

BIAS–VARIANCE TRADEOFF IN TENSOR ESTIMATION

Anonymous authors

Paper under double-blind review

ABSTRACT

We study denoising of a third-order tensor when the ground-truth tensor is **not** necessarily Tucker low-rank. Specifically, we observe

$$Y = X^* + Z \in \mathbb{R}^{p_1 \times p_2 \times p_3},$$

where X^* is the ground-truth tensor, and Z is the noise tensor. We propose a simple variant of the higher-order SVD (HOSVD) estimator, denoted \tilde{X} . We show that, uniformly over any user-specified Tucker ranks (r_1, r_2, r_3) for \tilde{X} , with high probability,

$$\|\tilde{X} - X^*\|_{\text{F}}^2 = O\left(\kappa^2 \left\{ r_1 r_2 r_3 + \sum_{k=1}^3 p_k r_k \right\} + \xi_{(r_1, r_2, r_3)}^2\right).$$

Here, the bias term $\xi_{(r_1, r_2, r_3)}$ corresponds to the best achievable approximation error of X^* over the class of tensors with Tucker ranks (r_1, r_2, r_3) ; κ^2 quantifies the noise level; and the variance term $\kappa^2 \{ r_1 r_2 r_3 + \sum_{k=1}^3 p_k r_k \}$ scales with the effective number of free parameters in the estimator \tilde{X} . Our analysis achieves a clean rank-adaptive bias–variance tradeoff: as we increase the ranks of estimator \tilde{X} , the bias $\xi_{(r_1, r_2, r_3)}$ decreases and the variance increases. As a byproduct we also obtain a convenient bias-variance decomposition for the vanilla low-rank SVD matrix estimators.

1 INTRODUCTION

Low-rank tensor models are a cornerstone of modern machine learning. They capture essential structure in high-dimensional signals while offering substantial gains in computation and storage. Applications include recommender systems (Koren et al., 2009), topic modeling (Blei et al., 2003), community detection (Abbe, 2018; Diakonikolas & Kane, 2024), and, more recently, latent variable learning (Anandkumar et al., 2014; Sherman & Kolda, 2020; Zhang & Kileel, 2023; Diakonikolas & Kane, 2024) and generative modeling (Hur et al., 2023; Peng et al., 2023). The effectiveness of these methods demonstrate that many real-world signals embedded in high-dimensional ambient spaces can be well-approximated by low-rank structures.

The study of exactly low-rank estimation is well-developed, particularly in the matrix setting. A large body of work has introduced efficient algorithms for tasks such as matrix completion and denoising, including nuclear norm relaxation (Koltchinskii et al., 2011), convex optimization (Candes & Recht, 2012), nonconvex methods (Chi et al., 2019), and SVD guarantees with unbalanced matrices (Cai & Zhang, 2018).

Parallel efforts in low-rank tensor estimation are well-developed, including the study of higher-order tensor singular value decompositions in the non-random setting (De Lathauwer et al., 2000b;a), **symmetric tensor decomposition in non-random setting** (Jin et al., 2024), and Riemannian optimization methods for low-rank Tucker tensor completion (Kressner et al., 2014). Nevertheless, the tensor setting poses additional challenges: many fundamental problems are computationally intractable, multiple inequivalent notions of rank (e.g. CP and Tucker) exist, and best low-rank approximations need not be unique (e.g. Hillar & Lim, 2013). Most theoretical guarantees in tensor estimation (e.g. Anandkumar et al., 2017; Zhang & Xia, 2018; Han et al., 2022), like their matrix counterparts, rely on the assumption that the underlying signal is exactly low-rank.

In practice, however, signals are not exactly low-rank: their singular spectra typically decay gradually rather than vanishing at a finite rank. This creates an interesting gap between theory and practice, since existing analyses provide guarantees under idealized exact low-rankness and leave open the question of how to analyze estimation error in more realistic settings.

We address this gap by designing and analyzing low-rank tensor estimation when the signal is not low-rank. In such settings, estimation error is governed by a *bias–variance tradeoff*: truncating to a smaller rank incurs approximation bias, while larger ranks inflate variance by amplifying noise. In particular, we propose a simple variant of the higher-order singular value decomposition (HOSVD) algorithm and provide the analysis of tensor estimation without assuming low-rankness. Our main result (Theorem 1) establishes an explicit *bias–variance decomposition* of the estimation error:

- The *bias term* quantifies the approximation error incurred by truncating a spectrum.
- The *variance term* matches the optimal rate for exactly low-rank models.

The success of SVD and HOSVD has been well established by their decades of widespread use. The objective of our manuscript is to provide mathematical justification for the information-efficient bias-variance tradeoff achieved by these two classical algorithms, and to establish uniform performance guarantees that hold for any user-specified ranks.

To achieve this clean bias-variance tradeoff, we utilize classical linear algebra results such as Mirsky’s and Ky Fan’s theorems to develop non-trivial extensions of classical perturbation tools to general tensors. As a convenient side result, we also recover a clean bias–variance characterization for matrices, thereby unifying the matrix and tensor perspectives. Finally, we validate our theory with simulations that confirm the existence of tradeoffs across different regimes of spectral decay and noise.

1.1 NOTATION

Matrices: For positive integers p, r , let $\mathbb{O}^{p \times r} = \{V \in \mathbb{R}^{p \times r} : V^\top V = I_r\}$, the set of column-orthonormal matrices. Let $M \in \mathbb{R}^{p \times q}$ and suppose the full singular value decomposition (SVD) satisfies $M = U\Sigma V^\top$. Here $s = \text{rank}(M)$, $U \in \mathbb{O}^{p \times s}$, $V \in \mathbb{O}^{q \times s}$, and $\Sigma \in \mathbb{R}^{s \times s}$ is diagonal with singular values $\sigma_1(M) \geq \sigma_2(M) \geq \dots \geq \sigma_s(M) \geq 0$. We write the smallest singular value of M as $\sigma_{\min}(M)$. The operator norm and Frobenius norm of M are defined as

$$\|M\| = \sigma_1(M), \quad \|M\|_F = \left(\sum_{i=1}^p \sum_{j=1}^q M_{i,j}^2 \right)^{1/2}.$$

For $r \leq \text{rank}(M)$, the rank- r truncated SVD is $M_{(r)} = U_{(r)}\Sigma_{(r)}V_{(r)}^\top$, where $U_{(r)} \in \mathbb{O}^{p \times r}$ and $V_{(r)} \in \mathbb{O}^{q \times r}$ corresponds to the leading r left and right singular vectors respectively, and $\Sigma_{(r)} = \text{diag}\{\sigma_1(M), \dots, \sigma_r(M)\}$. For brevity, we use $\text{SVD}_r(M)$ to denote the leading r left singular vectors, i.e.

$$\text{SVD}_r(M) = U_{(r)}.$$

For two matrices $M_1 \in \mathbb{R}^{p_1 \times q_1}$ and $M_2 \in \mathbb{R}^{p_2 \times q_2}$, their Kronecker product is $M_1 \otimes M_2 \in \mathbb{R}^{(p_1 p_2) \times (q_1 q_2)}$.

Let $U, \widehat{U} \in \mathbb{O}^{p \times r}$ be two singular subspaces. We write the principal angles between U and \widehat{U} as

$$\Theta(U, \widehat{U}) = \text{diag}\{\arccos(\sigma_1(U^\top \widehat{U})), \dots, \arccos(\sigma_r(U^\top \widehat{U}))\}.$$

We use $\|\sin \Theta(U, \widehat{U})\|$ and $\|\sin \Theta(U, \widehat{U})\|_F$ to measure the distance between the two singular subspaces.

Tensors: For any tensor $B \in \mathbb{R}^{p_1 \times p_2 \times p_3}$, the Frobenius norm is

$$\|B\|_F = \left(\sum_{i_1=1}^{p_1} \sum_{i_2=1}^{p_2} \sum_{i_3=1}^{p_3} B_{i_1, i_2, i_3}^2 \right)^{1/2}.$$

The mode-1 matricization $\mathcal{M}_1(B)$ is the unfolding of B into a $p_1 \times p_2 p_3$ matrix. The mode-2 and mode-3 matricization of B are defined similarly. The mode-1 product of B with a matrix

108 $M \in \mathbb{R}^{q \times p_1}$ is defined as

$$109 \quad (B \times_1 M)(j, i_2, i_3) = \sum_{i_1=1}^{p_1} M(i_1, j) B(i_1, i_2, i_3).$$

110 We write the Tucker rank of B as (r_1, r_2, r_3) if

$$111 \quad \text{rank}(\mathcal{M}_j(B)) = r_j, \text{ for } j = 1, 2, 3.$$

112 **Random variables:** For a random variable $X \in \mathbb{R}$, we denote the sub-Gaussian and sub-Exponential norms as

$$113 \quad \|X\|_{\psi_2} = \inf\{K > 0 : \mathbb{E} \exp(X^2/K^2) \leq 2\} \quad \text{and} \quad \|X\|_{\psi_1} = \inf\{K > 0 : \mathbb{E} \exp(|X|/K) \leq 2\}.$$

114 We write $X \sim \text{subGaussian}(0, \kappa^2)$ if $\mathbb{E}(X) = 0$ and $\|X\|_{\psi_2} \leq \kappa$.

115 **Universal constants:** We use C_1, C_2, \dots and c_1, c_2, \dots to denote positive constants whose values may differ from place to place.

124 2 BIAS-VARIANCE TRADEOFF IN TENSOR ESTIMATION

125 We consider the noisy tensor model

$$126 \quad Y = X^* + Z \in \mathbb{R}^{p_1 \times p_2 \times p_3}, \quad (1)$$

127 where X^* is an unknown signal and Z is a random perturbation. Even when X^* has a full Tucker rank (p_1, p_2, p_3) , it is common to approximate X^* by a Tucker rank- (r_1, r_2, r_3) estimator. This is because (i) low-rank structure yields dramatic gains in computation and memory, see Remark 2; and (ii) the approximation bias can decay quickly when approximating X^* by a Tucker low-rank tensor, so balancing bias and variance may lead to a smaller estimation error. Given a target Tucker rank (r_1, r_2, r_3) , we consider the following simple variant of the HOSVD algorithm (De Lathauwer et al., 2000a).

136 Algorithm 1 One-step HOSVD

137 **INPUT:** Tensor Y ; target Tucker rank (r_1, r_2, r_3) .

138 **for** $k = 1, 2, 3$ **do**

$$139 \quad U_k^{(0)} \leftarrow \text{SVD}_{r_k}(\mathcal{M}_k(Y)) \in \mathbb{O}^{p_k \times r_k}.$$

140 **end for**

$$141 \quad U_1^{(1)} \leftarrow \text{SVD}_{r_1}(\mathcal{M}_1(Y) \cdot \{U_2^{(0)} \otimes U_3^{(0)}\}) \in \mathbb{O}^{p_1 \times r_1},$$

$$142 \quad U_2^{(1)} \leftarrow \text{SVD}_{r_2}(\mathcal{M}_2(Y) \cdot \{U_1^{(0)} \otimes U_3^{(0)}\}) \in \mathbb{O}^{p_2 \times r_2},$$

$$143 \quad U_3^{(1)} \leftarrow \text{SVD}_{r_3}(\mathcal{M}_3(Y) \cdot \{U_1^{(0)} \otimes U_2^{(0)}\}) \in \mathbb{O}^{p_3 \times r_3}.$$

144 **OUTPUT:** $\tilde{X} \leftarrow Y \times_1 U_1^{(1)} U_1^{(1)\top} \times_2 U_2^{(1)} U_2^{(1)\top} \times_3 U_3^{(1)} U_3^{(1)\top}$.

148 Let $\mathcal{T}_{(r_1, r_2, r_3)}$ denote the class of tensors in $\mathbb{R}^{p_1 \times p_2 \times p_3}$ with Tucker rank at most (r_1, r_2, r_3) , i.e.

$$149 \quad \mathcal{T}_{(r_1, r_2, r_3)} = \{A \in \mathbb{R}^{p_1 \times p_2 \times p_3} : \text{rank}(\mathcal{M}_k(A)) \leq r_k, k = 1, 2, 3\}. \quad (2)$$

150 For any tensor $X^* \in \mathbb{R}^{p_1 \times p_2 \times p_3}$, let $\xi_{(r_1, r_2, r_3)}$ denote the best Tucker rank- (r_1, r_2, r_3) approximation error:

$$151 \quad \xi_{(r_1, r_2, r_3)} = \inf_{A \in \mathcal{T}_{(r_1, r_2, r_3)}} \|A - X^*\|_{\mathbb{F}}.$$

152 **Theorem 1.** Suppose Y follows (1) with $Z_{\mu_1, \mu_2, \mu_3} \stackrel{\text{i.i.d.}}{\sim} \text{subGaussian}(0, \kappa^2)$. Let \tilde{X} be the output of Algorithm 1 with target Tucker rank (r_1, r_2, r_3) . Define $r_{\max} = \max_k r_k$ and $p_{\min} = \min_k p_k$. Assume the singular gaps satisfy

$$153 \quad \{\sigma_{r_k}(\mathcal{M}_k(X^*)) - \sigma_{r_k+1}(\mathcal{M}_k(X^*))\}^2 \geq C_{\text{gap}} \kappa^2 \left(\sqrt{p_1 p_2 p_3 r_{\max}} + \sum_{k=1}^3 p_k r_{\max} \right), \quad (3)$$

for all $k = 1, 2, 3$ and for a sufficiently large constant $C_{\text{gap}} > 0$. Then, with probability at least $1 - C_1 p_1 p_2 p_3 \exp(-C_2 p_{\min})$,

$$\|\tilde{X} - X^*\|_{\text{F}} \leq C_3 \left\{ \sqrt{\kappa^2 \left(\sum_{k=1}^3 p_k r_k + r_1 r_2 r_3 \right)} + \xi_{(r_1, r_2, r_3)} \right\}. \quad (4)$$

Remark 1 (Assumptions in Theorem 1). *In Theorem 1, we assume that the entries of the noise tensor Z is i.i.d. sub-Gaussian. This is a commonly seen condition in the tensor literature (e.g. Zhang & Xia, 2018; Han et al., 2022).*

The signal-to-noise ratio (SNR) condition (3) in Theorem 1 is also commonly seen in literature. In fact, if the ground truth X^ is exactly Tucker low-rank with Tucker rank (r_1, r_2, r_3) , then $\sigma_{r_k+1}(\mathcal{M}_k(X^*)) = 0$ and (3) reduces to the SNR condition used in Zhang & Xia (2018), when r_{\max} is a bounded constant.*

Remark 2 (Computational efficiency). *Suppose $Y \in \mathbb{R}^{p_1 \times p_2 \times p_3}$ and let*

$$Y_{(r_1, r_2, r_3)} = S \times_1 U_1 \times_2 U_2 \times_3 U_3,$$

be a Tucker rank- (r_1, r_2, r_3) approximation of Y , where $U_k \in \mathbb{R}^{p_k \times r_k}$ satisfying $U_k^\top U_k = I_{r_k}$ for $k = 1, 2, 3$, and $S \in \mathbb{R}^{r_1 \times r_2 \times r_3}$ is the core tensor:

For dense multiplications with $v_k \in \mathbb{R}^{p_k}$ along each mode k , computing $Y \times_1 v_1 \times_2 v_2 \times_3 v_3$ costs $O(p_1 p_2 p_3)$ operations. In contrast, using the Tucker decomposition, we compute $Y_{(r_1, r_2, r_3)} \times_1 v_1 \times_2 v_2 \times_3 v_3$ successively as

$$t_k = U_k^\top v_k \in \mathbb{R}^{r_k} \quad (\text{cost} \approx 2 \sum_{k=1}^3 p_k r_k), \quad S' = S \times_1 t_1 \in \mathbb{R}^{r_2 \times r_3} \quad (\text{cost} \approx 2r_1 r_2 r_3),$$

$$s = t_2^\top S' \in \mathbb{R}^{r_3} \quad (\text{cost} \approx 2r_2 r_3), \quad w = s^\top t_3 \in \mathbb{R} \quad (\text{cost} \approx r_3).$$

Thus, vector multiplication with Tucker factors reduce the computational cost from $O(p_1 p_2 p_3)$ to $O(\sum_{k=1}^3 p_k r_k + r_1 r_2 r_3)$. The storage requirements of Tucker decomposition also drop significantly. Storing the full tensor Y requires $O(p_1 p_2 p_3)$ memory, while the Tucker representation requires only $\sum_{k=1}^3 p_k r_k + r_1 r_2 r_3$ scalar storage.

Therefore, both computation and storage of Tucker decomposition scale with $O(\sum_{k=1}^3 p_k r_k + r_1 r_2 r_3)$, which is substantially smaller than the dense case when $r_k \ll p_k$.

Adaptivity in Theorem 1. For any user specified target Tucker rank (r_1, r_2, r_3) , the estimator \tilde{X} satisfies (4). In other words, the result holds uniformly over all target Tucker ranks.

Bias–variance tradeoff. The error bound (4) exhibits a clear bias–variance decomposition for Tucker rank- (r_1, r_2, r_3) estimators. The bias term $\xi_{(r_1, r_2, r_3)}$ is the best achievable approximation error of X^* at the chosen ranks and decreases as the r_k increase. Therefore, as the ranks grow, variance increases while bias shrinks. Balancing these two leads to improved accuracy.

Proof techniques. For the bias, we leverage classical linear algebra tools, such as Mirsky’s and Ky Fan’s theorems, to characterize the optimal approximation error appearing in (4). To handle the variance of the estimator \tilde{X} , we adopt existing techniques from Zhang & Xia (2018).

Theorem 2. *Let $r_1, r_2, r_3, p_1, p_2, p_3$ be arbitrary integers such that $r_i \leq \min\{p_i, \prod_{j \neq i} p_j\}$ for all $i \in \{1, 2, 3\}$. Let $\mathcal{T}_{(r_1, r_2, r_3)}$ be the class of Tucker low-rank tensor as in (2) and for $\xi > 0$, let*

$$\mathcal{T}_{(r_1, r_2, r_3)}^\xi = \left\{ B \in \mathbb{R}^{p_1 \times p_2 \times p_3} : \inf_{A \in \mathcal{T}_{(r_1, r_2, r_3)}} \|B - A\|_{\text{F}} \leq \xi \right\}.$$

Consider the model $Y = X^ + Z \in \mathbb{R}^{p_1 \times p_2 \times p_3}$ where the entries of Z are i.i.d. subGaussian $(0, \kappa^2)$. For any ranks (r_1, r_2, r_3) and any $0 \leq \xi \leq \kappa^2 p_1 p_2 p_3$, there exist a universal constant $c > 0$ such that*

$$\inf_{\tilde{X}} \sup_{X^* \in \mathcal{T}_{(r_1, r_2, r_3)}^\xi} \mathbb{E} \|\tilde{X} - X^*\|_{\text{F}}^2 \geq c \left\{ \kappa^2 \left(\sum_{k=1}^3 p_k r_k + r_1 r_2 r_3 \right) + \xi^2 \right\}. \quad (5)$$

Tightness. Our upper bound in (4) attains the minimax-optimal rate (up to constants) over the class of tensors with Tucker rank at most (r_1, r_2, r_3) . In particular, the lower bound in Theorem 2 matches this rate over the same model class, showing that no estimator can uniformly achieve a strictly smaller error. Thus, our bound is sharp and the proposed procedure is minimax optimal.

3 BIAS–VARIANCE TRADEOFF IN MATRIX ESTIMATION

We consider the model

$$Y = X^* + Z \in \mathbb{R}^{m \times n}, \quad (6)$$

where X^* is an unknown matrix with arbitrary rank, and Z is a perturbation (noise) matrix. Even when $\text{rank}(X^*)$ is not small, it is common to approximate X^* by a rank- r estimator, because (i) low-rank decomposition can dramatically increase the computational and memory efficiency, see Remark 3 for more details; and (ii) the approximation bias can be small, so balancing the bias and variance may lead to smaller estimation error. Let

$$X^* = \sum_{i=1}^{\min\{m,n\}} \sigma_i(X^*) u_i^* v_i^{*\top}, \quad X_{(r)}^* = \sum_{i=1}^r \sigma_i(X^*) u_i^* v_i^{*\top}.$$

By the Eckart–Young–Mirsky theorem (Eckart & Young, 1936), $X_{(r)}^*$ is the best rank- r approximation to X^* in Frobenius norm in the sense that

$$\|X^* - X_{(r)}^*\|_F \leq \|X^* - W\|_F \text{ for any } W \in \mathbb{R}^{m \times n} \text{ with } \text{rank}(W) \leq r.$$

Suppose we observe Y instead of X^* , and write the SVD of Y as

$$Y = \sum_{i=1}^{\min\{m,n\}} \sigma_i(Y) u_i v_i^\top, \quad Y_{(r)} = \sum_{i=1}^r \sigma_i(Y) u_i v_i^\top. \quad (7)$$

That is, $Y_{(r)}$ is the rank- r truncation of the SVD of Y . We now provide an upper bound between $Y_{(r)}$ and the unknown matrix X^* .

Theorem 3. *Under (6), let $Y_{(r)}$ be defined as in (7). Then*

$$\|Y_{(r)} - X^*\|_F \leq C(\sqrt{r}\|Z\| + \xi_{(r)}), \quad (8)$$

where $\xi_{(r)} = \|X_{(r)}^* - X^*\|_F = \sqrt{\sum_{i=r+1}^{\min\{m,n\}} \sigma_i^2(X^*)}$.

Adaptivity in Theorem 3. For any prescribed rank r , the rank- r truncated SVD estimator $Y_{(r)}$ satisfies the guarantee in Theorem 3. In other words, the theorem holds uniformly over all $r \geq 1$, so it applies to the vanilla SVD at any user-specified target rank.

Constant in (8). Tracing the proof of Theorem 3 shows that one may take the explicit constant $C = 2 + \sqrt{2} \approx 3.414$ in (8). We did not attempt to optimize this further.

Bias–variance tradeoff. Theorem 3 exhibits a clear bias–variance tradeoff for rank- r estimation. The bias term $\xi_{(r)} = \|X^* - X_{(r)}^*\|_F$ is the best possible rank- r approximation error of X^* and decreases as r increases. The standard deviation term, being order $(\sqrt{r}\|Z\|)$, captures the cost of estimating additional singular components under noise and increases with r .

Remark 3. *Suppose $Y \in \mathbb{R}^{m \times n}$ and let $Y_{(r)} = U_{(r)} \Sigma_{(r)} V_{(r)}^\top$ be a rank- r factorization (e.g. the truncated SVD), where $U_{(r)} \in \mathbb{R}^{m \times r}$, $V_{(r)} \in \mathbb{R}^{n \times r}$ satisfy $U_{(r)}^\top U_{(r)} = I_r$, $V_{(r)}^\top V_{(r)} = I_r$, and $\Sigma_{(r)} = \text{diag}(\sigma_1, \dots, \sigma_r) \in \mathbb{R}^{r \times r}$. For a dense multiplication with $v \in \mathbb{R}^n$, computing Yv costs $O(mn)$ operations. In contrast, using the factorization we compute successively*

$$t = V_{(r)}^\top v \in \mathbb{R}^r \text{ (cost } \approx 2nr), \quad s = \Sigma_{(r)} t \in \mathbb{R}^r \text{ (cost } \approx r), \quad w = U_{(r)} s \in \mathbb{R}^m \text{ (cost } \approx 2mr).$$

Hence $Y_{(r)}v = U_{(r)} \Sigma_{(r)} V_{(r)}^\top v$ can be computed in $O(mr + nr)$ operations.

Beyond computational savings, the SVD factorization also reduces storage costs. Storing the full matrix Y requires $O(mn)$ memory, whereas storing the factors $U_{(r)} \in \mathbb{R}^{m \times r}$, $\Sigma_{(r)} \in \mathbb{R}^{r \times r}$, and $V_{(r)} \in \mathbb{R}^{n \times r}$ requires only $O(mr + nr)$ scalars.

Therefore, both computation and storage using the truncated SVD are significantly smaller than in the full matrix case, especially when $r \ll \min\{m, n\}$.

In the following, we illustrate the application of Theorem 3 in three matrix estimation settings.

Corollary 4. Suppose (6) holds and that the entries of Z are independent sub-Gaussian random variables with

$$\|Z_{ij}\|_{\psi_2} \leq \kappa.$$

Then with probability at least $1 - \exp(-(m+n))$, there exists a universal constant $C > 0$ such that

$$\|Y_{(r)} - X^*\|_F \leq C(\|X_{(r)}^* - X^*\|_F + \kappa\sqrt{r(m+n)}).$$

Proof of Corollary 4. From Vershynin (2018) section 4.4.2, it follows that $\|Z\| \leq C_1\kappa\sqrt{m+n}$ with probability at least $1 - \exp(-(m+n))$. The desired result follows immediately from Theorem 3. \square

Corollary 5. Suppose $Z \in \mathbb{R}^{m \times n}$ is a sub-Gaussian random matrix in the sense that for any $v \in \mathbb{R}^m$ with $\|v\|_2 = 1$, and $w \in \mathbb{R}^n$ with $\|w\|_2 = 1$, it holds that

$$\|v^\top Z w\|_{\psi_2} \leq \kappa.$$

Then with probability at least $1 - \exp(-(m+n))$, there exists a universal constant $C > 0$ such that

$$\|Y_{(r)} - X^*\|_F \leq C(\|X_{(r)}^* - X^*\|_F + \kappa\sqrt{r(m+n)}).$$

Proof of Corollary 5. By Lemma 15, $\|Z\| \leq C_1\sqrt{m+n}$ with probability at least $1 - \exp(-(m+n))$. The desired result follows immediately from Theorem 3. \square

Corollary 6. Let $Z_1, \dots, Z_N \in \mathbb{R}^n$ be i.i.d. mean-zero sub-Gaussian random vectors with covariance Σ . Define the sample covariance matrix

$$Y = \frac{1}{N} \sum_{k=1}^N Z_k Z_k^\top,$$

and let $X^* = \mathbb{E}[Y]$ be the population covariance matrix. Suppose Z_k is sub-Gaussian in the sense that

$$\|u^\top Z_k\|_{\psi_2} \leq \kappa \quad \text{for all } u \in \mathbb{R}^n, \|u\|_2 = 1.$$

Then with probability at least $1 - 2\exp(-n)$, there exists a universal constant $C > 0$ such that

$$\|Y_{(r)} - X^*\|_F \leq C\left(\|X_{(r)}^* - X^*\|_F + \kappa^2\sqrt{r}\left[\sqrt{\frac{n}{N}} + \frac{n}{N}\right]\right).$$

Proof of Corollary 6. Let $\tilde{Z} := Y - X^*$. By Remark 4.7.3 of Vershynin (2018), with probability at least $1 - \exp(-cn)$ one has $\|\tilde{Z}\| \leq C_1\kappa^2\left[\sqrt{\frac{n}{N}} + \frac{n}{N}\right]$. Applying Theorem 3 with Z replaced by \tilde{Z} yields the desired bound. \square

Theorem 7. Let r, m, n be arbitrary integers such that $r \leq \min\{m, n\}$. For $\xi > 0$, let \mathcal{M}_r^ξ be such that

$$\mathcal{M}_r^\xi = \{B \in \mathbb{R}^{m \times n} : \text{there exists } A \text{ such that } \text{rank}(A) \leq r \text{ and } \|B - A\|_F \leq \xi\}.$$

Consider the model $Y = X^* + Z \in \mathbb{R}^{m \times n}$ where the entries of Z are i.i.d. $\text{subGaussian}(0, \kappa^2)$. Suppose in addition that $\xi^2 \leq (m-r)n\kappa^2$. Then there exist a universal constant $c > 0$ such that

$$\inf_{\hat{X}} \sup_{X^* \in \mathcal{M}_r^\xi} \mathbb{E}\|\hat{X} - X^*\|_F^2 \geq c\left\{\kappa^2(m+n)r + \xi^2\right\}.$$

The upper bound in (8) attains the minimax-optimal rate (up to constants) over the class of matrices with rank at most r . The lower bound in Theorem 7 matches this rate on the same model class.

4 NUMERICAL EXPERIMENTS

In many applied settings, selecting the single *best* rank r is inherently difficult in finite samples with unknown latent structure. Our analysis instead provides guarantees for singular-value decompositions in both the matrix (SVD) and tensor (HOSVD) regimes across a broad range of r . Consequently, rather than hinging on a brittle rank-selection heuristic, we empirically demonstrate that both truncated SVD and one-step HOSVD (Algorithm 1) are robust whenever ranks are chosen neither too small (underfitting/bias) nor too large (overfitting/variance). This is consistent with the bias–variance tradeoff formalized in Theorems 1 and 3. Moreover, in practice the choice of r , is frequently driven by diverse aims other than merely minimizing mean-squared error, including visualization, interpretability, or downstream decision tasks. This is reflected in widespread, long-standing use across biology/genomics (Price et al., 2006), economics (Filmer & Pritchett, 2001), and computer vision (Turk & Pentland, 1991).

We evaluate these predictions on (i) a real 3D brain-MRI dataset with controlled additive noise (IXI; T1-weighted volumes (IXI, 2002)), and (ii) simulated data for both matrices and third–order tensors.

4.1 REAL DATA: 3D BRAIN MRI WITH CONTROLLED NOISE

We evaluate on **five** T1-weighted brain MRI volume drawn from the IXI dataset (IXI, 2002). The selected **volumes are third-order tensors** of shape $256 \times 256 \times 150$. To probe low-rank behavior in a realistic preprocessing pipeline, **all volume are** prepared in two wavelet-smoothed variants using standard orthonormal families—(a) Daubechies-6 (db6) and (b) Symlet-8 (sym8) (Daubechies, 1988; 1992).

We implement separable 3D multilevel decompositions (`wavedecn`) with periodic boundary handling, and apply subband-wise soft-thresholding via BayesShrink (Chang et al., 2000; Donoho & Johnstone, 1994). The literature supports wavelet shrinkage as an effective denoising/smoothing step for MR magnitude images (Nowak, 1999; Wood & Johnson, 1999).

Noise model and ranks. For our prepared **volumes** X^* and each noise level $\lambda \in \{0.2, 0.3, 0.4, 0.5, 0.6, 0.7\}$ we generate

$$Y = X^* + Z, \quad Z_{ijk} \stackrel{\text{i.i.d.}}{\sim} \mathcal{N}(0, \kappa^2), \quad \kappa = \lambda \frac{\|X^*\|_F}{\sqrt{p_1 p_2 p_3}},$$

which keeps the per-entry SNR comparable across images and variants. We compute one-step HOSVD reconstructions (Algorithm 1) \tilde{X} at Tucker rank (r, r, r) for $r \in \{35, 50, 60, 65, 70, 80, 90, 95\}$.

Findings. Figure 1 shows representative mid-sagittal slices from the reconstructions of the wavelet-smoothed variants of a selected volume. As intermediate ranks, the images sharpen and cortical detail appears while background fluctuations remain suppressed. We provide selected quantitative results in Section 4.1, which show that all variants exhibit this expected progression.

4.2 SYNTHETIC DATA: TENSORS

We generate third-order tensors $X^* \in \mathbb{R}^{p \times p \times p}$ in the Tucker form

$$X^* = \mathcal{G} \times_1 U_1 \times_2 U_2 \times_3 U_3, \quad U_1, U_2, U_3 \in \mathbb{R}^{p \times p}, \quad U_1^\top U_1 = U_2^\top U_2 = U_3^\top U_3 = I_p,$$

where U_1, U_2, U_3 are drawn independently with orthonormal columns. The core tensor $\mathcal{G} \in \mathbb{R}^{p \times p \times p}$ is defined entrywise by

$$\mathcal{G}_{ijk} = \gamma_{ijk} \beta^{i+j+k}, \quad \beta = 0.8,$$

where the signs γ_{ijk} are independent with $\mathbb{P}(\gamma_{ijk} = -1) = 1/3, \mathbb{P}(\gamma_{ijk} = 1) = 2/3$. We observe $Y = X^* + Z$ with i.i.d. Gaussian noise $Z_{ijk} \sim \mathcal{N}(0, 1)$ entry-wise. The SNR is controlled by scaling the signal such that

$$\|X^*\|_F = \lambda \sqrt{p^3}, \quad \lambda \in \{10, 50\}.$$

The estimator \tilde{X} is obtained via the one-step HOSVD (Algorithm 1) at the Tucker rank (r, r, r) .

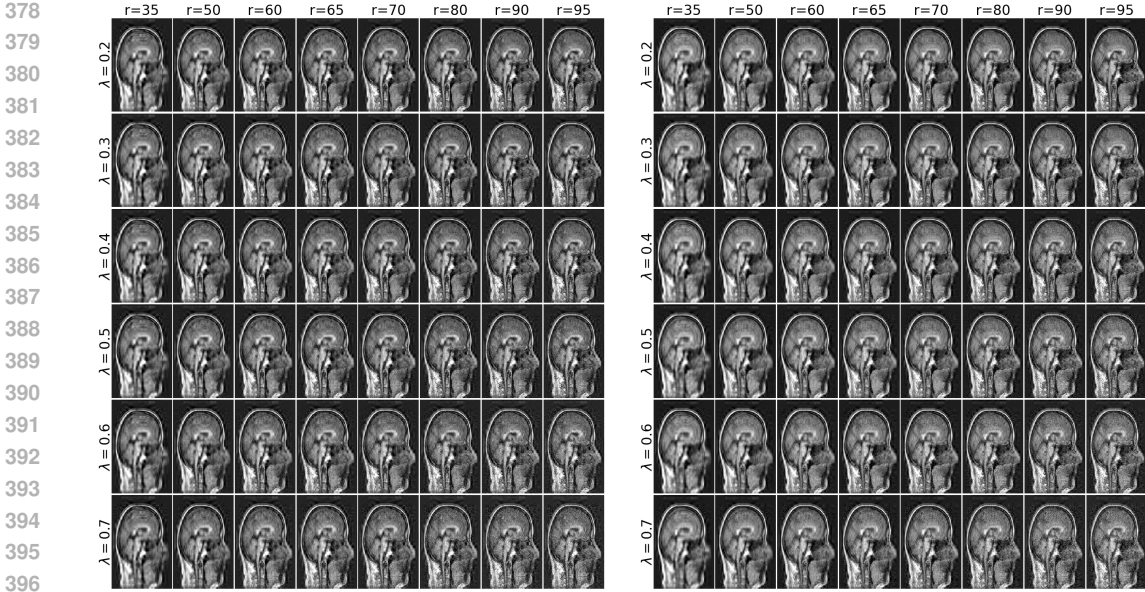


Figure 1: Mid-sagittal MRI slice of the reconstructed volume at selected ranks and noise levels λ for the `db6` variant (left), and for the `sym8` variant (right).

λ	r	original	db6	sym8
0.4	50	0.1753 (0.0330)	0.1747 (0.0537)	0.1681 (0.0466)
	65	0.1553 (0.0268)	0.1561 (0.0435)	0.1501 (0.0373)
	80	0.1490 (0.0183)	0.1499 (0.0291)	0.1451 (0.0245)
	95	0.1550 (0.0102)	0.1558 (0.0159)	0.1525 (0.0128)
0.5	50	0.1797 (0.0322)	0.1793 (0.0521)	0.1728 (0.0451)
	65	0.1655 (0.0248)	0.1668 (0.0400)	0.1610 (0.0341)
	80	0.1677 (0.0154)	0.1691 (0.0244)	0.1647 (0.0201)
	95	0.1840 (0.0073)	0.1847 (0.0116)	0.1819 (0.0089)
0.6	50	0.1851 (0.0311)	0.1850 (0.0503)	0.1787 (0.0434)
	65	0.1776 (0.0224)	0.1793 (0.0362)	0.1738 (0.0307)
	80	0.1890 (0.0124)	0.1904 (0.0200)	0.1866 (0.0162)
	95	0.2150 (0.0049)	0.2154 (0.0085)	0.2132 (0.0062)

Table 1: Mean relative reconstruction errors (entries are Mean (SE) across volumes).

Dimensions and evaluated ranks. We vary $p \in \{20, 50, 75, 100\}$ and for each p , we evaluate two ranks r as outlined in Table 2.

Each configuration (p, λ, r) is repeated $R = 50$ times with freshly drawn (U_1, U_2, U_3) and noise. Implementations use exact batched SVD on GPU via TensorLy (PyTorch backend) (Kossaifi et al., 2019). We report the sample mean and standard error of the relative Frobenius error

$$\text{RelErr}(\tilde{X}; X^*) = \frac{\|\tilde{X} - X^*\|_F}{\|X^*\|_F},$$

summarized on the right panel of Table 2. We observe that the error consistently decreases as the SNR parameter λ increases. Overall, one-step HOSVD is robust across the tested sizes and ranks, yielding accurate estimates on the synthetic tensors.

4.3 SYNTHETIC DATA: MATRICES

We examine the matrix analogue under a controlled SNR design that mirrors the tensor experiments. For each configuration, we draw an $m \times n$ latent signal

$$X^* = U \Sigma V^\top, \quad U \in \mathbb{R}^{m \times n}, \quad V \in \mathbb{R}^{n \times n}, \quad U^\top U = V^\top V = I_n,$$

where $\Sigma = \text{diag}(\sigma_1, \dots, \sigma_n)$ has exponentially decaying singular values, i.e. $\sigma_i = \beta^i$ with $\beta = 0.8$. We choose m and n as outlined in Table 2. We observe $Y = X^* + Z$, where $Z_{ij} \sim \mathcal{N}(0, 1)$ is i.i.d. Gaussian noise entry-wise. The SNR is controlled by scaling the signal such that

$$\|X^*\|_F = \lambda\sqrt{mn}, \quad \lambda \in \{10, 50\}.$$

This design varies SNR via λ while keeping the per-entry noise scale identical across sizes. Each configuration (m, n, λ, r) is repeated $R = 50$ times with independent draws of (U, V) and noise, and we report the mean and standard error of RelErr. We use the truncated SVD estimator $Y_{(r)}$ as in Theorem 3. Results are shown in the left panel of Table 2. The truncated SVD is robust over a wide range of selected ranks.

Table 2: Results for matrix (left) and tensor (right).

Matrix					Tensor					
λ	m	n	r	Mean (SE)	λ	p	r	Mean (SE)		
10	100	15	10	0.1178 (0.00125)	10	20	10	0.1420 (0.00350)		
			12	0.0906 (0.00164)			12	0.0977 (0.00208)		
	250	25	10	0.1254 (0.00044)		50	10	10	0.1370 (0.00347)	
			15	0.0868 (0.00077)				15	0.0459 (0.00102)	
	375	20	10	0.1279 (0.00045)		75	10	10	0.1367 (0.00348)	
			15	0.0927 (0.00076)				15	0.0427 (0.00113)	
	500	80	30	0.0698 (0.00032)		100	30	30	0.0203 (0.00008)	
			40	0.0795 (0.00035)				40	0.0294 (0.00009)	
	50	100	15	10		0.0844 (0.00007)	50	20	10	0.1361 (0.00381)
				12		0.0181 (0.00037)			12	0.0840 (0.00207)
250		25	10	0.1079 (0.00002)	50	10		10	0.1362 (0.00359)	
			15	0.0378 (0.00009)				15	0.0410 (0.00109)	
375		20	10	0.1068 (0.00002)	75	10		10	0.1368 (0.00374)	
			15	0.0349 (0.00008)				15	0.0412 (0.00098)	
500		80	30	0.0134 (0.00008)	100	30		30	0.0039 (0.00002)	
			40	0.0156 (0.00006)				40	0.0058 (0.00002)	

5 DISCUSSION

This work provides a rank-adaptive analysis of HOSVD-based tensor denoising without assuming exact low-rank. Our main theorem (Theorem 1) yields an explicit bias-variance decomposition

$$\|\tilde{X} - X^*\|_F \lesssim \kappa \sqrt{\sum_{k=1}^3 p_k r_k + r_1 r_2 r_3} + \xi_{(r_1, r_2, r_3)},$$

uniformly over all user-specified target Tucker ranks. The variance term scales with the effective degrees of freedom of the Tucker model, while the bias term is the best achievable approximation error at those ranks. Together with the matrix counterpart (Theorem 3), the results unify classical SVD intuition with the multilinear (tensor) setting and rigorously justify a practice common in applications: choose ranks large enough to suppress bias but not so large as to amplify noise. Our experiments on IXI brain MRI and controlled synthetic data corroborate this picture.

Future work. (i) *Beyond i.i.d. noise and full observations.* Extending the theory to heteroskedastic or correlated perturbations (e.g. spatially correlated fields, Rician-like MRI noise) and to incomplete observations (tensor completion, masked entries) is natural. We expect the variance term to inherit problem-dependent effective dimensions (e.g. leverage scores or sampling densities), while the bias term remains $\xi_{(r_1, r_2, r_3)}$. A key technical step is replacing isotropic concentration with mode-wise covariance-aware bounds for $\mathcal{M}_k(Z)$ and their projections. (ii) *Higher orders and higher ranks.* The proof strategy appears to scale to d th-order tensors with the clean generalization

$$\|\tilde{X} - X^*\|_F \lesssim \kappa \sqrt{\sum_{k=1}^d p_k r_k + \prod_{k=1}^d r_k} + \xi_{(r_1, \dots, r_d)},$$

under appropriately formulated spectral gaps for each unfolding. This would offer a unified bias-variance law across orders and ranks, and would help explain the strong empirical performance of HOSVD-style estimators in truly high-order settings.

REFERENCES

- 486
487
488 Ixi dataset. <http://brain-development.org/ixi-dataset/>, 2002. Public multi-
489 modal MRI; we randomly selected 50 T1-weighted volumes. Accessed 2025-09-09. Data col-
490 lected as part of the EPSRC GR/S21533/02. Available under a Creative Commons BY-SA 3.0
491 license.
- 492 Emmanuel Abbe. Community detection and stochastic block models: recent developments. *Journal*
493 *of Machine Learning Research*, 18(177):1–86, 2018.
- 494 Anima Anandkumar, Yuan Deng, Rong Ge, and Hossein Mobahi. Homotopy analysis for tensor
495 pca. In *Conference on Learning Theory*, pp. 79–104. PMLR, 2017.
- 496
497 Animashree Anandkumar, Rong Ge, Daniel J Hsu, Sham M Kakade, Matus Telgarsky, et al. Tensor
498 decompositions for learning latent variable models. *J. Mach. Learn. Res.*, 15(1):2773–2832, 2014.
- 499
500 J Calvin Berry. Minimax estimation of a bounded normal mean vector. *Journal of Multivariate*
501 *Analysis*, 35(1):130–139, 1990.
- 502
503 Rajendra Bhatia. *Matrix analysis*, volume 169. Springer Science & Business Media, 2013.
- 504 David M Blei, Andrew Y Ng, and Michael I Jordan. Latent dirichlet allocation. *Journal of machine*
505 *Learning research*, 3(Jan):993–1022, 2003.
- 506
507 T. Tony Cai and Anru Zhang. Rate-optimal perturbation bounds for singular subspaces with
508 applications to high-dimensional statistics. *Annals of Statistics*, 46(1):60–89, 2018. doi:
509 10.1214/17-AOS1541.
- 510 Emmanuel Candes and Benjamin Recht. Exact matrix completion via convex optimization. *Com-*
511 *munications of the ACM*, 55(6):111–119, 2012.
- 512
513 S. Grace Chang, Bin Yu, and Martin Vetterli. Adaptive wavelet thresholding for image denoising
514 and compression. *IEEE Transactions on Image Processing*, 9(9):1532–1546, 2000. doi: 10.1109/
515 83.862633.
- 516 Yuejie Chi, Yue M Lu, and Yuxin Chen. Nonconvex optimization meets low-rank matrix factoriza-
517 tion: An overview. *IEEE Transactions on Signal Processing*, 67(20):5239–5269, 2019.
- 518
519 Ingrid Daubechies. Orthonormal bases of compactly supported wavelets. *Communications on Pure*
520 *and Applied Mathematics*, 41(7):909–996, 1988. doi: 10.1002/cpa.3160410705.
- 521
522 Ingrid Daubechies. *Ten Lectures on Wavelets*. SIAM, Philadelphia, 1992. ISBN 978-0-89871-274-2.
- 523
524 Lieven De Lathauwer, Bart De Moor, and Joos Vandewalle. On the best rank-1 and rank-
525 (r_1, r_2, \dots, r_n) approximation of higher-order tensors. *SIAM Journal on Matrix Analysis and*
Applications, 21(4):1324–1342, 2000a. doi: 10.1137/S0895479898346995.
- 526
527 Lieven De Lathauwer, Bart De Moor, and Joos Vandewalle. A multilinear singular value decompo-
528 sition. *SIAM journal on Matrix Analysis and Applications*, 21(4):1253–1278, 2000b.
- 529
530 Ilias Diakonikolas and Daniel M Kane. Implicit high-order moment tensor estimation and learning
531 latent variable models. *arXiv preprint arXiv:2411.15669*, 2024.
- 532
533 David L. Donoho and Iain M. Johnstone. Ideal spatial adaptation by wavelet shrinkage. *Biometrika*,
534 81(3):425–455, 1994. doi: 10.1093/biomet/81.3.425.
- 535
536 David L Donoho, Richard C Liu, and Brenda MacGibbon. Minimax risk over hyperrectangles, and
537 implications. *The Annals of Statistics*, pp. 1416–1437, 1990.
- 538
539 Carl Eckart and Gale Young. The approximation of one matrix by another of lower rank. *Psychome-*
trika, 1(3):211–218, 1936.
- Deon Filmer and Lant H Pritchett. Estimating wealth effects without expenditure data—or tears: an
application to educational enrollments in states of india. *Demography*, 38(1):115–132, 2001.

- 540 Rungang Han, Rebecca Willett, and Anru R Zhang. An optimal statistical and computational frame-
541 work for generalized tensor estimation. *The Annals of Statistics*, 50(1):1–29, 2022.
- 542
- 543 Christopher J Hillar and Lek-Heng Lim. Most tensor problems are np-hard. *Journal of the ACM*
544 (*JACM*), 60(6):1–39, 2013.
- 545 Yoonhaeng Hur, Jeremy G Hoskins, Michael Lindsey, E Miles Stoudenmire, and Yuehaw Khoo.
546 Generative modeling via tensor train sketching. *Applied and Computational Harmonic Analysis*,
547 67:101575, 2023.
- 548
- 549 Ruhui Jin, Joe Kileel, Tamara G Kolda, and Rachel Ward. Scalable symmetric tucker tensor decom-
550 position. *SIAM Journal on Matrix Analysis and Applications*, 45(4):1746–1781, 2024.
- 551 Vladimir Koltchinskii, Karim Lounici, and Alexandre B. Tsybakov. Nuclear-norm penalization and
552 optimal rates for noisy low-rank matrix completion. *Annals of Statistics*, 39(5):2302–2329, 2011.
553 doi: 10.1214/11-AOS894.
- 554
- 555 Yehuda Koren, Robert Bell, and Chris Volinsky. Matrix factorization techniques for recommender
556 systems. *Computer*, 42(8):30–37, 2009.
- 557
- 558 Jean Kossaifi, Yannis Panagakis, Anima Anandkumar, and Maja Pantic. Tensorly: Tensor learning
559 in python. *Journal of Machine Learning Research*, 20(26):1–6, 2019.
- 560 Daniel Kressner, Michael Steinlechner, and Bart Vandereycken. Low-rank tensor completion by
561 riemannian optimization. *BIT Numerical Mathematics*, 54(2):447–468, 2014.
- 562
- 563 Leon Mirsky. Symmetric gauge functions and unitarily invariant norms. *The quarterly journal of*
564 *mathematics*, 11(1):50–59, 1960.
- 565 Robert D. Nowak. Wavelet-based rician noise removal for magnetic resonance imaging. *IEEE*
566 *Transactions on Image Processing*, 8(10):1408–1419, 1999. doi: 10.1109/83.791966.
- 567
- 568 Alp Ozdemir, Marisel Villafane-Delgado, David C Zhu, Mark A Iwen, and Selin Aviyente. Multi-
569 scale higher order singular value decomposition (ms-hosvd) for resting-state fmri compression
570 and analysis. In *2017 IEEE International Conference on Acoustics, Speech and Signal Processing*
571 (*ICASSP*), pp. 6299–6303. IEEE, 2017.
- 572 Alain Pajor. Metric entropy of the grassmann manifold. *Convex Geometric Analysis*, 34(181-188):
573 0942–46013, 1998.
- 574
- 575 Yifan Peng, Yian Chen, E Miles Stoudenmire, and Yuehaw Khoo. Generative modeling via hierar-
576 chical tensor sketching. *arXiv preprint arXiv:2304.05305*, 2023.
- 577
- 578 Alkes L Price, Nick J Patterson, Robert M Plenge, Michael E Weinblatt, Nancy A Shadick, and
579 David Reich. Principal components analysis corrects for stratification in genome-wide association
580 studies. *Nature genetics*, 38(8):904–909, 2006.
- 581 Samantha Sherman and Tamara G Kolda. Estimating higher-order moments using symmetric tensor
582 decomposition. *SIAM Journal on Matrix Analysis and Applications*, 41(3):1369–1387, 2020.
- 583
- 584 Matthew Turk and Alex Pentland. Eigenfaces for recognition. *Journal of cognitive neuroscience*, 3
585 (1):71–86, 1991.
- 586 Nick Vannieuwenhoven, Raf Vandebril, and Karl Meerbergen. A new truncation strategy for the
587 higher-order singular value decomposition. *SIAM Journal on Scientific Computing*, 34(2):A1027–
588 A1052, 2012.
- 589 Roman Vershynin. *High-dimensional probability: An introduction with applications in data science*,
590 volume 47. Cambridge university press, 2018.
- 591
- 592 Hermann Weyl. Das asymptotische verteilungsgesetz der eigenwerte linearer partieller differen-
593 tialgleichungen (mit einer anwendung auf die theorie der hohlraumstrahlung). *Mathematische*
Annalen, 71(4):441–479, 1912.

594 John C. Wood and William M. Johnson. Wavelet packet denoising of magnetic resonance images.
595 *Magnetic Resonance in Medicine*, 41(4):640–645, 1999. doi: 10.1002/(SICI)1522-2594(199904)
596 41:4<640::AID-MRM12>3.0.CO;2-8.
597
598 Anru Zhang and Dong Xia. Tensor svd: Statistical and computational limits. *IEEE Transactions on*
599 *Information Theory*, 64(11):7311–7338, 2018.
600 Yifan Zhang and Joe Kileel. Moment estimation for nonparametric mixture models through implicit
601 tensor decomposition. *SIAM Journal on Mathematics of Data Science*, 5(4):1130–1159, 2023.
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647

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

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

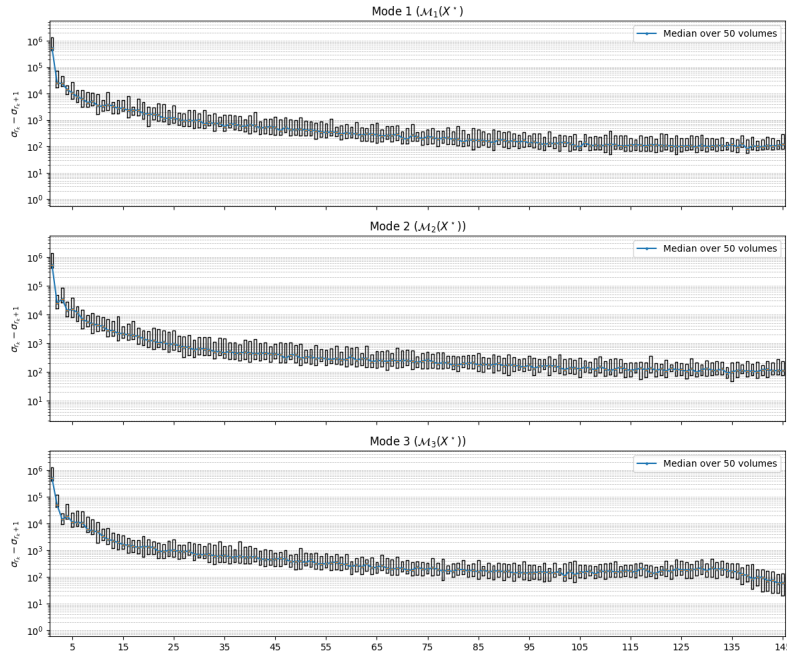


Figure 2: Empirical median and interquartile range (Q1–Q3) of the mode-wise singular-value gaps (log scale) over 50 T1-weighted IXI brain MRI volumes.

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 an 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 $\|Y - X^*\|_F$, we consider the synthetic tensor model of Section 4.2 with $p = 100$. For each candidate Tucker rank $r \in \{12, 15, 18, \dots, 84, 87\}$, we run 50 independent repetitions. The bias component is obtained as the error $\|\tilde{X} - X^*\|_F$, where \tilde{X} denotes the output of Algorithm 1 when applied directly to the noiseless tensor X^* . The variance component is obtained as

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

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

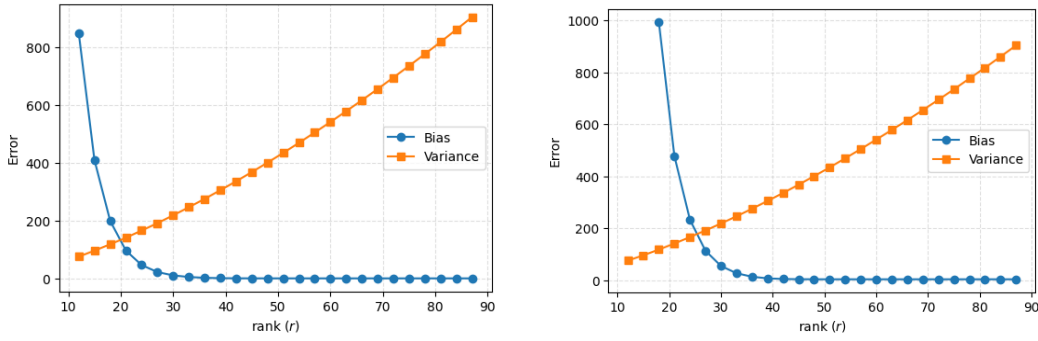


Figure 3: Mean bias and variance components of the error as functions of the target rank in the synthetic experiment with $p = 100$. The left panel represent the curve with scaling paramter $\lambda = 10$ and the right panel $\lambda = 50$.

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 $\tau = \alpha \|Y\|_F^2$ 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

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

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

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

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

$$\|B - B'\|_F^2 \leq r'_1 r_2 r_3 \|B\|_\infty^2 = r'_1 r_2 r_3 \kappa^4 \leq \xi^2.$$

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

$$\inf_{\hat{X}} \sup_{X^* \in \mathcal{T}_{(r_1, r_2, r_3)}^\xi} \mathbb{E} \|\hat{X} - X^*\|_F^2 \geq \inf_{\hat{X}} \sup_{X^* \in \mathcal{B}} \mathbb{E} \|\hat{X} - X^*\|_F^2. \quad (11)$$

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

$$\inf_{\hat{X}} \sup_{X^* \in \mathcal{B}} \mathbb{E} \|\hat{X} - X^*\|_F^2 = \inf_{\hat{V}} \sup_{V \in \mathcal{V}} \mathbb{E} \|\hat{V} - V\|^2.$$

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

$$\inf_{\hat{X}} \sup_{X^* \in \mathcal{B}} \mathbb{E} \|\hat{X} - X^*\|_F^2 = \inf_{\hat{V}} \sup_{V \in \mathcal{V}} \mathbb{E} \|\hat{V} - V\|^2 \geq c_2 \kappa^2 d = c_2 \kappa^2 (r_1 + r'_1) r_2 r_3.$$

Step 1.2. To show (9), there are two cases. **(a)** Suppose $\xi^2 \geq r_2 r_3 \kappa^2$, then $r'_1 = \lfloor \xi^2 / (r_2 r_3 \kappa^2) \rfloor \geq 1$ and therefore

$$\inf_{\hat{X}} \sup_{X^* \in \mathcal{T}_{(r_1, r_2, r_3)}^\xi} \mathbb{E} \|\hat{X} - X^*\|_F^2 \geq c_2 \kappa^2 (r_1 + r'_1) r_2 r_3 \geq c_2 r_1 r_2 r_3 + c_3 \xi^2.$$

(b) Suppose $\xi^2 < r_2 r_3 \kappa^2$. Then $r'_1 = 0$

$$\inf_{\hat{X}} \sup_{X^* \in \mathcal{T}_{(r_1, r_2, r_3)}^\xi} \mathbb{E} \|\hat{X} - X^*\|_F^2 \geq c_2 \kappa^2 r_1 r_2 r_3 \geq \frac{c_2}{2} (\kappa^2 r_1 r_2 r_3 + \xi^2),$$

where the first inequality follows from **Step 1.1**, the second inequality follows from $\xi^2 < r_2 r_3 \kappa^2$.

Step 2. Note that

$$\inf_{\hat{X}} \sup_{X^* \in \mathcal{T}_{(r_1, r_2, r_3)}^\xi} \mathbb{E} \|\hat{X} - X^*\|_F^2 \geq \inf_{\hat{X}} \sup_{X^* \in \mathcal{T}_{(r_1, r_2, r_3)}} \mathbb{E} \|\hat{X} - X^*\|_F^2 \geq c_3 \kappa^2 \sum_{k=1}^3 p_k r_k, \quad (12)$$

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

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

$$\inf_{\hat{X}} \sup_{X^* \in \mathcal{M}_r^\xi} \mathbb{E} \|\hat{X} - X^*\|_F^2 \geq c_1 \left\{ \kappa^2 n r + \xi^2 \right\}. \quad (13)$$

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

$$\mathcal{B} \subset \mathcal{M}_r^\xi. \quad (14)$$

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, $\text{rank}(B') \leq r$ and

$$\|B - B'\|_F^2 \leq r' n \|B\|_\infty^2 = r' n \kappa^2 \leq \xi^2.$$

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

$$\inf_{\widehat{X}} \sup_{X^* \in \mathcal{M}_r^\xi} \mathbb{E} \|\widehat{X} - X^*\|_F^2 \geq \inf_{\widehat{X}} \sup_{X^* \in \mathcal{B}} \mathbb{E} \|\widehat{X} - X^*\|_F^2. \quad (15)$$

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

$$\inf_{\widehat{X}} \sup_{X^* \in \mathcal{B}} \mathbb{E} \|\widehat{X} - X^*\|_F^2 = \inf_{\widehat{V}} \sup_{V \in \mathcal{V}} \mathbb{E} \|\widehat{V} - V\|^2.$$

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

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

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

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

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

$$\inf_{\widehat{X}} \sup_{X^* \in \mathcal{M}_r^\xi} \mathbb{E} \|\widehat{X} - X^*\|_F^2 \geq c_2 \kappa^2 r n \geq \frac{c_2}{2} (\kappa^2 r n + \xi^2), \quad (16)$$

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

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). \square

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:

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

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

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

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

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

Step 1. For the term \mathbf{I}_2 , observe that

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

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

Step 2. For the term \mathbf{I}_1 , we have that

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

where the last inequality follows from the observation that

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

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

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

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

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

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

$$\begin{aligned} & \left\| X^* \times_1 (I_{p_1} - \mathcal{P}_{U_1^{(1)}}) \times_2 \mathcal{P}_{U_2^{(0)}} \times_3 \mathcal{P}_{U_3^{(0)}} \right\|_{\mathbb{F}} \\ &= \left\| (I_{p_1} - \mathcal{P}_{U_1^{(1)}}) \cdot \mathcal{M}_1(X^*) \cdot (\mathcal{P}_{U_2^{(0)}} \otimes \mathcal{P}_{U_3^{(0)}}) \right\|_{\mathbb{F}} \\ &= \left\| (I_{p_1} - \mathcal{P}_{U_1^{(1)}}) \cdot \mathcal{M}_1(X^*) \cdot (U_2^{(0)} \otimes U_3^{(0)})(U_2^{(0)} \otimes U_3^{(0)})^\top \right\|_{\mathbb{F}} \\ &= \left\| (I_{p_1} - \mathcal{P}_{U_1^{(1)}}) \cdot \mathcal{M}_1(X^*) \cdot (U_2^{(0)} \otimes U_3^{(0)}) \right\|_{\mathbb{F}} \\ &\geq \left\| (I_{p_1} - \mathcal{P}_{U_1^{(1)}}) \cdot \mathcal{M}_1(X^*) \cdot \mathcal{P}_{U_2^* \otimes U_3^*} \cdot (U_2^{(0)} \otimes U_3^{(0)}) \right\|_{\mathbb{F}} \\ &\quad - \left\| (I_{p_1} - \mathcal{P}_{U_1^{(1)}}) \cdot \mathcal{M}_1(X^*) \cdot (I_{p_2 p_3} - \mathcal{P}_{U_2^* \otimes U_3^*}) \cdot (U_2^{(0)} \otimes U_3^{(0)}) \right\|_{\mathbb{F}} \\ &= \text{II}_1 - \text{II}_2, \end{aligned}$$

where the first inequality follows from that $(U_2^{(0)} \otimes U_3^{(0)}) \in \mathbb{O}^{(p_2 p_3) \times (r_2 r_3)}$, and [Lemma 22](#). Before analyzing the terms II_1 and II_2 , we firstly note that

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

which hold due to the properties of the Kronecker product and [Lemma 25](#).

For the term II_1 , we have that

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

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

$$\begin{aligned} \text{II}_2 &= \left\| (I_{p_1} - \mathcal{P}_{U_1^{(1)}}) \cdot \mathcal{M}_1(X^*) \cdot (I_{p_2 p_3} - \mathcal{P}_{U_2^* \otimes U_3^*}) \cdot (U_2^{(0)} \otimes U_3^{(0)}) \right\|_{\mathbb{F}} \\ &= \left\| (I_{p_1} - \mathcal{P}_{U_1^{(1)}}) \cdot \mathcal{M}_1(X^*) \cdot (U_2^* \otimes U_3^*)_{\perp} \cdot (U_2^* \otimes U_3^*)_{\perp}^\top \cdot (U_2^{(0)} \otimes U_3^{(0)}) \right\|_{\mathbb{F}} \\ &\leq \left\| I_{p_1} - \mathcal{P}_{U_1^{(1)}} \right\| \left\| \mathcal{M}_1(X^*) \cdot (U_2^* \otimes U_3^*)_{\perp} \right\|_{\mathbb{F}} \left\| (U_2^* \otimes U_3^*)_{\perp}^\top \cdot (U_2^{(0)} \otimes U_3^{(0)}) \right\|_{\mathbb{F}} \\ &= \left\| \mathcal{M}_1(X^*) \cdot (U_2^* \otimes U_3^*)_{\perp} \right\|_{\mathbb{F}}. \end{aligned}$$

Note that

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

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

where the last inequality follows from [Lemma 29](#). Therefore,

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

Combining II_1 and II_2 , we have

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

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

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

It suffices to apply [Lemma 23](#) to bound

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

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

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

Here

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

Note that

$$\begin{aligned}
&\left\| \mathcal{M}_1(Z) \cdot U_2^* \otimes U_3^{(0)} \right\| \|U_2^{*\top} U_2^{(0)}\| = \left\| \mathcal{M}_1(Z) \cdot U_2^* \otimes (U_3^* U_3^{*\top} + U_{3\perp}^* U_{3\perp}^{*\top}) U_3^{(0)} \right\| \\
&\leq \left\| \mathcal{M}_1(Z) \cdot U_2^* \otimes (U_3^* U_3^{*\top} U_3^{(0)}) \right\| + \left\| \mathcal{M}_1(Z) \cdot U_2^* \otimes (U_{3\perp}^* U_{3\perp}^{*\top}) U_3^{(0)} \right\| \\
&\leq \left\| \mathcal{M}_1(Z) \cdot U_2^* \otimes U_3^* \right\| \|U_3^{*\top} U_3^{(0)}\| + \left\| \mathcal{M}_1(Z) \cdot U_2^* \otimes U_{3\perp}^* \right\| \|U_{3\perp}^{*\top} U_3^{(0)}\| \\
&\leq C_4 \kappa(\sqrt{p_1 + r_2 r_3}) + C_5 \kappa(\sqrt{p_1 + p_3 r_2}) r_{\max}^{-1/2}, \quad (23)
\end{aligned}$$

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

$$\begin{aligned}
&\left\| \mathcal{M}_1(Z) \cdot U_{2\perp}^* \otimes U_3^{(0)} \right\| \|U_{2\perp}^{*\top} U_2^{(0)}\| \leq \left\| \mathcal{M}_1(Z) \cdot U_{2\perp}^* \otimes (U_3^* U_3^{*\top} + U_{3\perp}^* U_{3\perp}^{*\top}) U_3^{(0)} \right\| \|U_{2\perp}^{*\top} U_2^{(0)}\| \\
&\leq \left\| \mathcal{M}_1(Z) \cdot U_{2\perp}^* \otimes (U_3^* U_3^{*\top} U_3^{(0)}) \right\| \|U_{2\perp}^{*\top} U_2^{(0)}\| + \left\| \mathcal{M}_1(Z) \cdot U_{2\perp}^* \otimes (U_{3\perp}^* U_{3\perp}^{*\top} U_3^{(0)}) \right\| \|U_{2\perp}^{*\top} U_2^{(0)}\| \\
&\leq \left\| \mathcal{M}_1(Z) \cdot U_{2\perp}^* \otimes U_3^* \right\| \|U_3^{*\top} U_3^{(0)}\| \|U_{2\perp}^{*\top} U_2^{(0)}\| + \left\| \mathcal{M}_1(Z) \cdot U_{2\perp}^* \otimes U_{3\perp}^* \right\| \|U_{3\perp}^{*\top} U_3^{(0)}\| \|U_{2\perp}^{*\top} U_2^{(0)}\| \\
&\leq C \kappa(\sqrt{p_1 + p_2 r_3}) r_{\max}^{-1/2} + C \kappa(\sqrt{p_1 + p_3 r_2}) r_{\max}^{-1}. \quad (24)
\end{aligned}$$

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

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

In addition,

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

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 ,

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

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

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

Step 5. The conclusions of [Step 3](#) and [Step 4](#) lead to

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

This bound together with [Step 1](#) and [Step 2](#) leads to

$$\|\tilde{X} - X^*\|_F \leq C_7 \kappa (\sqrt{p_1 r_1} + \sqrt{p_2 r_2} + \sqrt{p_3 r_3} + \sqrt{r_1 r_2 r_3}) + C_7 \xi_{(r_1, r_2, r_3)}.$$

□

D PROOF OF THEOREM 3

Proof of Theorem 3. By the triangle inequality,

$$\|Y_{(r)} - X^*\|_F \leq \|X_{(r)}^* - X^*\|_F + \|Y_{(r)} - X_{(r)}^*\|_F = \xi_{(r)} + \|Y_{(r)} - X_{(r)}^*\|_F. \quad (27)$$

Since $\text{rank}(Y_{(r)} - X_{(r)}^*) \leq 2r$, we have

$$\|Y_{(r)} - X_{(r)}^*\|_F^2 = \sum_{k=1}^{2r} \sigma_k^2(Y_{(r)} - X_{(r)}^*). \quad (28)$$

Write

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

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

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

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

$$\|Y_{(r)} - X_{(r)}^*\|_F \leq \sqrt{\sum_{i=1}^{2r} \sigma_{r+i}^2(X^*)} + \sqrt{\sum_{i=1}^{2r} \sigma_i^2(Z)} + \sqrt{\sum_{i=1}^{2r} \sigma_{r+i}^2(Y)}. \quad (29)$$

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

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

Substituting this into (29), we get

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

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 $\|Y_{(r)} - X_{(r)}^*\|_{\text{F}}$ into (27), we obtain

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

□

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

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

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

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

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

$$\mathbb{P}\left(\|\mathcal{M}_1(Z)(W_2 \otimes W_3)\| \geq C_1 \kappa (\sqrt{p_1 + s_2 s_3})\right) \leq C_2 \exp(-cp_1),$$

Proof. It suffices to observe that $W_2 \otimes W_3 \in \mathbb{O}^{p_2 p_3 \times r_2 r_3}$. The desired result is a direct consequence of Lemma 13. \square

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

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

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

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,

$$\mathcal{M}(Z) \in \mathbb{R}^{p_1 \times p_2 p_3}, (\mathcal{M}(Z))_{i,j} \stackrel{i.i.d}{\sim} \text{subGaussian}(0, \kappa^2), \quad \text{and} \quad \|V_2 \otimes V_3\| = \|V_2\| \|V_3\| = 1.$$

It follows from Lemma 13 that

$$\mathbb{P}\left(\|\mathcal{M}_1(Z) \cdot (V_2 \otimes V_3)\| > x\right) \leq 2 \exp(C(p_1 + r_2 r_3) - cx^2 \kappa^{-2}).$$

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 $\mathcal{N}_{p_3, r_3}(\epsilon)$ 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

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

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

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

Then

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

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

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

1188 Similarly

$$1189 \quad \|\mathcal{M}_1(Z)V_2 \otimes (V_3 - \tilde{V}_3)\| \leq \frac{\psi}{4}.$$

1191 In addition,

$$1192 \quad \|\mathcal{M}_1(Z)\tilde{V}_2 \otimes \tilde{V}_3\| \leq \sup_{V_2 \in \mathcal{N}_{p_2, r_2}(1/4), b \in \mathcal{N}_{p_3, r_3}(1/4)} \|\mathcal{M}_1(Z)V_2 \otimes V_3\|.$$

1195 So for any V_2 and V_3 ,

$$1196 \quad \|\mathcal{M}_1(Z)V_2 \otimes V_3\| \leq \frac{1}{2}\psi + \sup_{V_2 \in \mathcal{N}_{p_2, r_2}(1/4), b \in \mathcal{N}_{p_3, r_3}(1/4)} \|\mathcal{M}_1(Z)V_2 \otimes V_3\|.$$

1199 Taking sup over all $V_2 \in \{A \in \mathbb{R}^{p_2 \times r_2} : \|A\| \leq 1\}$ and $V_3 \in \{B \in \mathbb{R}^{p_3 \times r_3} : \|B\| \leq 1\}$, it follows that

$$1201 \quad \psi \leq \frac{1}{2}\psi + \sup_{V_2 \in \mathcal{N}_{p_2, r_2}(1/4), b \in \mathcal{N}_{p_3, r_3}(1/4)} \|\mathcal{M}_1(Z)V_2 \otimes V_3\|,$$

1203 or simply

$$1204 \quad \psi \leq 2 \sup_{V_2 \in \mathcal{N}_{p_2, r_2}(1/4), b \in \mathcal{N}_{p_3, r_3}(1/4)} \|\mathcal{M}_1(Z)V_2 \otimes V_3\|.$$

1206 **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

$$1209 \quad \mathbb{P}(\psi \geq 2t) \leq \mathbb{P}\left(\sup_{V_2 \in \mathcal{N}_{p_2, r_2}(1/4), b \in \mathcal{N}_{p_3, r_3}(1/4)} \|\mathcal{M}_1(Z)V_2 \otimes V_3\| \geq t\right)$$

$$1211 \quad \leq C_2^{p_2 r_2} C_3^{p_3 r_3} \sup_{V_2 \in \mathcal{N}_{p_2, r_2}(1/4), b \in \mathcal{N}_{p_3, r_3}(1/4)} \mathbb{P}\left(\|\mathcal{M}_1(Z)V_2 \otimes V_3\| \geq t\right)$$

$$1213 \quad \leq 2 \exp(C(p_1 + r_2 r_3 + p_2 r_2 + p_3 r_3) - ct^2 \kappa^{-2}).$$

1215 Here C and C_3 are positive constants. The desired result follows by noting

$$1216 \quad \psi = \sup_{\substack{V_2 \in \mathbb{R}^{p_2 \times r_2}, \|V_2\| \leq 1 \\ V_3 \in \mathbb{R}^{p_3 \times r_3}, \|V_3\| \leq 1}} \|\mathcal{M}_1(Z)(V_2 \otimes V_3)\|.$$

1220 □

1221 **Lemma 13.** Suppose $Z \in \mathbb{R}^{n \times m}$, with $Z_{ij} \stackrel{i.i.d.}{\sim} \text{subGaussian}(0, \kappa^2)$. Let $A \in \mathbb{R}^{p \times n}$ and $B \in \mathbb{R}^{m \times q}$ be non-random matrices. Then for any $t > 0$

$$1224 \quad \mathbb{P}(\|AZB\| > t) \leq C_1 \exp\left(C(p+q) - \frac{ct^2}{\kappa^2 \|A\|^2 \|B\|^2}\right). \quad (30)$$

1227 *Proof.*

1228 **Step 1.** Let $u \in \mathbb{R}^n$ and $v \in \mathbb{R}^m$ be non-random. Then that $u^\top Z v = \sum_{i=1}^n \sum_{j=1}^m u_i Z_{ij} v_j$. 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 $\kappa^2 \|u\|_2 \|v\|_2$. Consequently by Hoeffding's inequality,

$$1232 \quad \mathbb{P}(|u^\top Z v| > t) \leq 2 \exp\left(-\frac{ct^2}{\kappa^2 \|u\|_2^2 \|v\|_2^2}\right).$$

1235 **Step 2.** Let $a \in \mathbb{R}^p$ and $b \in \mathbb{R}^q$ be non-random vectors such that $\|a\|_2, \|b\|_2 \leq 1$. By **Step 1**, it follows that

$$1237 \quad \mathbb{P}(|a^\top AZB| > t) \leq 2 \exp\left(-\frac{ct^2}{\kappa^2 \|Aa\|_2^2 \|Bb\|_2^2}\right) \leq 2 \exp\left(-\frac{ct^2}{\kappa^2 \|A\|^2 \|B\|^2}\right).$$

1240 **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 $\tilde{a} \in \mathcal{N}_p(\epsilon)$ such that

$$1241 \quad \|a - \tilde{a}\|_2 \leq \epsilon.$$

1242 Similarly let $\mathcal{N}_q(\epsilon)$ be an ϵ -net of the unit ball in \mathbb{R}^q .

1243

1244 Denote the random quantity

1245

$$1246 \quad \psi = \|AZB\| = \sup_{a \in \mathbb{R}^p, b \in \mathbb{R}^q, \|a\|_2 = \|b\|_2 = 1} |a^\top AZBb|.$$

1247

1248 For any given $a \in \mathbb{R}^p$ and $b \in \mathbb{R}^q$, let $\tilde{a} \in \mathcal{N}_p(1/4)$ and $\tilde{b} \in \mathcal{N}_q(1/4)$ be such that

1249

$$1250 \quad \|a - \tilde{a}\|_2 \leq 1/4 \quad \text{and} \quad \|b - \tilde{b}\|_2 \leq 1/4.$$

1251

1252 Then

$$1253 \quad |a^\top AZBb| \leq |(a - \tilde{a})^\top AZBb| + |\tilde{a}^\top AZB(\tilde{b} - b)| + |\tilde{a}^\top AZB\tilde{b}|.$$

1254

1255 Note that $\|a - \tilde{a}\| \leq \frac{1}{4}$. So

1256

$$1257 \quad |(a - \tilde{a})^\top AZBb| = \frac{1}{4} |4(a - \tilde{a})^\top AZBb| \leq \frac{\psi}{4}.$$

1258

1259 Similarly

1260

$$1261 \quad |\tilde{a}^\top AZB(\tilde{b} - b)| \leq \frac{\psi}{4}.$$

1262

1263 In addition,

$$1264 \quad |\tilde{a}^\top AZB\tilde{b}| \leq \sup_{a \in \mathcal{N}_p(1/4), b \in \mathcal{N}_q(1/4)} |a^\top AZBb|.$$

1265

1266 So for any a and b ,

1267

$$1268 \quad |a^\top AZBb| \leq \frac{1}{2}\psi + \sup_{a \in \mathcal{N}_p(1/4), b \in \mathcal{N}_q(1/4)} |a^\top AZBb|.$$

1269

1270 Taking sup over all unit vectors $a \in \mathbb{R}^p$ and $b \in \mathbb{R}^q$, it follows that

1271

$$1272 \quad \psi \leq \frac{1}{2}\psi + \sup_{a \in \mathcal{N}_p(1/4), b \in \mathcal{N}_q(1/4)} |a^\top AZBb|,$$

1273

1274 or simply

1275

$$1276 \quad \psi \leq 2 \sup_{a \in \mathcal{N}_p(1/4), b \in \mathcal{N}_q(1/4)} |a^\top AZBb|.$$

1277

1278 **Step 4.** By Vershynin (2018), the cardinality of $\mathcal{N}_p(\epsilon)$ is bounded by $(\frac{C}{\epsilon})^p$, and the cardinality of

1279

1280 $\mathcal{N}_q(\epsilon)$ is bounded by $(\frac{C}{\epsilon})^q$. Therefore

1281

$$1282 \quad \mathbb{P}(\psi \geq 2t) = \mathbb{P}\left(\sup_{a \in \mathcal{N}_p(1/4), b \in \mathcal{N}_q(1/4)} |a^\top AZBb| \geq t\right) \leq C_2^p C_2^q \sup_{a \in \mathcal{N}_p(1/4), b \in \mathcal{N}_q(1/4)} \mathbb{P}(|a^\top AZBb| \geq t)$$

$$1283 \quad \leq 2C_2^p C_2^q \exp\left(-\frac{ct^2}{\kappa^2 \|A\|^2 \|B\|^2}\right),$$

1284

1285

1286

1287 where c, C are positive constants. The desired result follows by noting $\psi = \|AZB\|$. \square

1288

1289 **Lemma 14.** Suppose $Z \in \mathbb{R}^{n \times m}$, with $Z_{ij} \stackrel{i.i.d}{\sim}$ subGaussian($0, \kappa^2$). Let $A \in \mathbb{R}^{m \times p}$ and $B \in$

1290

1291 $\mathbb{R}^{m \times q}$ be non-random matrices. Then for any $t > 0$

1292

$$1293 \quad \mathbb{P}\left(\|A^\top Z^\top ZB - n\kappa^2 A^\top B\| > t\right) \leq C_1 \exp\left(C(p+q) - \min\left(\frac{t^2}{n\kappa^4 \|B\|^2 \|A\|^2}, \frac{t}{\kappa^2 \|B\| \|A\|}\right)\right),$$

1294

1295

where C and C_1 are positive constants.

1296 *Proof.* For any non-random $u, v \in \mathbb{R}^m$, it follows that

$$1297 \quad u^\top Z^\top Z v - n\kappa^2 u^\top v = \sum_{j=1}^n (u^\top Z_j)(v^\top Z_j) - \mathbb{E}\{(u^\top Z_j)(v^\top Z_j)\},$$

1300 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
 1301 $(v^\top Z_j)$ is sub-Gaussian with parameter $\kappa^2 \|v\|_2^2$. Since Z have i.i.d. entries, it follows that
 1302 $\{(u^\top Z_j)(v^\top Z_j)\}_{j=1}^n$ are i.i.d. sub-exponential with parameter $\kappa^4 \|u\|_2^2 \|v\|_2^2$. So

$$1303 \quad \mathbb{P}\left(|u^\top Z^\top Z v - n\kappa^2 u^\top v| \geq t\right) \leq 2 \exp\left(-c \min\left\{\frac{t^2}{n\kappa^4 \|u\|_2^2 \|v\|_2^2}, \frac{t}{\kappa^2 \|u\|_2 \|v\|_2}\right\}\right).$$

1304 **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 $\|a\|_2 = 1$,
 1305 there exists $\tilde{a} \in \mathcal{N}_p(\epsilon)$ such that

$$1306 \quad \|a - \tilde{a}\|_2 \leq \epsilon.$$

1307 Similarly let $\mathcal{N}_q(\epsilon)$ be the ϵ -net of the unit ball in \mathbb{R}^q .

1308 Denote the random quantity

$$1309 \quad \psi = \|A^\top Z^\top Z B - n\kappa^2 A^\top B\| = \sup_{a \in \mathbb{R}^p, b \in \mathbb{R}^q, \|a\|_2 = \|b\|_2 = 1} |a^\top (A^\top Z^\top Z B - n\kappa^2 A^\top B)b|.$$

1310 For any given $a \in \mathbb{R}^p$ and $b \in \mathbb{R}^q$, let $\tilde{a} \in \mathcal{N}_p(1/4)$ and $\tilde{b} \in \mathcal{N}_q(1/4)$ be such that

$$1311 \quad \|a - \tilde{a}\|_2 \leq 1/4 \quad \text{and} \quad \|b - \tilde{b}\|_2 \leq 1/4.$$

1312 Then

$$1313 \quad |a^\top (A^\top Z^\top Z B - n\kappa^2 A^\top B)b| \leq |(a - \tilde{a})^\top (A^\top Z^\top Z B - n\kappa^2 A^\top B)b|$$

$$1314 \quad + |\tilde{a}^\top (A^\top Z^\top Z B - n\kappa^2 A^\top B)(\tilde{b} - b)| + |\tilde{a}^\top (A^\top Z^\top Z B - n\kappa^2 A^\top B)\tilde{b}|.$$

1315 Note that $\|a - \tilde{a}\|_2 \leq \frac{1}{4}$. So

$$1316 \quad |(a - \tilde{a})^\top (A^\top Z^\top Z B - n\kappa^2 A^\top B)b| = \frac{1}{4} |\{4(a - \tilde{a})^\top\} (A^\top Z^\top Z B - n\kappa^2 A^\top B)b| \leq \frac{\psi}{4}.$$

1317 Similarly

$$1318 \quad |\tilde{a}^\top (A^\top Z^\top Z B - n\kappa^2 A^\top B)(\tilde{b} - b)| \leq \frac{\psi}{4}.$$

1319 In addition,

$$1320 \quad |\tilde{a}^\top (A^\top Z^\top Z B - n\kappa^2 A^\top B)\tilde{b}| \leq \sup_{a \in \mathcal{N}_p(1/4), b \in \mathcal{N}_q(1/4)} |a^\top (A^\top Z^\top Z B - n\kappa^2 A^\top B)b|.$$

1321 So for any a and b ,

$$1322 \quad |a^\top (A^\top Z^\top Z B - n\kappa^2 A^\top B)b| \leq \frac{1}{2}\psi + \sup_{a \in \mathcal{N}_p(1/4), b \in \mathcal{N}_q(1/4)} |a^\top (A^\top Z^\top Z B - n\kappa^2 A^\top B)b|.$$

1323 Taking sup over all unit vectors $a \in \mathbb{R}^p$ and $b \in \mathbb{R}^q$, it follows that

$$1324 \quad \psi \leq \frac{1}{2}\psi + \sup_{a \in \mathcal{N}_p(1/4), b \in \mathcal{N}_q(1/4)} |a^\top (A^\top Z^\top Z B - n\kappa^2 A^\top B)b|,$$

1325 or simply

$$1326 \quad \psi \leq 2 \sup_{a \in \mathcal{N}_p(1/4), b \in \mathcal{N}_q(1/4)} |a^\top (A^\top Z^\top Z B - n\kappa^2 A^\top B)b|.$$

1327 **Step 2.** By Vershynin (2018), the cardinality of $\mathcal{N}_p(\epsilon)$ is bounded by $(\frac{C}{\epsilon})^p$, and the cardinality of
 1328 $\mathcal{N}_q(\epsilon)$ is bounded by $(\frac{C}{\epsilon})^q$, for a positive constant C . Therefore

$$1329 \quad \mathbb{P}(\psi \geq 2t) = \mathbb{P}\left(\sup_{a \in \mathcal{N}_p(1/4), b \in \mathcal{N}_q(1/4)} |a^\top (A^\top Z^\top Z B - n\kappa^2 A^\top B)b| \geq t\right)$$

$$\begin{aligned}
1350 & \leq C_2^p C_2^q \sup_{a \in \mathcal{N}_p(1/4), b \in \mathcal{N}_q(1/4)} \mathbb{P}(|a^\top (A^\top Z^\top Z B - n\kappa^2 A^\top B)b| \geq t) \\
1351 & = C_2^p C_2^q \sup_{a \in \mathcal{N}_p(1/4), b \in \mathcal{N}_q(1/4)} \mathbb{P}(|(Aa)^\top Z^\top Z (Bb) - n\kappa^2 (Aa)^\top (Bb)| \geq t) \\
1352 & \leq 2C_2^p C_2^q \exp\left(-c \min\left\{\frac{t^2}{n\kappa^4 \|Aa\|_2^2 \|Bb\|_2^2}, \frac{t}{\kappa^2 \|Aa\|_2 \|Bb\|_2}\right\}\right), \\
1353 & \\
1354 & \\
1355 & \\
1356 &
\end{aligned}$$

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

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

$$\mathbb{P}(\|Z\| > t) \leq C_1 \exp\left(C(m+n) - \frac{ct^2}{\kappa^2}\right),$$

where c, C and C_1 are positive constants.

Proof. By assumption,

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

The rest of the proof is similar and simpler than [Lemma 13](#) and is omitted. \square

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

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

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

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

Then it holds that

$$Y V_r M = (X_r + X_\perp + Z) V_r M = (X_r + Z) V_r M,$$

$$1404 \quad M^\top \mathbb{E} [V_r^\top Y^\top Y V_r] M = M^\top \mathbb{E} [V_r^\top (X_r + Z)^\top (X_r + Z) V_r] M = I_r.$$

1405
1406 **Step 1.** For $\sigma_r(YV_r)$, observe that

$$1407 \quad \sigma_r^2(YV_r) = \sigma_r^2(\{X_r + Z\}V_r) = \sigma_r^2(\{X_r + Z\}V_r M M^{-1}) \geq \sigma_r^2(\{X_r + Z\}V_r M) \sigma_{\min}^2(M^{-1})$$

$$1408 \quad = \sigma_r^2(\{X_r + Z\}V_r M) \{\sigma_r^2(X) + n\kappa^2\}$$

$$1409 \quad = \sigma_r(M^\top V_r^\top \{X_r + Z\}^\top \{X_r + Z\} V_r M) \{\sigma_r^2(X) + n\kappa^2\}$$

$$1410 \quad (32)$$

1411 where the inequality follows from [Lemma 20](#).

1412 Consider the term $\sigma_r(M^\top V_r^\top \{X_r + Z\}^\top \{X_r + Z\} V_r M)$. Note that

$$1413 \quad M^\top V_r^\top (X_r + Z)^\top (X_r + Z) V_r M - I_r$$

$$1414 \quad = M^\top V_r^\top (X_r + Z)^\top (X_r + Z) V_r M - \mathbb{E} [M^\top V_r^\top (X_r + Z)^\top (X_r + Z) V_r M]$$

$$1415 \quad = \underbrace{M^\top V_r^\top X_r^\top X_r V_r M - \mathbb{E} [M^\top V_r^\top X_r^\top X_r V_r M]}_{=0} + M^\top V_r^\top X_r^\top Z V_r M$$

$$1416 \quad + \underbrace{M^\top V_r^\top Z^\top X_r V_r M}_{=0} + \underbrace{M^\top V_r^\top Z^\top Z V_r M - \mathbb{E} [M^\top V_r^\top Z^\top Z V_r M]}_{=0}$$

$$1417 \quad + \underbrace{\mathbb{E} [M^\top V_r^\top X_r^\top Z V_r M]}_{=0} + \underbrace{\mathbb{E} [M^\top V_r^\top Z^\top X_r V_r M]}_{=0}$$

$$1418 \quad = M^\top V_r^\top X_r^\top Z V_r M + M^\top V_r^\top Z^\top X_r V_r M + M^\top V_r^\top (Z^\top Z - n\kappa^2 I_m) V_r M. \quad (33)$$

1419 Since,

$$1420 \quad \|X_r V_r M\|^2 = \max_{k=1, \dots, r} \frac{\sigma_k^2(X)}{\sigma_k^2(X) + n\kappa^2} \leq 1, \quad \text{and} \quad \|V_r M\|^2 = \|M\|^2 = \frac{1}{\sigma_r^2(X) + n\kappa^2},$$

1421 it follows by [Lemma 13](#) that

$$1422 \quad \mathbb{P}\left(\|M^\top V_r^\top X_r^\top Z V_r M\| \geq x\right) \leq 2 \exp\left(Cr - cx^2 \frac{\sigma_r^2(X) + n\kappa^2}{\kappa^2}\right). \quad (34)$$

1423 Similarly, [Lemma 14](#) implies that

$$1424 \quad \mathbb{P}\left(\|M^\top V_r^\top (Z^\top Z - n\kappa^2 I_m) V_r M\| \geq x\right)$$

$$1425 \quad \leq 2 \exp\left(Cr - c \min\left\{x^2 \frac{(\sigma_r^2(X) + n\kappa^2)^2}{n\kappa^4}, x \frac{\sigma_r^2(X) + n\kappa^2}{\kappa^2}\right\}\right), \quad (35)$$

$$1426 \quad \leq 2 \exp\left(Cr - c \frac{\sigma_r^2(X) + n\kappa^2}{\kappa^2} \min\{x^2, x\}\right),$$

1427 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

$$1428 \quad \mathbb{P}\left(\|M^\top V_r^\top (X_r + Z)^\top (X_r + Z) V_r M - I_r\| \geq x\right) \leq 6 \exp\left(Cr - c \frac{\sigma_r^2(X) + n\kappa^2}{\kappa^2} \min\{x^2, x\}\right).$$

$$1429 \quad (36)$$

1430 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

$$1431 \quad \mathbb{P}\left(\sigma_r(M^\top V_r^\top \{X_r + Z\}^\top \{X_r + Z\} V_r M) \geq 1 - x\right)$$

$$1432 \quad \geq 1 - 6 \exp\left(Cr - c \frac{\sigma_r^2(X) + n\kappa^2}{\kappa^2} \min\{x^2, x\}\right), \quad (37)$$

(37) and (32) together imply that

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

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

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

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

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

Moreover,

$$\begin{aligned} \sigma_{\max}^2(YV_\perp) &= \|V_\perp^\top Y^\top Y V_\perp\| = \|V_\perp^\top (X_r + X_\perp + Z)^\top (X_r + X_\perp + Z) V_\perp\| \\ &= \|V_\perp^\top (X_\perp + Z)^\top (X_\perp + Z) V_\perp\| \\ &\leq \|V_\perp^\top Z^\top Z V_\perp\| + \|V_\perp^\top Z^\top X_\perp V_\perp\| + \|V_\perp^\top X_\perp^\top Z V_\perp\| + \|V_\perp^\top X_\perp^\top X_\perp V_\perp\| \\ &= \underbrace{\|V_\perp^\top Z^\top Z V_\perp\|}_{I_1} + 2 \underbrace{\|V_\perp^\top Z^\top X_\perp V_\perp\|}_{I_2} + \underbrace{\|V_\perp^\top X_\perp^\top X_\perp V_\perp\|}_{I_3}. \end{aligned} \quad (39)$$

For the term I_3 , we have

$$I_3 = \sigma_1^2(X_\perp V_\perp) = \sigma_1^2(U_\perp \Sigma_\perp V_\perp^\top V_\perp) = \sigma_1^2(\Sigma_\perp) = \sigma_{r+1}^2(X). \quad (40)$$

For I_2 , note that

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

it follows from [Lemma 13](#) that

$$\mathbb{P}(I_2 \geq x) \leq 2 \exp\left(C_1 m - \frac{c_1 x^2}{\kappa^2 \sigma_{r+1}^2(X)}\right). \quad (41)$$

For I_1 , note that by [Lemma 14](#) and the fact that $V_\perp^\top V_\perp = I_{m-r}$, we have

$$\begin{aligned} & \mathbb{P}\left(\|V_\perp^\top Z^\top Z V_\perp - n\kappa^2 I_{m-r}\| \geq t\right) \leq 2 \exp\left(C_2 m - c_2 \min\left\{\frac{t^2}{n\kappa^4}, \frac{t}{\kappa^2}\right\}\right), \\ & \implies \mathbb{P}(I_1 \geq n\kappa^2(1+t)) \leq 2 \exp(C_2 m - c_2 n \min\{t^2, t\}). \end{aligned} \quad (42)$$

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

$$\begin{aligned} & \mathbb{P}\left(\sigma_{r+1}^2(Y) \geq n\kappa^2 + \frac{\sigma_r^2(X) - \sigma_{r+1}^2(X)}{6} + \kappa \sigma_{r+1}(X) \sqrt{\frac{2C_1}{c_1} m} + \sigma_{r+1}^2(X)\right) \\ & \leq 2 \exp\left(C_2 m - c_2 \min\left\{\frac{(\sigma_r^2(X) - \sigma_{r+1}^2(X))^2}{36\kappa^4 n}, \frac{\sigma_r^2(X) - \sigma_{r+1}^2(X)}{6\kappa^2}\right\}\right) + 2 \exp(-C_3 m). \end{aligned} \quad (43)$$

Step 3. Recall we define

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

We have

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

where the first equality follows from the fact that YV_r and $YV_r M$ 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 $YV_r M$ are in nonincreasing order so that its smallest singular value equals $\sigma_r(YV_r M)$.

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

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

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

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

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

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

Since,

$$\|X_r V_r M\|^2 \leq 1, \|V_r M\|^2 = \frac{1}{\sigma_r^2(X) + n\kappa^2} \text{ and } \|X_\perp V_\perp\|^2 = \|X_\perp\|^2 = \sigma_{r+1}^2(X),$$

it follows from Lemma 13 that

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

Similarly, Lemma 14 implies that

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

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

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

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

Step 4. Define the event

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

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

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

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

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

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

In what follows, we show that under the SNR assumption

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

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

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

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

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

We have

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

So

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

In addition,

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

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

Step 5. Under the event \mathcal{E} , by [Lemma 24](#), we have that

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

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,

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

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

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

where the third inequality follows from the observation that

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

and last display follows from

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

□

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 $1 - C_1 \exp(-C_2 p_k)$,*

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

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

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

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

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

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

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

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

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

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 $\text{rank}(A) = r$. Therefore $A^\top A$ is invertible. Let the SVD of A satisfies $A = U_A \Sigma_A V_A^\top$, then

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

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

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

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

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

The desired result follows from (53). \square

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](#). \square

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

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

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

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

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

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

Then

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

□

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

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

Proof. Observe that

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

□

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

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

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

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

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

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

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

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

$$\begin{aligned}
& \left\| \mathcal{P}_{\hat{U}_\perp} B \right\|_{\mathbb{F}} \leq \left\| \mathcal{P}_{\hat{U}_\perp} B_{(R)} \right\|_{\mathbb{F}} + \left\| \mathcal{P}_{\hat{U}_\perp} (B - B_{(R)}) \right\|_{\mathbb{F}} = \sqrt{\sum_{j=1}^R \sigma_j^2(\mathcal{P}_{\hat{U}_\perp} B_{(R)})} + \left\| \mathcal{P}_{\hat{U}_\perp} (B - B_{(R)}) \right\|_{\mathbb{F}} \\
& \leq \sqrt{\sum_{j=1}^R \sigma_j^2(\mathcal{P}_{\hat{U}_\perp} B_{(R)})} + \|B - B_{(R)}\|_{\mathbb{F}} = \sqrt{\sum_{j=1}^R \sigma_j^2(\mathcal{P}_{\hat{U}_\perp} B_{(R)})} + \sqrt{\sum_{j=R+1}^n \sigma_j^2(B)} \\
& \leq \left\| (\sigma_1(\mathcal{P}_{\hat{U}_\perp} B_{(R)}) - \sigma_1(\mathcal{P}_{\hat{U}_\perp} B), \dots, \sigma_R(\mathcal{P}_{\hat{U}_\perp} B_{(R)}) - \sigma_R(\mathcal{P}_{\hat{U}_\perp} B))^\top \right\|_2 + \left\| (\sigma_1(\mathcal{P}_{\hat{U}_\perp} B), \dots, \sigma_R(\mathcal{P}_{\hat{U}_\perp} B))^\top \right\|_2 \\
& \quad + \sqrt{\sum_{j=R+1}^n \sigma_j^2(B)} \\
& \leq \left\| \Sigma(\mathcal{P}_{\hat{U}_\perp} B_{(R)}) - \Sigma(\mathcal{P}_{\hat{U}_\perp} B) \right\|_{\mathbb{F}} + \left\| (\sigma_1(\mathcal{P}_{\hat{U}_\perp} B), \dots, \sigma_R(\mathcal{P}_{\hat{U}_\perp} B))^\top \right\|_2 + \sqrt{\sum_{j=R+1}^n \sigma_j^2(B)} \\
& \leq \left\| \mathcal{P}_{\hat{U}_\perp} (B_{(R)} - B) \right\|_{\mathbb{F}} + \sqrt{\sum_{j=1}^R \sigma_j^2(\mathcal{P}_{\hat{U}_\perp} B)} + \sqrt{\sum_{j=R+1}^n \sigma_j^2(B)} \\
& \leq \sqrt{\sum_{j=1}^R \sigma_j^2(\mathcal{P}_{\hat{U}_\perp} B)} + 2\sqrt{\sum_{j=R+1}^n \sigma_j^2(B)},
\end{aligned}$$

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

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

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

$$\begin{aligned}
& \sqrt{\sum_{j=1}^R \sigma_j^2(\mathcal{P}_{\hat{U}_\perp} A)} = \sqrt{\sum_{j=R+1}^{2R} \sigma_j^2(A)} = \|(\sigma_{R+1}(A), \dots, \sigma_{2R}(A))^\top\| \\
& \leq \|(\sigma_{R+1}(A) - \sigma_{R+1}(B), \dots, \sigma_{2R}(A) - \sigma_{2R}(B))^\top\| + \|(\sigma_{R+1}(B), \dots, \sigma_{2R}(B))^\top\| \\
& \leq \min \left\{ \sqrt{R} \|Z\|, \|Z\|_{\mathbb{F}} \right\} + \sqrt{\sum_{j=R+1}^n \sigma_j^2(B)}, \tag{55}
\end{aligned}$$

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

$$\|(\sigma_{R+1}(A) - \sigma_{R+1}(B), \dots, \sigma_{2R}(A) - \sigma_{2R}(B))^\top\| \leq \|\Sigma(A) - \Sigma(B)\|_{\mathbb{F}} \leq \|Z\|_{\mathbb{F}},$$

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

$$\begin{aligned}
& \sqrt{\sum_{j=1}^R \sigma_j^2(\mathcal{P}_{\hat{U}_\perp} B)} = \left\| (\sigma_1(\mathcal{P}_{\hat{U}_\perp} (A - Z)), \dots, \sigma_R(\mathcal{P}_{\hat{U}_\perp} (A - Z)))^\top \right\| \\
& \leq \left\| (\sigma_1(\mathcal{P}_{\hat{U}_\perp} (A - Z)) - \sigma_1(\mathcal{P}_{\hat{U}_\perp} A), \dots, \sigma_R(\mathcal{P}_{\hat{U}_\perp} (A - Z)) - \sigma_R(\mathcal{P}_{\hat{U}_\perp} A))^\top \right\| \\
& \quad + \left\| (\sigma_1(\mathcal{P}_{\hat{U}_\perp} A), \dots, \sigma_R(\mathcal{P}_{\hat{U}_\perp} A))^\top \right\|
\end{aligned}$$

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

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

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

□

Lemma 24 (Proposition 1 of Cai & Zhang (2018)). *Suppose $Y \in \mathbb{R}^{m \times n}$, $\hat{V} = [\hat{V}_r \ \hat{V}_\perp] \in \mathbb{O}_n$ where $\hat{V}_r \in \mathbb{O}_{n,r}$, $\hat{V}_\perp \in \mathbb{O}_{n,n-r}$ 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*

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

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

Lemma 25 (Properties of the $\sin \Theta$ distances).

The following properties hold for the $\sin \Theta$ distances.

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

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

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

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

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

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

3. (Equivalence with Other Metrics)

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

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

$$\|\sin \Theta(\hat{V}, V)\| \leq \|\hat{V} \hat{V}^T - V V^T\| \leq 2 \|\sin \Theta(\hat{V}, V)\|,$$

$$\|\hat{V} \hat{V}^T - V V^T\|_{\mathbb{F}} = \sqrt{2} \|\sin \Theta(\hat{V}, V)\|_{\mathbb{F}}.$$

Theorem 26 (Mirsky's singular value inequality in Mirsky (1960)). *For any matrices $A, B \in \mathbb{R}^{m \times n}$, let $A = V_1 \Sigma(A) W_1^T$ and $B = V_2 \Sigma(B) W_2^T$ 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*

$$\|\Sigma(A) - \Sigma(B)\| \leq \|A - B\| \quad (57)$$

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 $1 \leq k \leq n$,

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

Therefore

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

and so

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

Then

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

□

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

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

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*

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

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

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

where

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

1944 *Proof.* By symmetry, it suffices to consider $k = 1$. Note that

$$1945 \quad \|X^* \times_k (I_{p_k} - \mathcal{P}_{U_k^*})\|_F = \|X^* \times_k \mathcal{P}_{U_{k\perp}^*}\|_F = \|X^* \times_k U_{k\perp}^*\|_F.$$

1947 In addition

$$1948 \quad \begin{aligned} 1949 \quad \|X^* \times_1 (I_{p_1} - \mathcal{P}_{U_1^*})\|_F &= \|(I_{p_1} - \mathcal{P}_{U_1^*}) \cdot \mathcal{M}_1(X^*)\|_F = \|U_{1\perp}^* U_{1\perp}^{*\top} \cdot (U_1^* \Sigma_1^* V_1^{*\top} + U_{1\perp}^* \Sigma_{1\perp}^* V_{1\perp}^{*\top})\|_F \\ 1950 \quad &= \|U_{1\perp}^* U_{1\perp}^{*\top} \cdot (U_{1\perp}^* \Sigma_{1\perp}^* V_{1\perp}^{*\top})\|_F = \sqrt{\sum_{j=r_1+1}^{\text{rank}(\mathcal{M}_1(X^*))} \sigma_j^2(\mathcal{M}_1(X^*))}. \end{aligned}$$

1954 Note that by the properties of SVD, for any $W \in \mathbb{R}^{p_1 \times p_2 p_3}$ such that $\text{rank}(W) \leq r_1$, it holds that

$$1955 \quad \sqrt{\sum_{j=r_1+1}^{\text{rank}(\mathcal{M}_1(X^*))} \sigma_j^2(\mathcal{M}_1(X^*))} \leq \|\mathcal{M}_1(X^*) - W\|_F.$$

1959 For any $A \in \mathcal{T}_{(r_1, r_2, r_3)}$, it holds that $\text{rank}(\mathcal{M}_1(A)) \leq r_1$. Therefore for any $A \in \mathcal{T}_{(r_1, r_2, r_3)}$,

$$1960 \quad \sqrt{\sum_{j=r_1+1}^{\text{rank}(\mathcal{M}_1(X^*))} \sigma_j^2(\mathcal{M}_1(X^*))} \leq \|\mathcal{M}_1(X^*) - A\|_F.$$

1964 Taking the inf over all $A \in \mathcal{T}_{(r_1, r_2, r_3)}$, it follows that

$$1965 \quad \sqrt{\sum_{j=r_1+1}^{\text{rank}(\mathcal{M}_1(X^*))} \sigma_j^2(\mathcal{M}_1(X^*))} \leq \xi_{(r_1, r_2, r_3)}.$$

1970 \square

1971
1972
1973
1974
1975
1976
1977
1978
1979
1980
1981
1982
1983
1984
1985
1986
1987
1988
1989
1990
1991
1992
1993
1994
1995
1996
1997