise for Y, we obtain

$$\|Y_{(r)} - X^*_{(r)}\|_{\mathrm{F}} \le \sqrt{\sum_{i=1}^{2r} \sigma_{r+i}^2(X^*)} + \sqrt{\sum_{i=1}^{2r} \sigma_i^2(Z)} + \sqrt{\sum_{i=1}^{2r} \sigma_{r+i}^2(Y)}. \tag{29}$$

To further simplify the third term, we apply Weyl's inequality (Theorem 27) to obtain

$$\sigma_{r+i}(Y) \leq \sigma_{r+i}(X^*) + \sigma_1(Z),$$

which gives

$$\sigma_{r+i}^2(Y) \leq 2\left\{\sigma_{r+i}^2(X^*) + \sigma_1^2(Z)\right\}.$$

Substituting this into (29), we get

$$\begin{aligned}
\|Y_{(r)} - X_{(r)}^*\|_{\mathrm{F}} &\leq \sqrt{\sum_{i=1}^{2r}\sigma_{r+i}^2(X^*)} + \sqrt{\sum_{i=1}^{2r}\sigma_i^2(Z)} + \sqrt{2\sum_{i=1}^{2r}\sigma_{r+i}^2(X^*)} + 2\sqrt{r}\sigma_1(Z) \\
&\leq (1+\sqrt{2})\sqrt{\sum_{i=1}^{2r}\sigma_{r+i}^2(X^*)} + (2+\sqrt{2})\sqrt{r}\|Z\| \\
&= (1+\sqrt{2})\sqrt{\sum_{i=r+1}^{3r}\sigma_i^2(X^*)} + (2+\sqrt{2})\sqrt{r}\|Z\|,
\end{aligned}$$

where the second inequality follows from the fact that for any $k \geq 1$, we have $\sigma_k(Z) \leq \sigma_1(Z) = \|Z\|$. Substituting the bound of $\|Y_{(r)} - X_{(r)}^*\|_{\mathrm{F}}$ into (27), we obtain

$$\|Y_{(r)} - X^*\|_{\mathrm{F}} \leq (2+\sqrt{2})\sqrt{\sum_{i=r+1}^{\min\{m,n\}}\sigma_i^2(X^*)} + (2+\sqrt{2})\sqrt{r}\|Z\|.$$

□

E DEVIATION BOUNDS

Throughout this appendix we work with centered sub-Gaussian random variables with parameter κ^2, and we assume (without loss of generality) that each such variable X satisfies $\mathbb{E}X^2 = \kappa^2$. Any other case can be handled with additional constants in the bounds.

Lemma 9 (Theorem 4.4.3 in Vershynin (2018))**.** *Assume all the entries of $Z \in \mathbb{R}^{m\times n}$ are independent mean-zero sub-Gaussian random variables, i.e.*

$$\|Z_{ij}\|_{\psi_2} = \sup_{q\geq 1}\mathbb{E}(|Z_{ij}|^q)^{1/q}/q^{1/2} \leq \kappa.$$

Then there exist some universal constant $C > 0$, such that

$$\|Z\| \leq C\kappa\left(\sqrt{m} + \sqrt{n}\right)$$

with probability at least $1 - \exp(-(m+n))$.

Lemma 10. *Suppose all the entries of $Z \in \mathbb{R}^{p_1\times p_2\times p_3}$ are independent mean-zero sub-Gaussian random variables, i.e.*

$$\|Z_{ijk}\|_{\psi_2} = \sup_{q\geq 1}\mathbb{E}(|Z_{ijk}|^q)^{1/q}/q^{1/2} \leq \kappa.$$

Then there exist some universal constants $C, c > 0$, such that

$$\sup_{\substack{A\in\mathbb{R}^{p_1\times p_2\times p_3},\\ \|A\|_{\mathrm{F}}\leq 1, A\in\mathcal{T}_{(r_1,r_2,r_3)}}} \langle Z, A\rangle \leq C\kappa\left(r_1 r_2 r_3 + \sum_{k=1}^{3} p_k r_k\right)^{1/2}$$

with probability at least $1 - \exp(-c\sum_{k=1}^{3} p_k r_k)$.

Proof. This directly follows from Lemma E.5 in Han et al. (2022). □

Lemma 11. *Suppose all the entries of $Z \in \mathbb{R}^{p_1\times p_2\times p_3}$ are independent mean-zero sub-Gaussian random variables, i.e.*

$$\|Z_{ijk}\|_{\psi_2} = \sup_{q\geq 1}\mathbb{E}(|Z_{ijk}|^q)^{1/q}/q^{1/2} \leq \kappa.$$

Let $W_2 \in \mathbb{O}^{p_2\times s_2}$ and $W_3 \in \mathbb{O}^{p_3\times s_3}$ be non-random. Then there exists absolute positive constants C_1, C_2 and c such that

$$\mathbb{P}\Big(\big\|\mathcal{M}_1(Z)\,(W_2\otimes W_3)\big\| \geq C_1\kappa(\sqrt{p_1+s_2s_3})\Big) \leq C_2\exp\left(-cp_1\right),$$

Proof. It suffices to observe that $W_2\otimes W_3 \in \mathbb{O}^{p_2p_3\times r_2r_3}$. The desired result is a direct consequence of Lemma 13. □

Lemma 12. *Suppose all the entries of $Z \in \mathbb{R}^{p_1\times p_2\times p_3}$ are independent mean-zero sub-Gaussian random variables, i.e.*

$$\|Z_{ijk}\|_{\psi_2} = \sup_{q\geq 1}\mathbb{E}(|Z_{ijk}|^q)^{1/q}/q^{1/2} \leq \kappa.$$

Then there exists absolute positive constants C_1, C_2 and c such that

$$\mathbb{P}\Bigg(\sup_{\substack{V_2\in\mathbb{R}^{p_2\times r_2},\,\|V_2\|\leq 1\\ V_3\in\mathbb{R}^{p_3\times r_3},\,\|V_3\|\leq 1}}\big\|\mathcal{M}_1(Z)\,(V_2\otimes V_3)\big\| \geq C_1\kappa\Big(\sqrt{p_1+r_2r_3+p_2r_2+p_3r_3}\Big)\Bigg)$$
$$\leq C_2\exp\left(-c(p_1+p_2+p_3)\right),$$

Proof. It follows from the assumption that $(\mathcal{M}_1(Z))_{i,j} \overset{i.i.d}{\sim} \text{subGaussian}(0,\kappa^2)$.

Step 1. For fixed $V_2 \in \mathbb{R}^{p_2\times r_2}$ with $\|V_2\|\leq 1$ and $V_3\in\mathbb{R}^{p_3\times r_3}$ with $\|V_3\|\leq 1$, it follows that

$$\|V_2\otimes V_3\| = \|V_2\|\,\|V_3\| \leq 1,$$

We upper bound $\|\mathcal{M}_1(Z)\cdot(V_2\otimes V_3)\|$ for any fixed V_2 and V_3. Since,

$$\mathcal{M}(Z)\in\mathbb{R}^{p_1\times p_2p_3}, (\mathcal{M}(Z))_{i,j} \overset{i.i.d}{\sim} \text{subGaussian}(0,\kappa^2), \quad \text{and} \quad \|V_2\otimes V_3\| = \|V_2\|\,\|V_3\| = 1.$$

It follows from Lemma 13 that

$$\mathbb{P}\Big(\,\|\mathcal{M}_1(Z)\cdot(V_2\otimes V_3)\| > x\Big) \leq 2\exp\left(C(p_1+r_2r_3)-cx^2\kappa^{-2}\right).$$

Step 2. Let $\mathcal{N}_{p_2,r_2}(\epsilon)$ denote an ϵ net of the set

$$\{A\in\mathbb{R}^{p_2\times r_2} : \|A\|\leq 1\}$$

with respect to the operator norm $\|\cdot\|$. Similarly, let Let $\mathcal{N}_{p_3,r_3}(\epsilon)$ denote an ϵ net of the set

$$\{B\in\mathbb{R}^{p_3\times r_3} : \|B\|\leq 1\}$$

with respect to the operator norm $\|\cdot\|$. Denote the random quantity

$$\psi = \sup_{\substack{V_2\in\mathbb{R}^{p_2\times r_2},\,\|V_2\|\leq 1\\ V_3\in\mathbb{R}^{p_3\times r_3},\,\|V_3\|\leq 1}}\big\|\mathcal{M}_1(Z)\,(V_2\otimes V_3)\big\|.$$

For any given $V_2\in\mathbb{R}^{p_2\times r_2}$ and $V_3\in\mathbb{R}^q$ with $\|V_2\|\leq 1$ and $\|V_3\|\leq 1$, let $\widetilde{V}_2\in\mathcal{N}_{p_2,r_2}(1/4)$ and $\widetilde{V}_3\in\mathcal{N}_{p_3,r_3}(1/4)$ be such that

$$\|V_2-\widetilde{V}_2\|\leq 1/4 \quad\text{and}\quad \|V_3-\widetilde{V}_3\|\leq 1/4.$$

Then

$$\|\mathcal{M}_1(Z)V_2\otimes V_3\| \leq \|\mathcal{M}_1(Z)(V_2-\widetilde{V}_2)\otimes V_3\| + \|\mathcal{M}_1(Z)\widetilde{V}_2\otimes(V_3-\widetilde{V}_3)\| + \|\mathcal{M}_1(Z)\widetilde{V}_2\otimes\widetilde{V}_3\|.$$

Note that $\|V_2-\widetilde{V}_2\|\leq 1/4$. So

$$\|\mathcal{M}_1(Z)(V_2-\widetilde{V}_2)\otimes V_3\| = \frac{1}{4}\|\mathcal{M}_1(Z)\{4(V_2-\widetilde{V}_2)\}\otimes V_3\| \leq \frac{\psi}{4}.$$

Similarly

$$\|\mathcal{M}_1(Z)V_2 \otimes (V_3 - \widetilde{V}_3)\| \le \frac{\psi}{4}.$$

In addition,

$$\|\mathcal{M}_1(Z)\widetilde{V}_2 \otimes \widetilde{V}_3\| \le \sup_{V_2 \in \mathcal{N}_{p_2,r_2}(1/4), b \in \mathcal{N}_{p_3,r_3}(1/4)} \|\mathcal{M}_1(Z)V_2 \otimes V_3\|.$$

So for any V_2 and V_3,

$$\|\mathcal{M}_1(Z)V_2 \otimes V_3\| \le \frac{1}{2}\psi + \sup_{V_2 \in \mathcal{N}_{p_2,r_2}(1/4), b \in \mathcal{N}_{p_3,r_3}(1/4)} \|\mathcal{M}_1(Z)V_2 \otimes V_3\|.$$

Taking sup over all $V_2 \in \{A \in \mathbb{R}^{p_2 \times r_2} : \|A\| \le 1\}$ and $V_3 \in \{B \in \mathbb{R}^{p_3 \times r_3} : \|B\| \le 1\}$, it follows that

$$\psi \le \frac{1}{2}\psi + \sup_{V_2 \in \mathcal{N}_{p_2,r_2}(1/4), b \in \mathcal{N}_{p_3,r_3}(1/4)} \|\mathcal{M}_1(Z)V_2 \otimes V_3\|,$$

or simply

$$\psi \le 2 \sup_{V_2 \in \mathcal{N}_{p_2,r_2}(1/4), b \in \mathcal{N}_{p_3,r_3}(1/4)} \|\mathcal{M}_1(Z)V_2 \otimes V_3\|.$$

Step 3. By Proposition 8 in Pajor (1998), the cardinality of $\mathcal{N}_{p_2,r_2}(\epsilon)$ is bounded $(\frac{C}{\epsilon})^{p_2 r_2}$, and $\mathcal{N}_{p_3,r_3}(\epsilon)$ is bounded $(\frac{C}{\epsilon})^{p_3 r_3}$. Therefore

$$\begin{aligned}
\mathbb{P}(\psi \ge 2t) \le & \mathbb{P}\left(\sup_{V_2 \in \mathcal{N}_{p_2,r_2}(1/4), b \in \mathcal{N}_{p_3,r_3}(1/4)} \|\mathcal{M}_1(Z)V_2 \otimes V_3\| \ge t \right) \\
\le & C_2^{p_2 r_2} C_3^{p_3 r_3} \sup_{V_2 \in \mathcal{N}_{p_2,r_2}(1/4), b \in \mathcal{N}_{p_3,r_3}(1/4)} \mathbb{P}\left(\|\mathcal{M}_1(Z)V_2 \otimes V_3\| \ge t \right) \\
\le & 2 \exp\left(C(p_1 + r_2 r_3 + p_2 r_2 + p_3 r_3) - ct^2 \kappa^{-2} \right).
\end{aligned}$$

Here C and C_3 are positive constants. The desired result follows by noting

$$\psi = \sup_{\substack{V_2 \in \mathbb{R}^{p_2 \times r_2}, \|V_2\| \le 1 \\ V_3 \in \mathbb{R}^{p_3 \times r_3}, \|V_3\| \le 1}} \big\|\mathcal{M}_1(Z)\,(V_2 \otimes V_3)\big\|.$$

□

Lemma 13. *Suppose $Z \in \mathbb{R}^{n \times m}$, with $Z_{ij} \overset{i.i.d}{\sim} \mathrm{subGaussian}(0, \kappa^2)$. Let $A \in \mathbb{R}^{p \times n}$ and $B \in \mathbb{R}^{m \times q}$ be non-random matrices. Then for any $t > 0$*

$$\mathbb{P}\left(\|A\,Z\,B\| > t\right) \le C_1 \exp\left(C(p+q) - \frac{c\,t^2}{\kappa^2\ \|A\|^2\ \|B\|^2} \right). \tag{30}$$

Proof.
Step 1. Let $u \in \mathbb{R}^n$ and $v \in \mathbb{R}^m$ be non-random. Then that $u^\top Z\, v = \sum_{i=1}^n \sum_{j=1}^n u_i Z_{ij} v_j$. Since Z_{ij} are i.i.d. sub-Gaussian with parameter κ^2, it follows that $u^\top Z\, v$ is sub-Gaussian with parameter $\kappa^2 \|u\|_2^2 \|v\|_2^2$. Consequently by Hoeffding's inequality,

$$\mathbb{P}\left(\left|u^\top Z\, v\right| > t \right) \le 2 \exp\left(-\frac{c\,t^2}{\kappa^2\ \|u\|_2^2\ \|v\|_2^2} \right).$$

Step 2. Let $a \in \mathbb{R}^p$ and $b \in \mathbb{R}^q$ be non-random vectors such that $\|a\|_2, \|b\|_2 \le 1$. By **Step 1**, it follows that

$$\mathbb{P}\left(\left|a^\top AZ\, Bb\right| > t \right) \le 2 \exp\left(-\frac{ct^2}{\kappa^2\ \|Aa\|_2^2\ \|Bb\|_2^2} \right) \le 2 \exp\left(-\frac{ct^2}{\kappa^2\ \|A\|^2\ \|B\|^2} \right).$$

Step 3. Let $\mathcal{N}_p(\epsilon)$ be an ϵ-net of the unit ball in $\mathbb{R}^p$. It follows that for any $a \in \mathbb{R}^p$ with $\|a\|_2 = 1$, there exists $\widetilde{a} \in \mathcal{N}_p(\epsilon)$ such that

$$\|a - \widetilde{a}\|_2 \le \epsilon.$$

Similarly let $\mathcal{N}_q(\epsilon)$ be an ϵ-net of the unit ball in $\mathbb{R}^q$.

Denote the random quantity

$$\psi = \|AZB\| = \sup_{a\in\mathbb{R}^p, b\in\mathbb{R}^q \|a\|_2=\|b\|_2=1} |a^\top AZBb|.$$

For any given $a \in \mathbb{R}^p$ and $b \in \mathbb{R}^q$, let $\widetilde{a} \in \mathcal{N}_p(1/4)$ and $\widetilde{b} \in \mathcal{N}_q(1/4)$ be such that

$$\|a - \widetilde{a}\|_2 \le 1/4 \quad \text{and} \quad \|b - \widetilde{b}\|_2 \le 1/4.$$

Then

$$|a^\top AZBb| \le |(a-\widetilde{a})^\top AZBb| + |\widetilde{a}^\top AZB(\widetilde{b} - b)| + |\widetilde{a}^\top AZB\widetilde{b}|.$$

Note that $\|a - \widetilde{a}\| \le \frac{1}{4}$. So

$$|(a-\widetilde{a})^\top AZBb| = \frac{1}{4}|\{4(a-\widetilde{a})^\top\}AZBb| \le \frac{\psi}{4}.$$

Similarly

$$|\widetilde{a}^\top AZB(\widetilde{b} - b)| \le \frac{\psi}{4}.$$

In addition,

$$|\widetilde{a}^\top AZB\widetilde{b}| \le \sup_{a\in\mathcal{N}_p(1/4), b\in\mathcal{N}_q(1/4)} |a^\top AZBb|.$$

So for any a and b,

$$|a^\top AZBb| \le \frac{1}{2}\psi + \sup_{a\in N_p(1/4), b\in\mathcal{N}_q(1/4)} |a^\top AZBb|.$$

Taking sup over all unit vectors $a \in \mathbb{R}^p$ and $b \in \mathbb{R}^q$, it follows that

$$\psi \le \frac{1}{2}\psi + \sup_{a\in\mathcal{N}_p(1/4), b\in\mathcal{N}_q(1/4)} |a^\top AZBb|,$$

or simply

$$\psi \le 2 \sup_{a\in\mathcal{N}_p(1/4), b\in\mathcal{N}_q(1/4)} |a^\top AZBb|.$$

Step 4. By Vershynin (2018), the cardinality of $\mathcal{N}_p(\epsilon)$ is bounded by $(\frac{C}{\epsilon})^p$, and the cardinality of $\mathcal{N}_q(\epsilon)$ is bounded by $(\frac{C}{\epsilon})^q$. Therefore

$$\begin{aligned}\mathbb{P}(\psi \ge 2t) = \mathbb{P}\left(\sup_{a\in\mathcal{N}_p(1/4), b\in\mathcal{N}_q(1/4)} |a^\top AZBb| \ge t\right) \le & C_2{}^p C_2{}^q \sup_{a\in\mathcal{N}_p(1/4), b\in\mathcal{N}_q(1/4)} \mathbb{P}(|a^\top AZBb| \ge t) \\ \le & 2C_2{}^p C_2{}^q \exp\left(-\frac{ct^2}{\kappa^2 \|A\|^2 \|B\|^2}\right),\end{aligned}$$

where c, C are positive constants. The desired result follows by noting $\psi = \|AZB\|$. □

Lemma 14. *Suppose $Z \in \mathbb{R}^{n\times m}$, with $Z_{ij} \overset{i.i.d}{\sim} \text{subGaussian}(0, \kappa^2)$. Let $A \in \mathbb{R}^{m\times p}$ and $B \in \mathbb{R}^{m\times q}$ be non-random matrices. Then for any $t > 0$*

$$\mathbb{P}\left(\left\|A^\top Z^\top ZB - n\kappa^2 A^\top B\right\| > t\right) \le C_1 \exp\left(C(p+q) - \min\left(\frac{t^2}{n\,\kappa^4 \|B\|^2 \|A\|^2}, \frac{t}{\kappa^2 \|B\| \|A\|}\right)\right), \tag{31}$$

where C and C_1 are positive constants.

Proof. For any non-random $u, v \in \mathbb{R}^m$, it follows that

$$u^\top Z^\top Z v - n\kappa^2 u^\top v = \sum_{j=1}^{n} (u^\top Z_j)(v^\top Z_j) - \mathbb{E}\{(u^\top Z_j)(v^\top Z_j)\},$$

where Z_j is the j-th row of Z. Note that $(u^\top Z_j)$ is sub-Gaussian with parameter $\kappa^2\|u\|_2^2$, and $(v^\top Z_j)$ is sub-Gaussian with parameter $\kappa^2\|v\|_2^2$. Since Z have i.i.d. entries, it follows that $\{(u^\top Z_j)(v^\top Z_j)\}_{j=1}^n$ are i.i.d. sub-exponential with parameter $\kappa^4\|u\|_2^2\|v\|_2^2$. So

$$\mathbb{P}\bigg(|u^\top Z^\top Z v - n\kappa^2 u^\top v| \geq t\bigg) \leq 2\exp\bigg(-c\min\bigg\{\frac{t^2}{n\kappa^4\|u\|_2^2\|v\|_2^2}, \frac{t}{\kappa^2\|u\|_2\|v\|_2}\bigg\}\bigg).$$

Step 1. Let $\mathcal{N}_p(\epsilon)$ be the ϵ-net of the unit ball in $\mathbb{R}^p$. It follows that for any $a \in \mathbb{R}^p$ with $\|a\|_2 = 1$, there exists $\widetilde{a} \in \mathcal{N}_p(\epsilon)$ such that

$$\|a - \widetilde{a}\|_2 \leq \epsilon.$$

Similarly let $\mathcal{N}_q(\epsilon)$ be the ϵ-net of the unit ball in $\mathbb{R}^q$.

Denote the random quantity

$$\psi = \|A^\top Z^\top Z B - n\kappa^2 A^\top B\| = \sup_{a\in\mathbb{R}^p, b\in\mathbb{R}^q \|a\|_2=\|b\|_2=1} \big|a^\top(A^\top Z^\top Z B - n\kappa^2 A^\top B)b\big|.$$

For any given $a \in \mathbb{R}^p$ and $b \in \mathbb{R}^q$, let $\widetilde{a} \in \mathcal{N}_p(1/4)$ and $\widetilde{b} \in \mathcal{N}_q(1/4)$ be such that

$$\|a - \widetilde{a}\|_2 \leq 1/4 \quad \text{and} \quad \|b - \widetilde{b}\|_2 \leq 1/4.$$

Then

$$\begin{aligned} |a^\top(A^\top Z^\top Z B - n\kappa^2 A^\top B)b| &\leq |(a-\widetilde{a})^\top(A^\top Z^\top Z B - n\kappa^2 A^\top B)b| \\ +|\widetilde{a}^\top(A^\top Z^\top Z B - n\kappa^2 A^\top B)(\widetilde{b} - b)| &+ |\widetilde{a}^\top(A^\top Z^\top Z B - n\kappa^2 A^\top B)\widetilde{b}|. \end{aligned}$$

Note that $\|a - \widetilde{a}\|_2 \leq \frac{1}{4}$. So

$$|(a-\widetilde{a})^\top(A^\top Z^\top Z B - n\kappa^2 A^\top B)b| = \frac{1}{4}|\{4(a-\widetilde{a})^\top\}(A^\top Z^\top Z B - n\kappa^2 A^\top B)b| \leq \frac{\psi}{4}.$$

Similarly

$$|\widetilde{a}^\top(A^\top Z^\top Z B - n\kappa^2 A^\top B)(\widetilde{b} - b)| \leq \frac{\psi}{4}.$$

In addition,

$$|\widetilde{a}^\top(A^\top Z^\top Z B - n\kappa^2 A^\top B)\widetilde{b}| \leq \sup_{a\in\mathcal{N}_p(1/4), b\in\mathcal{N}_q(1/4)} |a^\top(A^\top Z^\top Z B - n\kappa^2 A^\top B)b|.$$

So for any a and b,

$$|a^\top(A^\top Z^\top Z B - n\kappa^2 A^\top B)b| \leq \frac{1}{2}\psi + \sup_{a\in N_p(1/4), b\in\mathcal{N}_q(1/4)} |a^\top(A^\top Z^\top Z B - n\kappa^2 A^\top B)b|.$$

Taking sup over all unit vectors $a \in \mathbb{R}^p$ and $b \in \mathbb{R}^q$, it follows that

$$\psi \leq \frac{1}{2}\psi + \sup_{a\in\mathcal{N}_p(1/4), b\in\mathcal{N}_q(1/4)} |a^\top(A^\top Z^\top Z B - n\kappa^2 A^\top B)b|,$$

or simply

$$\psi \leq 2 \sup_{a\in\mathcal{N}_p(1/4), b\in\mathcal{N}_q(1/4)} |a^\top(A^\top Z^\top Z B - n\kappa^2 A^\top B)b|.$$

Step 2. By Vershynin (2018), the cardinality of $\mathcal{N}_p(\epsilon)$ is bounded by $(\frac{C}{\epsilon})^p$, and the cardinality of $\mathcal{N}_q(\epsilon)$ is bounded by $(\frac{C}{\epsilon})^q$, for a positive constant C. Therefore

$$\mathbb{P}(\psi \geq 2t) = \mathbb{P}\bigg(\sup_{a\in\mathcal{N}_p(1/4), b\in\mathcal{N}_q(1/4)} |a^\top(A^\top Z^\top Z B - n\kappa^2 A^\top B)b| \geq t\bigg)$$

$$\leq {C_2}^p {C_2}^q \sup_{a \in \mathcal{N}_p(1/4), b \in \mathcal{N}_q(1/4)} \mathbb{P}(|a^\top (A^\top Z^\top Z B - n\kappa^2 A^\top B) b| \geq t)$$
$$= {C_2}^p {C_2}^q \sup_{a \in \mathcal{N}_p(1/4), b \in \mathcal{N}_q(1/4)} \mathbb{P}(|(Aa)^\top Z^\top Z (Bb) - n\kappa^2 (Aa)^\top (Bb)| \geq t)$$
$$\leq 2 {C_2}^p {C_2}^q \exp\left(- c \min \left\{ \frac{t^2}{n\kappa^4 \|Aa\|_2^2 \|Bb\|_2^2}, \frac{t}{\kappa^2 \|Aa\|_2 \|Bb\|_2} \right\} \right),$$

where C_2 is a positive constant. The desired result follows from the observation that $\|Aa\|_2 \leq \|A\| \|a\|_2 \leq \|A\|$, and $\|Bb\|_2 \leq \|B\| \|b\|_2 \leq \|B\|$. □

Lemma 15. *Suppose $Z \in \mathbb{R}^{m \times n}$ is a sub-Gaussian random matrix in the sense that for any $u \in \mathbb{R}^m, v \in \mathbb{R}^n$, it holds that*
$$\|u^\top Z v\|_{\psi_2} \leq \kappa \|u\|_2 \|v\|_2.$$
Then with probability at least $1 - \exp(-c(m+n))$, it holds that
$$\mathbb{P}\left(\|Z\| > t\right) \leq C_1 \exp\left(C(m+n) - \frac{c\, t^2}{\kappa^2} \right),$$
where c, C and C_1 are positive constants.

Proof. By assumption,
$$\mathbb{P}\left(\left| u^\top Z\, v \right| > t \right) \leq 2 \exp\left(- \frac{c\, t^2}{\kappa^2\, \|u\|_2^2\, \|v\|_2^2} \right).$$
The rest of the proof is similar and simpler than Lemma 13 and is omitted.

□

E.1 SVD for Unbalanced Matrices

Lemma 16. *Suppose*
$$Y = X + Z \in \mathbb{R}^{n \times m},$$
where X is a non-random matrix of arbitrary rank, and Z is a random matrix whose entries are i.i.d. sub-Gaussian random variables with mean zero and the sub-Gaussian norm $\|Z_{ij}\|_{\psi_2} = \kappa < \infty$. For any $r \leq \min\{n, m\}$, write the full SVD of X as
$$X = U \Sigma V^\top = [U_r\ U_\perp] \cdot \begin{bmatrix} \Sigma_r & \\ & \Sigma_\perp \end{bmatrix} \cdot \begin{bmatrix} V_r^\top \\ V_\perp^\top \end{bmatrix} = X_r + X_\perp.$$
Here $U_r \in \mathbb{O}_{m,r}$, $V_r \in \mathbb{O}_{m,r}$ correspond to the leading r left and right singular vectors of X. Suppose that
$$\{\sigma_r(X) - \sigma_{r+1}(X)\}^2 \geq C_{\text{gap}} \kappa^2 \{\sqrt{mn} + m\}$$
where $C_{\text{gap}} > 0$ is a sufficient large constant. Then with probability at least $1 - C_1 \exp(-C_2 n)$, it holds that
$$\left\| \sin \Theta\left(\widehat{V}_r, V_r \right) \right\|^2 \leq C_3 \left\{ \frac{m\kappa^2}{(\sigma_r(X) - \sigma_{r+1}(X))^2} + \frac{\kappa^4 nm}{(\sigma_r(X) - \sigma_{r+1}(X))^4} \right\},$$
where $C_1, C_2, C_3 > 0$ are absolute constants only depending on C_{gap}.

Proof. Note that by assumption, we have
$$\mathbb{E}[Z^\top Z] = n\kappa^2 I_m, \quad \mathbb{E}[Y^\top Y] = V_r \Sigma_r^2 V_r^\top + V_\perp \Sigma_\perp^2 V_\perp^\top + n\kappa^2 I_m, \quad \mathbb{E}[V_r^\top Y^\top Y V_r] = \Sigma_r^2 + n\kappa^2 I_r.$$
Define the diagonal weighting matrix
$$M = \operatorname{diag}\left((\sigma_1^2 + n\kappa^2)^{-1/2}, \ldots, (\sigma_r^2 + n\kappa^2)^{-1/2} \right) \in \mathbb{R}^{r \times r}.$$
Then it holds that
$$Y V_r M = (X_r + X_\perp + Z)\, V_r M = (X_r + Z)\, V_r M,$$

$$M^\top \mathbb{E}\left[V_r^\top Y^\top Y V_r\right] M = M^\top \mathbb{E}\left[V_r^\top \left(X_r + Z\right)^\top \left(X_r + Z\right) V_r\right] M = I_r.$$

Step 1. For $\sigma_r(YV_r)$, observe that

$$\begin{aligned}\sigma_r^2(YV_r) =& \sigma_r^2\left(\{X_r + Z\}V_r\right) = \sigma_r^2\left(\{X_r + Z\}V_r M M^{-1}\right) \geq \sigma_r^2(\{X_r + Z\}V_r M)\sigma_{\min}^2(M^{-1}) \\ =& \sigma_r^2\left(\{X_r + Z\}V_r M\right)\{\sigma_r^2(X) + n\kappa^2\} \\ =& \sigma_r\big(M^\top V_r^\top \{X_r + Z\}^\top \{X_r + Z\}V_r M\big)\{\sigma_r^2(X) + n\kappa^2\} \end{aligned} \tag{32}$$

where the inequality follows from Lemma 20.

Consider the term $\sigma_r\big(M^\top V_r^\top \{X_r + Z\}^\top \{X_r + Z\}V_r M\big)$. Note that

$$\begin{aligned} & M^\top V_r^\top \left(X_r + Z\right)^\top \left(X_r + Z\right) V_r M - I_r \\ =& M^\top V_r^\top \left(X_r + Z\right)^\top \left(X_r + Z\right) V_r M - \mathbb{E}\left[M^\top V_r^\top \left(X_r + Z\right)^\top \left(X_r + Z\right) V_r M\right] \\ =& \underbrace{M^\top V_r^\top X_r^\top X_r V_r M - \mathbb{E}\left[M^\top V_r^\top X_r^\top X_r V_r M\right]}_{=0} + M^\top V_r^\top X_r^\top Z V_r M \\ & + M^\top V_r^\top Z^\top X_r V_r M + M^\top V_r^\top Z^\top Z V_r M - \mathbb{E}\left[M^\top V_r^\top Z^\top Z V_r M\right] \\ & + \underbrace{\mathbb{E}\left[M^\top V_r^\top X_r^\top Z V_r M\right]}_{=0} + \underbrace{\mathbb{E}\left[M^\top V_r^\top Z^\top X_r V_r M\right]}_{=0} \\ =& M^\top V_r^\top X_r^\top Z V_r M + M^\top V_r^\top Z^\top X_r V_r M + M^\top V_r^\top \left(Z^\top Z - n\kappa^2 I_m\right) V_r M. \end{aligned} \tag{33}$$

Since,

$$\|X_r V_r M\|^2 = \max_{k=1,\dots,r} \frac{\sigma_k^2(X)}{\sigma_k^2(X) + n\kappa^2} \leq 1, \qquad \text{and} \qquad \|V_r M\|^2 = \|M\|^2 = \frac{1}{\sigma_r^2(X) + n\kappa^2},$$

it follows by Lemma 13 that

$$\mathbb{P}\Big(\big\|M^\top V_r^\top X_r^\top Z V_r M\big\| \geq x\Big) \leq 2\,\exp\left(Cr - cx^2\, \frac{\sigma_r^2(X) + n\kappa^2}{\kappa^2}\right). \tag{34}$$

Similarly, Lemma 14 implies that

$$\begin{aligned} & \mathbb{P}\Big(\big\|M^\top V_r^\top \left(Z^\top Z - n\kappa^2 I_m\right) V_r M\big\| \geq x\Big) \\ & \leq 2\,\exp\left(Cr - c\min\left\{x^2\, \frac{\left(\sigma_r^2(X) + n\kappa^2\right)^2}{n\kappa^4},\, x\, \frac{\sigma_r^2(X) + n\kappa^2}{\kappa^2}\right\}\right), \\ & \leq 2\,\exp\left(Cr - c\,\frac{\sigma_r^2(X) + n\kappa^2}{\kappa^2}\,\min\left\{x^2,\, x\right\}\right), \end{aligned} \tag{35}$$

where the last inequality follows from the fact that $\frac{\{\sigma_r^2(X)+n\kappa^2\}^2}{\kappa^4 n} \geq \frac{\sigma_r^2(X)+n\kappa^2}{\kappa^2}$. Thus, combining (33), (34), and (35), we have

$$\mathbb{P}\big(\big\|M^\top V_r^\top (X_r + Z)^\top (X_r + Z) V_r M - I_r\big\| \geq x\big) \leq 6\exp\left(Cr - c\frac{\sigma_r^2(X) + n\kappa^2}{\kappa^2}\,\min\{x^2, x\}\right). \tag{36}$$

Here we apply a union bound to the three terms in (33) with thresholds $x/3$ each. The same tail bound as in (34) holds for $M^\top V_r^\top Z^\top X_r V_r M$ by symmetry, and (35) controls the centered quadratic term $M^\top V_r^\top (Z^\top Z - n\kappa^2 I_m) V_r M$. All absolute constants are absorbed into C, c. This implies that

$$\begin{aligned} & \mathbb{P}\Big(\sigma_r\big(M^\top V_r^\top \{X_r + Z\}^\top \{X_r + Z\}V_r M\big) \geq 1 - x\Big) \\ & \geq 1 - 6\exp\left(Cr - c\,\frac{\sigma_r^2(X) + n\kappa^2}{\kappa^2}\,\min\left\{x^2,\, x\right\}\right), \end{aligned} \tag{37}$$

(37) and (32) together imply that

$$\mathbb{P}\Big(\sigma_r^2(YV_r) \geq \{\sigma_r^2(X) + n\kappa^2\}(1-x)\Big) \geq 1 - 6\,\exp\left(Cr - c\,\frac{\sigma_r^2(X)+n\kappa^2}{\kappa^2}\min\left\{x^2,\, x\right\}\right).$$

Setting $x = \frac{1}{6}\frac{\sigma_r^2(X)-\sigma_{r+1}^2(X)}{\sigma_r^2(X)+n\kappa^2}$, we have

$$\begin{aligned}&\mathbb{P}\left(\sigma_r^2(YV_r) \geq \sigma_r^2(X) + n\kappa^2 - \frac{\sigma_r^2(X)-\sigma_{r+1}^2(X)}{6}\right)\\ &\geq 1 - 6\exp\left(Cr - c\min\left\{\frac{1}{36\kappa^2}\,\frac{\big(\sigma_r^2(X)-\sigma_{r+1}^2(X)\big)^2}{\sigma_r^2(X)+n\kappa^2},\, \frac{\sigma_r^2(X)-\sigma_{r+1}^2(X)}{6\kappa^2}\right\}\right).\end{aligned} \tag{38}$$

Step 2. We upper bound the term $\sigma_{r+1}^2(Y)$. Note that

$$\sigma_{r+1}(Y) = \min_{\mathrm{rank}(B)\leq r}\|Y - B\| \leq \left\|Y - Y\cdot V_rV_r^\top\right\| = \sigma_{\max}(YV_\perp).$$

Moreover,

$$\begin{aligned}\sigma_{\max}^2(YV_\perp) &= \left\|V_\perp^\top Y^\top YV_\perp\right\| = \left\|V_\perp^\top(X_r+X_\perp+Z)^\top(X_r+X_\perp+Z)V_\perp\right\|\\ &= \left\|V_\perp^\top(X_\perp+Z)^\top(X_\perp+Z)V_\perp\right\|\\ &\leq \left\|V_\perp^\top Z^\top ZV_\perp\right\| + \left\|V_\perp^\top Z^\top X_\perp V_\perp\right\| + \left\|V_\perp^\top X_\perp^\top ZV_\perp\right\| + \left\|V_\perp^\top X_\perp^\top X_\perp V_\perp\right\|\\ &= \underbrace{\left\|V_\perp^\top Z^\top ZV_\perp\right\|}_{\mathrm{I}_1} + 2\underbrace{\left\|V_\perp^\top Z^\top X_\perp V_\perp\right\|}_{\mathrm{I}_2} + \underbrace{\left\|V_\perp^\top X_\perp^\top X_\perp V_\perp\right\|}_{\mathrm{I}_3}.\end{aligned} \tag{39}$$

For the term I_3, we have

$$\mathrm{I}_3 = \sigma_1^2(X_\perp V_\perp) = \sigma_1^2(U_\perp\Sigma_\perp V_\perp^\top V_\perp) = \sigma_1^2(\Sigma_\perp) = \sigma_{r+1}^2(X). \tag{40}$$

For I_2, note that

$$\|X_\perp V_\perp\|^2 = \|U_\perp\Sigma_\perp\|^2 = \sigma_{r+1}^2(X), \qquad \text{and} \qquad \|V_\perp\|^2 = 1.$$

it follows from Lemma 13 that

$$\mathbb{P}\Big(\mathrm{I}_2 \geq x\Big) \leq 2\,\exp\left(C_1m - \frac{c_1x^2}{\kappa^2\,\sigma_{r+1}^2(X)}\right). \tag{41}$$

For I_1, note that by Lemma 14 and the fac that $V_\perp^\top V_\perp = I_{m-r}$, we have

$$\begin{aligned}&\mathbb{P}\Big(\left\|V_\perp^\top Z^\top ZV_\perp - n\kappa^2 I_{m-r}\right\| \geq t\Big) \leq 2\,\exp\left(C_2m - c_2\min\left\{\frac{t^2}{n\,\kappa^4},\, \frac{t}{\kappa^2}\right\}\right),\\ &\Longrightarrow \mathbb{P}\Big(\mathrm{I}_1 \geq n\kappa^2(1+t)\Big) \leq 2\,\exp\left(C_2m - c_2n\min\left\{t^2,\, t\right\}\right).\end{aligned} \tag{42}$$

Combining the calculations in this step, with $x = \kappa\sigma_{r+1}(X)\sqrt{\frac{2C_1}{c_1}m}$ in (41), and with $t = \frac{\sigma_r^2(X)-\sigma_{r+1}^2(X)}{6n\kappa^2}$ in (42), it follows that

$$\begin{aligned}&\mathbb{P}\left(\sigma_{r+1}^2(Y) \geq n\kappa^2 + \frac{\sigma_r^2(X)-\sigma_{r+1}^2(X)}{6} + \kappa\sigma_{r+1}(X)\sqrt{\frac{2C_1}{c_1}m} + \sigma_{r+1}^2(X)\right)\\ &\leq 2\,\exp\left(C_2m - c_2\,\min\left\{\frac{\big(\sigma_r^2(X)-\sigma_{r+1}^2(X)\big)^2}{36\,\kappa^4\,n},\, \frac{\sigma_r^2(X)-\sigma_{r+1}^2(X)}{6\,\kappa^2}\right\}\right) + 2\,\exp\big(-C_3\,m\big).\end{aligned} \tag{43}$$

Step 3. Recall we define

$$M = \mathrm{diag}\left((\sigma_1^2+n\kappa^2)^{-1/2}, \ldots, (\sigma_r^2+n\kappa^2)^{-1/2}\right) \in \mathbb{R}^{r\times r}.$$

We have

$$\begin{aligned}
\|\mathcal{P}_{YV_r}YV_\perp\| &= \|\mathcal{P}_{YV_rM}YV_\perp\| \\
&= \left\|(YV_rM)\left((YV_rM)^\top(YV_rM)\right)^{-1}(YV_rM)^\top YV_\perp\right\| \\
&\le \left\|(YV_rM)\left((YV_rM)^\top(YV_rM)\right)^{-1}\right\| \left\|M^\top V_r^\top Y^\top YV_\perp\right\| \\
&\le \sigma_{\min}^{-1}(YV_rM)\left\|M^\top V_r^\top Y^\top YV_\perp\right\| = \sigma_r^{-1}(YV_rM)\left\|M^\top V_r^\top Y^\top YV_\perp\right\|, \qquad (44)
\end{aligned}$$

where the first equality follows from the fact that YV_r and YV_rM have the same column spaces (since M is invertible), the last inequality follows from Lemma 18, and for the last equality we use that the singular values of YV_rM are in nonincreasing order so that its smallest singular value equals $\sigma_r(YV_rM)$.

By (36), we have for every $x > 0$

$$\mathbb{P}\left(\left\|M^\top V_r^\top (X_r+Z)^\top (X_r+Z)V_rM - I_r\right\| \ge x\right) \le 6\exp\left(Cr - c\,\frac{\sigma_r^2(X)+n\kappa^2}{\kappa^2}\min\{x^2,x\}\right).$$

Taking $x = \frac{1}{2}$ gives

$$\mathbb{P}\left(\left\|M^\top V_r^\top (X_r+Z)^\top (X_r+Z)V_rM - I_r\right\| < \tfrac{1}{2}\right) \ge 1 - 6\exp\left(Cr - c\,\frac{\sigma_r^2(X)+n\kappa^2}{4\kappa^2}\right). \qquad (45)$$

In particular, on this event all eigenvalues of $M^\top V_r^\top (X_r+Z)^\top (X_r+Z)V_rM$ are at least $1/2$, so $\sigma_r^2(YV_rM) \ge 1/2$ with the same probability bound. Consider $\left\|M^\top V_r^\top Y^\top YV_\perp\right\|$. Since $V_r^\top X_\perp^\top = 0$, $X_rV_\perp = 0$ and $X_\perp^\top X_r = 0$, it follows that

$$\begin{aligned}
M^\top V_r^\top Y^\top YV_\perp &= M^\top V_r^\top \left(X_r + X_\perp + Z\right)^\top \left(X_r + X_\perp + Z\right)V_\perp \\
&= M^\top V_r^\top X_r^\top ZV_\perp + M^\top V_r^\top Z^\top X_\perp V_\perp + M^\top V_r^\top Z^\top ZV_\perp \\
&= M^\top V_r^\top X_r^\top ZV_\perp + M^\top V_r^\top Z^\top X_\perp V_\perp + M^\top V_r^\top Z^\top ZV_\perp - \underbrace{M^\top V_r^\top\left(n\kappa^2 I_m\right)V_\perp}_{=0},
\end{aligned}$$

Since,

$$\|X_rV_rM\|^2 \le 1,\ \|V_rM\|^2 = \frac{1}{\sigma_r^2(X)+n\kappa^2} \text{ and } \|X_\perp V_\perp\|^2 = \|X_\perp\|^2 = \sigma_{r+1}^2(X),$$

it follows from Lemma 13 that

$$\begin{aligned}
&\mathbb{P}\left(\left\|M^\top V_r^\top X_r^\top ZV_\perp\right\| \ge x\right) \le 2\exp\left(Cm - \frac{cx^2}{\kappa^2}\right), \\
&\mathbb{P}\left(\left\|M^\top V_r^\top Z^\top X_\perp V_\perp\right\| \ge x\right) \le 2\exp\left(Cm - \frac{cx^2}{\kappa^2}\,\frac{\sigma_r^2(X)+n\kappa^2}{\sigma_{r+1}^2(X)}\right) \le 2\exp\left(Cm - \frac{cx^2}{\kappa^2}\right).
\end{aligned} \qquad (46)$$

Similarly, Lemma 14 implies that

$$\begin{aligned}
&\mathbb{P}\left(\left\|M^\top V_r^\top\left(Z^\top Z - n\kappa^2 I_m\right)V_\perp\right\| \ge x\right) \\
&\le 2\exp\left(Cm - c\min\left\{x^2\,\frac{\sigma_r^2(X)+n\kappa^2}{n\kappa^4},\ x\,\frac{\sqrt{\sigma_r^2(X)+n\kappa^2}}{\kappa^2}\right\}\right), \qquad (47) \\
&\le 2\exp\left(Cm - c\min\left\{\frac{x^2}{\kappa^2},\ x\,\frac{\sqrt{\sigma_r^2(X)+n\kappa^2}}{\kappa^2}\right\}\right),
\end{aligned}$$

where the last inequality uses $\frac{\sigma_R^2(X)+n\kappa^2}{n\,\kappa^2} \ge 1$. Thus, combining (44), (45), (46) and (47) with $x = \kappa\sqrt{\frac{2C}{c}m}$, we have

$$\mathbb{P}\left(\|\mathcal{P}_{YV_r}YV_\perp\|^2 \ge \frac{36C}{c}\,m\kappa^2\right) \le 6\exp\left(Cr - c\frac{\sigma_r^2(X)+n\kappa^2}{4\kappa^2}\right) + 4\exp(-Cm) \qquad (48)$$

$$+ 2\exp\Big(Cm - c\min\Big\{ \tfrac{2C}{c}\, m,\ \sqrt{\tfrac{2C}{c}\, m\ \tfrac{\sigma_r^2(X)+n\kappa^2}{\kappa^2}} \Big\}\Big).$$

Step 4. Define the event

$$\mathcal{E} = \Bigg\{\sigma_r^2(YV_r) \geq \sigma_r^2(X) + n\kappa^2 - \frac{\sigma_r^2(X) - \sigma_{r+1}^2(X)}{6}\,;$$
$$\sigma_{r+1}^2(Y) \leq n\kappa^2 + \frac{\sigma_r^2(X) - \sigma_{r+1}^2(X)}{6} + \kappa\sigma_{r+1}(X)\sqrt{\frac{2C}{c}m} + \sigma_{r+1}^2(X)\,;$$
$$\|\mathcal{P}_{YV_r} YV_\perp\|^2 \leq \frac{8C}{c}\, m\kappa^2\Bigg\}.$$

It follows from (38), (43) and (48) that by the union bound,

$$\mathbb{P}(\mathcal{E}^c) \leq 6\exp\left(Cr - c\min\left\{\frac{1}{36\kappa^2}\,\frac{\big(\sigma_r^2(X) - \sigma_{r+1}^2(X)\big)^2}{\sigma_r^2(X) + n\kappa^2},\ \frac{\sigma_r^2(X) - \sigma_{r+1}^2(X)}{6\kappa^2}\right\}\right) \tag{49}$$

$$+ 2\ \exp\left(Cm\ -\ c\ \min\left\{\frac{\big(\sigma_r^2(X) - \sigma_{r+1}^2(X)\big)^2}{36\,\kappa^4\, n}, \frac{\sigma_r^2(X) - \sigma_{r+1}^2(X)}{6\,\kappa^2}\right\}\right) + 2\ \exp\big(-C\, m\big) \tag{50}$$

$$+ 6\exp\left(Cr - c\frac{\sigma_r^2(X) + n\kappa^2}{4\kappa^2}\right) + 4\exp(-Cm) \tag{51}$$

$$+ 2\exp\left(Cm - c\min\left\{\frac{2C}{c}\, m,\ \sqrt{\frac{2C}{c}\, m\frac{\sigma_r^2(X) + n\kappa^2}{\kappa^2}}\right\}\right). \tag{52}$$

In what follows, we show that under the SNR assumption

$$\Big(\sigma_r(X) - \sigma_{r+1}(X)\Big)^2 \geq C_{\text{gap}}\kappa^2\Big(\sqrt{nm} + m\Big)$$

with sufficient large absolute constant $C_{\text{gap}} > 0$, we have

$$\mathbb{P}(\mathcal{E}^c) \leq C\exp\left(-Cm\right),$$

where $C > 0$ is an absolute constant, appropriately scaled to absorb the other constants.

We illustrate how to bound (49), as the rest of the terms can be bounded in a similar and simpler way. Note that

$$\frac{1}{36\kappa^2}\,\frac{\big(\sigma_r^2(X) - \sigma_{r+1}^2(X)\big)^2}{\sigma_r^2(X) + n\kappa^2} \geq \frac{\big(\sigma_r^2(X) - \sigma_{r+1}^2(X)\big)^2}{72\kappa^2}\min\left\{\frac{1}{\sigma_r^2(X)}, \frac{1}{n\kappa^2}\right\}$$
$$\geq \min\left\{\frac{\big(\sigma_r^2(X) - \sigma_{r+1}^2(X)\big)^2}{72\,\kappa^2\sigma_r^2(X)}, \frac{\big(\sigma_r^2(X) - \sigma_{r+1}^2(X)\big)^2}{72n\kappa^4}\right\}.$$

We have

$$\frac{\big(\sigma_r^2(X) - \sigma_{r+1}^2(X)\big)^2}{72\,\kappa^2\sigma_r^2(X)} = \frac{\big(\sigma_r(X) + \sigma_{r+1}(X)\big)^2\big(\sigma_r(X) - \sigma_{r+1}(X)\big)^2}{72\,\kappa^2\sigma_r^2(X)} \geq \frac{\big(\sigma_r(X) - \sigma_{r+1}(X)\big)^2}{72\,\kappa^2} \geq \frac{2C}{c}m,$$
$$\frac{\big(\sigma_r^2(X) - \sigma_{r+1}^2(X)\big)^2}{72\,\kappa^4\, n} \geq \frac{\big(\sigma_r(X) - \sigma_{r+1}(X)\big)^4}{72\,\kappa^4\, n} \geq \frac{2C}{c}m.$$

So

$$\frac{1}{36\kappa^2}\,\frac{\big(\sigma_r^2(X) - \sigma_{r+1}^2(X)\big)^2}{\sigma_r^2(X) + n\kappa^2} \geq \frac{2C}{c}m.$$

In addition,

$$\frac{\sigma_R^2(X) - \sigma_{R+1}^2(X)}{6\kappa^2} \geq \frac{\big(\sigma_R^2(X) - \sigma_{R+1}^2(X)\big)^2}{72\,\kappa^2\sigma_R^2(X)} \geq \frac{2C}{c}m.$$

So (49) $\leq C\exp\left(-Cm\right)$, for some absolute constant $C>0$.

Step 5. Under the event $\mathcal{E}$, by Lemma 24, we have that

$$\begin{aligned}\left\|\sin\Theta\left(\widehat{V}_r,V_r\right)\right\|^2 &\leq \frac{\sigma_r^2(YV_r)\left\|\mathcal{P}_{YV_r}YV_\perp\right\|^2}{\left(\sigma_r^2(YV_r)-\sigma_{r+1}^2(Y)\right)^2}\leq C_7\frac{\sigma_r^2(YV_r)m\kappa^2}{\left(\sigma_r^2(YV_r)-\sigma_{r+1}^2(Y)\right)^2}\\
&\leq C_8\frac{\left(\sigma_r^2(X)+n\kappa^2-\frac{\sigma_r^2(X)-\sigma_{r+1}^2(X)}{6}\right)m\kappa^2}{\left(\sigma_r^2(X)+n\kappa^2-\frac{\sigma_r^2(X)-\sigma_{r+1}^2(X)}{6}-\sigma_{r+1}^2(Y)\right)^2}\\
&\leq C_8\frac{\left(\sigma_r^2(X)+n\kappa^2-\frac{\sigma_r^2(X)-\sigma_{r+1}^2(X)}{6}\right)m\kappa^2}{\left(\left(1-\frac{1}{3}\right)\left(\sigma_r^2(X)-\sigma_{r+1}^2(X)\right)-\kappa\sigma_{r+1}(X)\sqrt{\frac{2C}{c}m}\right)^2}\\
&\leq C_9\frac{\left(\sigma_r^2(X)+n\kappa^2-\frac{\sigma_r^2(X)-\sigma_{r+1}^2(X)}{6}\right)m\kappa^2}{(1-\frac{1}{2})^2\left(\sigma_r^2(X)-\sigma_{r+1}^2(X)\right)^2}\\
&\leq C_{10}\frac{\left(\sigma_r^2(X)+n\kappa^2-\frac{\sigma_r^2(X)-\sigma_{r+1}^2(X)}{6}\right)m\kappa^2}{\left(\sigma_r^2(X)-\sigma_{r+1}^2(X)\right)^2},\end{aligned}$$

Here, the third inequality follows from the fact that $x^2/(x^2-y^2)^2$ is a decreasing function of x and an increasing function of y when $x>y\geq 0$, together with the fact that the event $\mathcal{E}$ holds. The fifth inequality follows from the fact that, under the assumption $\left(\sigma_r(X)-\sigma_{r+1}(X)\right)^2\geq C_{\text{gap}}\kappa^2(\sqrt{nm}+m)$ with $C_{\text{gap}}>0$ being large enough,

$$\begin{aligned}\frac{\left(\sigma_r^2(X)-\sigma_{r+1}^2(X)\right)^2}{36} &=\left(\sigma_r(X)+\sigma_{r+1}(X)\right)^2\frac{\left(\sigma_r(X)-\sigma_{r+1}(X)\right)^2}{36}\\
&\geq\sigma_r^2(X)\frac{C_{\text{gap}}m\kappa^2}{36}\geq Cm\kappa^2\sigma_r^2(X)\geq Cm\kappa^2\sigma_{r+1}^2(X).\end{aligned}$$

Therefore, with probability at least $1-C\exp(-Cm)$,

$$\begin{aligned}\left\|\sin\Theta\left(\widehat{V}_r,V_r\right)\right\|^2 &\leq C_3\frac{\left(\sigma_r^2(X)+n\kappa^2-\frac{\sigma_r^2(X)-\sigma_{r+1}^2(X)}{6}\right)m\kappa^2}{(\sigma_r^2(X)-\sigma_{r+1}^2(X))^2}\\
&\leq C_3\frac{\left(\sigma_r^2(X)+n\kappa^2\right)m\kappa^2}{(\sigma_r^2(X)-\sigma_{r+1}^2(X))^2}=C_3\frac{\sigma_r^2(X)m\kappa^2}{(\sigma_r^2(X)-\sigma_{r+1}^2(X))^2}+C_3\frac{nm\kappa^4}{(\sigma_r^2(X)-\sigma_{r+1}^2(X))^2}\\
&\leq C_4\left\{\frac{m\kappa^2}{(\sigma_r(X)-\sigma_{r+1}(X))^2}+\frac{\kappa^4nm}{(\sigma_r(X)-\sigma_{r+1}(X))^4}\right\},\end{aligned}$$

where the third inequality follows from the observation that

$$\frac{\sigma_r^2(X)}{(\sigma_r^2(X)-\sigma_{r+1}^2(X))^2}=\frac{\sigma_r^2(X)}{(\sigma_r(X)+\sigma_{r+1}(X))^2(\sigma_r(X)-\sigma_{r+1}(X))^2}\leq\frac{1}{(\sigma_r(X)-\sigma_{r+1}(X))^2},$$

and last display follows from

$$\sigma_r^2(X)-\sigma_{r+1}^2(X)=(\sigma_r(X)+\sigma_{r+1}(X))(\sigma_r(X)-\sigma_{r+1}(X))\geq\left(\sigma_r(X)-\sigma_{r+1}(X)\right)^2.$$

□

Corollary 17. *Suppose the conditions of Theorem 1 hold, in particular the condition in Equation (3) for each mode-k unfolding $\mathcal{M}_k(X^*)$ with target rank r_k. Let $U_k^{(0)}$ be the matrix of top r_k left singular vectors of the mode-k unfolding $Y^{(k)}:=\mathcal{M}_k(Y)$. Then for each $k\in\{1,2,3\}$, with probability at least $1-C_1\exp(-C_2p_k)$,*

$$\left\|\sin\Theta\left(U_k^{(0)},U_k^*\right)\right\|\ \leq\ \frac{1}{2\sqrt{r_{\max}}}.$$

Proof. Apply Lemma 16 to $Y^{(k)} = \mathcal{M}_k(Y) \in \mathbb{R}^{p_k \times p_{\neg k}}$ with signal $X^{(k)} = \mathcal{M}_k(X^*)$ and noise $Z^{(k)} = \mathcal{M}_k(Z)$, where $p_{\neg k} := \prod_{j \neq k} p_j$. Use rank $r = r_k$ and identify $n := p_k$ (rows) and $m := p_{\neg k}$ (columns). The lemma (applied to left singular vectors, or equivalently to the transpose) gives

$$\left\|\sin\Theta(U_k^{(0)}, U_k^*)\right\|^2 \;\leq\; C\left\{\frac{p_k\,\kappa^2}{\Delta_k^2} + \frac{\kappa^4\,p_k\,p_{\neg k}}{\Delta_k^4}\right\},$$

where $\Delta_k := \sigma_{r_k}\big(\mathcal{M}_k(X^*)\big) - \sigma_{r_k+1}\big(\mathcal{M}_k(X^*)\big)$ is the mode-k spectral gap.

By the assumption (3) in Theorem 1 (applied to mode k),

$$\Delta_k^2 \;\geq\; C_{\text{gap}}\kappa^2\left(\sqrt{p_k\,p_{\neg k}\,r_{\max}} + r_{\max}\sum_{j=1}^{3} p_j\right).$$

Choosing C_{gap} sufficiently large makes the right–hand side above at most $1/(4r_{\max})$, hence

$$\left\|\sin\Theta(U_k^{(0)}, U_k^*)\right\| \;\leq\; \frac{1}{2\sqrt{r_{\max}}}.$$

The probability bound $1 - C_1\exp(-C_2 p_k)$ matches the row dimension in the matrix lemma. □

F Matrix Perturbation Bounds

Lemma 18. *Suppose that $A \in \mathbb{R}^{n\times r}$. Then*

$$\|A(A^\top A)^{-1}\| \leq \sigma_r^{-1}(A).$$

Proof. If $\sigma_r(A) = 0$, then the desired result trivially follows. So suppose $\operatorname{rank}(A) = r$. Therefore $A^\top A$ is invertible. Let the SVD of A satisfies $A = U_A\Sigma_A V_A^\top$, then

$$\left\|A(A^\top A)^{-1}\right\| = \left\|U_A\Sigma_A V_A^\top(V_A\Sigma_A^2 V_A^\top)^{-1}\right\| = \left\|U_A\Sigma_A^{-1}V_A^\top\right\| = \sigma_{\min}^{-1}(A) = \sigma_r^{-1}(A).$$

□

Lemma 19. *Suppose $A \in \mathbb{R}^{m\times n}$ and $B \in \mathbb{R}^{n\times k}$ are any two matrices. Then*

$$\sigma_j(AB) \leq \sigma_j(A)\sigma_{\max}(B).$$

Proof. Let $\lambda_j(M)$ denote the $j-th$ eigenvalues of M in the absolute value order. Then

$$\sigma_j(AB) = \sqrt{\lambda_j(ABB^\top A^\top)} \quad \text{and} \quad \sigma_j(A) = \sqrt{\lambda_j(AA^\top)}. \tag{53}$$

Since $BB^\top \preceq \sigma_{\max}^2(B)I_n$, it follows that

$$ABB^\top A^\top \preceq A(\sigma_{\max}^2(B)I_n)A^\top = \sigma_{\max}^2(B)AA^\top.$$

By the monotonicity of eigenvalues under the positive definite matrices, it follows that

$$\lambda_j(ABB^\top A^\top) \leq \sigma_{\max}^2(B)\lambda_j(AA^\top). \tag{54}$$

The desired result follows from (53). □

Lemma 20. *Suppose $A \in \mathbb{R}^{m\times n}$ and $B \in \mathbb{R}^{n\times n}$ are any two matrices. Then*

$$\sigma_j(AB) \geq \sigma_j(A)\sigma_{\min}(B).$$

Proof. Suppose $\sigma_{\min}(B) = 0$. Then the desired result immediately follows. Therefore it suffices to assume $\sigma_{\min}(B) > 0$ and B is invertible. It suffices to observe that

$$\sigma_j(A) = \sigma_j(ABB^{-1}) \leq \sigma_j(AB)\sigma_{\max}(B^{-1}) = \sigma_j(AB)\sigma_{\min}^{-1}(B),$$

where the inequality follows from Lemma 19.

□

Lemma 21. *For any real matrices $A \in \mathbb{R}^{n\times m}$ and $B \in \mathbb{R}^{m\times m}$, it holds that*

$$\|AB\|_{\mathrm{F}} \geq \sigma_{\min}(B)\|A\|_{\mathrm{F}}.$$

Proof of Lemma 21. Since B is a square matrix, it follows that

$$\lambda_{\min}(BB^\top) = \sigma^2_{\min}(B),$$

where $\lambda_{\min}(\cdot)$ denotes the minimum eigenvalue. Note that

$$BB^\top \succeq \lambda_{\min}(BB^\top)I_m.$$

Therefore

$$ABB^\top A^\top \succeq A\{\lambda_{\min}(BB^\top)I_m\}A^\top,$$

and so

$$\mathrm{tr}(ABB^\top A^\top) \geq \mathrm{tr}(A\{\lambda_{\min}(BB^\top)I_m\}A^\top).$$

Then

$$\|AB\|^2_{\mathrm{F}} = \mathrm{tr}(ABB^\top A^\top) \geq \mathrm{tr}(A\{\lambda_{\min}(BB^\top)I_m\}A^\top) = \lambda_{\min}(BB^\top)\,\mathrm{tr}(AA^\top) = \sigma^2_{\min}(B)\|A\|^2_{\mathrm{F}}.$$

□

Lemma 22. *Let $A \in \mathbb{R}^{p\times q}$ and $U \in \mathbb{O}^{q\times r}$. Then*

$$\|AUU^\top\|_{\mathrm{F}} = \|AU\|_{\mathrm{F}}.$$

Proof. Observe that

$$\|AUU^\top\|^2_{\mathrm{F}} = \mathrm{tr}(AUU^\top UU^\top A^\top) = \mathrm{tr}(AUU^\top A^\top) = \|AU\|^2_{\mathrm{F}}.$$

□

Lemma 23. *Suppose $B, Z \in \mathbb{R}^{n\times m}$. For all $1 \leq R \leq \min\{n, m\}$, write the full SVD of A as*

$$A = B + Z = \widehat{U}\widehat{\Sigma}\widehat{V}^\top = \begin{bmatrix}\widehat{U}_{(R)} & \widehat{U}_\perp\end{bmatrix} \cdot \begin{bmatrix}\widehat{\Sigma}_{(R)} & \\ & \widehat{\Sigma}_\perp\end{bmatrix} \cdot \begin{bmatrix}\widehat{V}^\top_{(R)} \\ \widehat{V}^\top_\perp\end{bmatrix},$$

where $\widehat{U}_{(R)} \in \mathbb{O}_{n,R}$, $\widehat{V}_{(R)} \in \mathbb{O}_{m,R}$ correspond to the leading R left and right singular vectors; and $\widehat{U}_\perp \in \mathbb{O}_{n,n-R}$, $\widehat{V}_\perp \in \mathbb{O}_{m,m-R}$ correspond to their orthonormal complement. We have

$$\begin{aligned}\left\|\mathcal{P}_{\widehat{U}_\perp}B\right\|_{\mathrm{F}} \leq& 3\sqrt{\sum_{j=R+1}^{\min\{n,m\}}\sigma_j^2(B)} + 2\min\left\{\sqrt{R}\|Z\|, \|Z\|_{\mathrm{F}}\right\} \\ =& 3\left\|B_{(R)} - B\right\|_{\mathrm{F}} + 2\min\left\{\sqrt{R}\|Z\|, \|Z\|_{\mathrm{F}}\right\},\end{aligned}$$

where $B_{(R)}$ denote the rank-R truncated SVD of B, this is, if $B = U\Sigma V^\top$ then $B_{(R)} := U_{(R)}\Sigma_{(R)}V^\top_{(R)}$.

Proof. Without loss of generality, assume $n \leq m$. For $A \in \mathbb{R}^{n\times m}$, let $\Sigma(A) \in \mathbb{R}^{n\times m}$ denote the non-negative diagonal matrices whose diagonal entries are the non-increasingly ordered singular values of A. For all $1 \leq R \leq n$, let $B_{(R)}$ denote the truncated SVD of B with rank R, and we have by the Eckart–Young–Mirsky theorem

$$\left\|B_{(R)} - B\right\|_{\mathrm{F}} = \sqrt{\sum_{j=R+1}^{n}\sigma_j^2(B)}.$$

For a matrix $A \in \mathbb{R}^{m\times n}$, let $\Sigma(A) \in \mathbb{R}^{m\times n}$ be a non-negative (rectangular) diagonal matrix whose diagonal entries are the non-increasingly ordered singular values of A.

We have that

$$\left\|\mathcal{P}_{\widehat{U}_\perp} B\right\|_{\mathrm{F}} \le \left\|\mathcal{P}_{\widehat{U}_\perp} B_{(R)}\right\|_{\mathrm{F}} + \left\|\mathcal{P}_{\widehat{U}_\perp}(B - B_{(R)})\right\|_{\mathrm{F}} = \sqrt{\sum_{j=1}^{R} \sigma_j^2(\mathcal{P}_{\widehat{U}_\perp} B_{(R)})} + \left\|\mathcal{P}_{\widehat{U}_\perp}(B - B_{(R)})\right\|_{\mathrm{F}}$$

$$\le \sqrt{\sum_{j=1}^{R} \sigma_j^2(\mathcal{P}_{\widehat{U}_\perp} B_{(R)})} + \left\|B - B_{(R)}\right\|_{\mathrm{F}} = \sqrt{\sum_{j=1}^{R} \sigma_j^2(\mathcal{P}_{\widehat{U}_\perp} B_{(R)})} + \sqrt{\sum_{j=R+1}^{n} \sigma_j^2(B)}$$

$$\le \left\|(\sigma_1(\mathcal{P}_{\widehat{U}_\perp} B_{(R)}) - \sigma_1(\mathcal{P}_{\widehat{U}_\perp} B), \ldots, \sigma_R(\mathcal{P}_{\widehat{U}_\perp} B_{(R)}) - \sigma_R(\mathcal{P}_{\widehat{U}_\perp} B))^\top\right\|_2 + \left\|(\sigma_1(\mathcal{P}_{\widehat{U}_\perp} B), \ldots, \sigma_R(\mathcal{P}_{\widehat{U}_\perp} B))^\top\right\|_2$$

$$+ \sqrt{\sum_{j=R+1}^{n} \sigma_j^2(B)}$$

$$\le \left\|\Sigma(\mathcal{P}_{\widehat{U}_\perp} B_{(R)}) - \Sigma(\mathcal{P}_{\widehat{U}_\perp} B)\right\|_{\mathrm{F}} + \left\|(\sigma_1(\mathcal{P}_{\widehat{U}_\perp} B), \ldots, \sigma_R(\mathcal{P}_{\widehat{U}_\perp} B))^\top\right\|_2 + \sqrt{\sum_{j=R+1}^{n} \sigma_j^2(B)}$$

$$\le \left\|\mathcal{P}_{\widehat{U}_\perp}(B_{(R)} - B)\right\|_{\mathrm{F}} + \sqrt{\sum_{j=1}^{R} \sigma_j^2(\mathcal{P}_{\widehat{U}_\perp} B)} + \sqrt{\sum_{j=R+1}^{n} \sigma_j^2(B)}$$

$$\le \sqrt{\sum_{j=1}^{R} \sigma_j^2(\mathcal{P}_{\widehat{U}_\perp} B)} + 2\sqrt{\sum_{j=R+1}^{n} \sigma_j^2(B)},$$

where the first equality follows from $\mathrm{rank}(B_{(R)}) = R$, and the fifth inequality follows from Theorem 26. To upper bound $\sqrt{\sum_{j=1}^{R} \sigma_j^2(\mathcal{P}_{\widehat{U}_\perp} B)}$, we first consider $\sqrt{\sum_{j=1}^{R} \sigma_j^2(\mathcal{P}_{\widehat{U}_\perp} A)}$. Note that

$$\mathcal{P}_{\widehat{U}_\perp} A = \sum_{j=R+1}^{n} \sigma_j(A) \widehat{u}_j \widehat{v}_j^\top,$$

where $\widehat{u}_j$ and $\widehat{v}_j$ are the left and right singular vectors associated with the jth largest singular value $\sigma_j(A)$. Note that $\sigma_j(A) = \sigma_j(B) = 0$ for $j > n$. It follows that

$$\sqrt{\sum_{j=1}^{R} \sigma_j^2(\mathcal{P}_{\widehat{U}_\perp} A)} = \sqrt{\sum_{j=R+1}^{2R} \sigma_j^2(A)} = \left\|(\sigma_{R+1}(A), \ldots, \sigma_{2R}(A))^\top\right\|$$

$$\le \left\|(\sigma_{R+1}(A) - \sigma_{R+1}(B), \ldots, \sigma_{2R}(A) - \sigma_{2R}(B))^\top\right\| + \left\|(\sigma_{R+1}(B), \ldots, \sigma_{2R}(B))^\top\right\|$$

$$\le \min\left\{\sqrt{R}\|Z\|, \|Z\|_{\mathrm{F}}\right\} + \sqrt{\sum_{j=R+1}^{n} \sigma_j^2(B)}, \tag{55}$$

where the first inequality follows from the triangle inequality, and second inequality follows from Weyl's inequality (Weyl, 1912), i.e. $|\sigma_j(A) - \sigma_j(B)| \le \|A - B\|$ for all $1 \le j \le n$, as well as the fact that

$$\left\|(\sigma_{R+1}(A) - \sigma_{R+1}(B), \ldots, \sigma_{2R}(A) - \sigma_{2R}(B))^\top\right\| \le \|\Sigma(A) - \Sigma(B)\|_{\mathrm{F}} \le \|Z\|_{\mathrm{F}},$$

where the last inequality follows from Theorem 26. It then follows from (55),

$$\sqrt{\sum_{j=1}^{R} \sigma_j^2(\mathcal{P}_{\widehat{U}_\perp} B)} = \left\|(\sigma_1(\mathcal{P}_{\widehat{U}_\perp}(A - Z)), \ldots, \sigma_R(\mathcal{P}_{\widehat{U}_\perp}(A - Z))^\top\right\|$$

$$\le \left\|(\sigma_1(\mathcal{P}_{\widehat{U}_\perp}(A - Z)) - \sigma_1(\mathcal{P}_{\widehat{U}_\perp} A), \ldots, \sigma_R(\mathcal{P}_{\widehat{U}_\perp}(A - Z)) - \sigma_R(\mathcal{P}_{\widehat{U}_\perp} A))^\top\right\|$$

$$+ \left\|(\sigma_1(\mathcal{P}_{\widehat{U}_\perp} A), \ldots, \sigma_R(\mathcal{P}_{\widehat{U}_\perp} A))^\top\right\|$$

$$
\begin{aligned}
&\leq \min\left\{\sqrt{R}\|\mathcal{P}_{\widehat{U}_\perp}Z\|, \|\mathcal{P}_{\widehat{U}_\perp}Z\|_{\mathrm{F}}\right\} + \sqrt{\sum_{j=1}^{R}\sigma_j^2(\mathcal{P}_{\widehat{U}_\perp}A)} \\
&\leq \min\left\{\sqrt{R}\|Z\|, \|Z\|_{\mathrm{F}}\right\} + \sqrt{\sum_{j=1}^{R}\sigma_j^2(P_{\widehat{U}_\perp}A)} \\
&\leq 2\min\left\{\sqrt{R}\|Z\|, \|Z\|_{\mathrm{F}}\right\} + \sqrt{\sum_{j=R+1}^{n}\sigma_j^2(B)},
\end{aligned}
$$

where the first two inequalities follow from the same arguments as in (55). Consequently,

$$
\left\|\mathcal{P}_{\widehat{U}_\perp}B\right\|_{\mathrm{F}} \leq 3\sqrt{\sum_{j=R+1}^{n}\sigma_j^2(B)} + 2\min\left\{\sqrt{R}\|Z\|, \|Z\|_{\mathrm{F}}\right\}.
$$

□

Lemma 24 (Proposition 1 of Cai & Zhang (2018))**.** *Suppose $Y \in \mathbb{R}^{m\times n}$, $\widehat{V} = [\widehat{V}_r\ \widehat{V}_\perp] \in \mathbb{O}_n$ where $\widehat{V}_r \in \mathbb{O}_{n,r}$, $\widehat{V}_\perp \in \mathbb{O}_{n,n-r}$ correspond to the first r and last $(n-r)$ right singular vectors of Y respectively. Let $V = [V_r\ V_\perp] \in \mathbb{O}_{n,n}$ be any orthogonal matrix with $V_r \in \mathbb{O}_{n,r}$, $V_\perp \in \mathbb{O}_{n,n-r}$. Given that $\sigma_R(YV_r) > \sigma_{r+1}(Y)$, we have*

$$
\|\sin\Theta(V_r, \widehat{V}_r)\| \leq \frac{\sigma_r(YV_r)\|\mathcal{P}_{YV_r}YV_\perp\|}{\sigma_r^2(YV_r) - \sigma_{r+1}^2(Y)} \wedge 1, \tag{56}
$$

where $\mathcal{P}_A$ is the projection operator onto the column space of A.

Lemma 25 (Properties of the sinΘ distances)**.**
The following properties hold for the $\sin\Theta$ *distances.*

1. *(Equivalent Expressions) Suppose $V, \widehat{V} \in \mathbb{O}_{p,R}$. If $V_\perp$ is an orthogonal extension of V, namely $[V \quad V_\perp] \in \mathbb{O}_p$, we have the following equivalent forms for $\|\sin\Theta(\widehat{V}, V)\|$ and $\|\sin\Theta(\widehat{V}, V)\|_{\mathrm{F}}$,*

$$
\|\sin\Theta(\widehat{V}, V)\| = \sqrt{1 - \sigma_{\min}^2(\widehat{V}^T V)} = \|\widehat{V}^T V_\perp\|,
$$

$$
\|\sin\Theta(\widehat{V}, V)\|_{\mathrm{F}} = \sqrt{r - \|V^T\widehat{V}\|_{\mathrm{F}}^2} = \|\widehat{V}^T V_\perp\|_{\mathrm{F}}.
$$

2. *(Triangle Inequality) For all $V_1, V_2, V_3 \in \mathbb{O}_{p,R}$,*

$$
\|\sin\Theta(V_2, V_3)\| \leq \|\sin\Theta(V_1, V_2)\| + \|\sin\Theta(V_1, V_3)\|,
$$

$$
\|\sin\Theta(V_2, V_3)\|_{\mathrm{F}} \leq \|\sin\Theta(V_1, V_2)\|_{\mathrm{F}} + \|\sin\Theta(V_1, V_3)\|_{\mathrm{F}}.
$$

3. *(Equivalence with Other Metrics)*

$$
\|\sin\Theta(\widehat{V}, V)\| \leq \sqrt{2}\|\sin\Theta(\widehat{V}, V)\|,
$$

$$
\|\sin\Theta(\widehat{V}, V)\|_{\mathrm{F}} \leq \sqrt{2}\|\sin\Theta(\widehat{V}, V)\|_{\mathrm{F}},
$$

$$
\|\sin\Theta(\widehat{V}, V)\| \leq \|\widehat{V}\widehat{V}^\top - VV^\top\| \leq 2\|\sin\Theta(\widehat{V}, V)\|,
$$

$$
\|\widehat{V}\widehat{V}^\top - VV^\top\|_{\mathrm{F}} = \sqrt{2}\|\sin\Theta(\widehat{V}, V)\|_{\mathrm{F}}.
$$

Theorem 26 (Mirsky's singular value inequality in Mirsky (1960))**.** *For any matrices $A, B \in \mathbb{R}^{m\times n}$, let $A = V_1\Sigma(A)W_1^\top$ and $B = V_2\Sigma(B)W_2^\top$ be the full SVDs of A and B, respectively. Note that $\Sigma(A), \Sigma(B) \in \mathbb{R}^{m\times n}$ are non-negative (rectangular) diagonal matrices whose diagonal entries are the non-increasingly ordered singular values of A and B, respectively. Then*

$$
\|\Sigma(A) - \Sigma(B)\| \leq \|A - B\| \tag{57}
$$

for any unitarily invariant norm $\|\cdot\|$ on $\mathbb{R}^{m\times n}$.

Theorem 27 (Weyl's Inequality for Singular Values). *Let $A, B \in \mathbb{R}^{m\times n}$ and denote their singular values (in nonincreasing order) by $\{\sigma_i(A)\}$ and $\{\sigma_i(B)\}$ respectively. In addition denote the singular values of $A+B$ as $\{\sigma_i(A+B)\}$. Then for all indices i, j satisfying $i+j-1 \leq \min\{m, n\}$,*

$$\sigma_{i+j-1}(A+B) \leq \sigma_i(A) + \sigma_j(B).$$

Lemma 28 (Ky Fan-type Inequality for Sums of Matrices). *Let $A, B \in \mathbb{R}^{m\times n}$ and denote their singular values (in nonincreasing order) by $\{\sigma_i(A)\}$ and $\{\sigma_i(B)\}$ respectively. In addition denote the singular values of $A+B$ as $\{\sigma_i(A+B)\}$. Then for any $1 \leq k \leq \min\{m, n\}$, it holds that*

$$\sqrt{\sum_{i=1}^{k} \sigma_i^2(A+B)} \leq \sqrt{\sum_{i=1}^{k} \sigma_i^2(A)} + \sqrt{\sum_{i=1}^{k} \sigma_i^2(B)}.$$

Proof. For a symmetric matrix $M \in \mathbb{R}^{n\times n}$, by Ky Fan's maximum principle (see e.g. II.1.13 in Bhatia (2013)), for any $1 \leq k \leq n$,

$$\sum_{i=1}^{k} \lambda_i(M) = \sup_{P \in \mathbb{O}^{n\times k}} \operatorname{tr}(P^\top M P).$$

Therefore

$$\sum_{i=1}^{k} \sigma_i^2(A) = \sum_{i=1}^{k} \lambda_i(A^\top A) = \sup_{P \in \mathbb{O}^{n\times k}} \operatorname{tr}(P^\top A^\top A P) = \sup_{P \in \mathbb{O}^{n\times k}} \|AP\|_{\mathrm{F}}^2,$$

and so

$$\sqrt{\sum_{i=1}^{k} \sigma_i^2(A)} = \sup_{P \in \mathbb{O}^{n\times k}} \|AP\|_{\mathrm{F}}.$$

Then

$$\begin{aligned}\sqrt{\sum_{i=1}^{k} \sigma_i^2\,(A+B)} &= \max_{U \in \mathbb{O}_{n,k}} \|(A+B)U\|_{\mathrm{F}} \leq \max_{U \in \mathbb{O}_{n,k}} \|AU\|_{\mathrm{F}} + \max_{U \in \mathbb{O}_{n,k}} \|BU\|_{\mathrm{F}} \\ &= \sqrt{\sum_{i=1}^{k} \sigma_i^2\,(A)} + \sqrt{\sum_{i=1}^{k} \sigma_i^2\,(B)}.\end{aligned}$$

□

Let $\mathcal{T}_{(r_1,r_2,r_3)}$ denote the class of tensor in $\mathbb{R}^{p_1\times p_2\times p_3}$ with tucker ranks at most (r_1, r_2, r_3). More precisely

$$\mathcal{T}_{(r_1,r_2,r_3)} = \{A \in \mathbb{R}^{p_1\times p_2\times p_3} : \operatorname{rank}(\mathcal{M}_k(A)) \leq r_k, k = 1, 2, 3\}.$$

Lemma 29. *Let $X^* \in \mathbb{R}^{p_1\times p_2\times p_3}$. For $k \in \{1, 2, 3\}$, suppose the k-th matricization of X^* satisfies*

$$\mathcal{M}_k(X^*) = \begin{bmatrix} U_k^* & U_{k\perp}^* \end{bmatrix} \begin{bmatrix} \Sigma_k^* & 0 \\ 0 & \Sigma_{k\perp}^* \end{bmatrix} \begin{bmatrix} V_k^* & V_{k\perp}^* \end{bmatrix}^\top$$

where $U_k^* \in \mathbb{O}_{p_k,r_k}$ corresponds to the the top r_k singular vectors of $\mathcal{M}_k(X^*)$. Then for $k \in \{1, 2, 3\}$, it holds that

$$\|X^* \times_k U_{k\perp}^*\|_{\mathrm{F}} = \left\|X^* \times_k (I_{p_k} - \mathcal{P}_{U_k^*})\right\|_{\mathrm{F}} = \sqrt{\sum_{j=r_k+1}^{\operatorname{rank}(\mathcal{M}_k(X^*))} \sigma_j^2(\mathcal{M}_k(X^*))} \leq \xi_{(r_1,r_2,r_3)},$$

where

$$\xi_{(r_1,r_2,r_3)} = \inf_{A \in \mathcal{T}_{(r_1,r_2,r_3)}} \|A - X^*\|_{\mathrm{F}}.$$

Proof. By symmetry, it suffices to consider $k = 1$. Note that

$$\left\|X^* \times_k (I_{p_k} - \mathcal{P}_{U_k^*})\right\|_{\mathrm{F}} = \|X^* \times_k \mathcal{P}_{U_{k\perp^*}}\|_{\mathrm{F}} = \|X^* \times_k U_{k\perp}^*\|_{\mathrm{F}} .$$

In addition

$$\begin{aligned}\left\|X^* \times_1 (I_{p_1} - \mathcal{P}_{U_1^*})\right\|_{\mathrm{F}} &= \left\|(I_{p_1} - \mathcal{P}_{U_1^*}) \cdot \mathcal{M}_1(X^*)\right\|_{\mathrm{F}} = \left\|U_{1\perp}^* U_{1\perp}^{*\top} \cdot (U_1^* \Sigma_1^* V_1^{*\top} + U_{1\perp}^* \Sigma_{1\perp}^* V_{1\perp}^{*\top})\right\|_{\mathrm{F}} \\ &= \left\|U_{1\perp}^* U_{1\perp}^{*\top} \cdot (U_{1\perp}^* \Sigma_{1\perp}^* V_{1\perp}^{*\top})\right\|_{\mathrm{F}} = \sqrt{\sum_{j=r_1+1}^{\mathrm{rank}(\mathcal{M}_1(X^*))} \sigma_j^2(\mathcal{M}_1(X^*))}.\end{aligned}$$

Note that by the properties of SVD, for any $W \in \mathbb{R}^{p_1 \times p_2 p_3}$ such that $\mathrm{rank}(W) \leq r_1$, it holds that

$$\sqrt{\sum_{j=r_1+1}^{\mathrm{rank}(\mathcal{M}_1(X^*))} \sigma_j^2(\mathcal{M}_1(X^*))} \leq \|\mathcal{M}_1(X^*) - W\|_{\mathrm{F}}.$$

For any $A \in \mathcal{T}_{(r_1,r_2,r_3)}$, it holds that $\mathrm{rank}(\mathcal{M}_1(A)) \leq r_1$. Therefore for any $A \in \mathcal{T}_{(r_1,r_2,r_3)}$,

$$\sqrt{\sum_{j=r_1+1}^{\mathrm{rank}(\mathcal{M}_1(X^*))} \sigma_j^2(\mathcal{M}_1(X^*))} \leq \|\mathcal{M}_1(X^*) - A\|_{\mathrm{F}}.$$

Taking the inf over all $A \in \mathcal{T}_{(r_1,r_2,r_3)}$, it follows that

$$\sqrt{\sum_{j=r_1+1}^{\mathrm{rank}(\mathcal{M}_1(X^*))} \sigma_j^2(\mathcal{M}_1(X^*))} \leq \xi_{(r_1,r_2,r_3)}.$$

□