Appendix: Tensor Decompositions via Two-Mode Higher-Order SVD (HOSVD)

Miaoyan Wang[1] and Yun S. Song[1,2]

[1]University of Pennsylvania
[2]University of California, Berkeley

A Characterization of Robust Eigenvectors

Proof of Theorem 3.2. The necessity is obvious. To prove the sufficiency, note that the tensor decomposition $\mathcal{T} = \sum_{i=1}^r \lambda_i \boldsymbol{u}_i^{\otimes k}$ implies the two-mode HOSVD:

$$\mathcal{T}_{(12)(3\ldots k)} = \sum_{i=1}^{r} \lambda_i \operatorname{Vec}(\boldsymbol{u}_i^{\otimes 2}) \operatorname{Vec}(\boldsymbol{u}_i^{\otimes (k-2)})^T, \tag{1}$$

where each $\lambda_i > 0$ and $\operatorname{Vec}(\boldsymbol{u}_i^{\otimes 2})$ is the ith left singular vector corresponding to λ_i. Now suppose $\operatorname{Vec}(\boldsymbol{a}^{\otimes 2})$ is the left singular vector of $\mathcal{T}_{(12)(3\ldots k)}$ corresponding to a non-zero singular value $\lambda \in \mathbb{R}\backslash\{0\}$. Then, by (1), we must have

$$\operatorname{Vec}(\boldsymbol{a}^{\otimes 2}) \in \operatorname{Span}\{\operatorname{Vec}(\boldsymbol{u}_i^{\otimes 2})\colon i \in [r] \text{ for which } \lambda_i = \lambda\}.$$

Hence, there exist coefficients $\{\alpha_i\}$ such that $\operatorname{Vec}(\boldsymbol{a}^{\otimes 2}) = \sum_{i\in[r]\colon \lambda_i=\lambda} \alpha_i \operatorname{Vec}(\boldsymbol{u}_i^{\otimes 2})$. In matrix form, this reads

$$\boldsymbol{a}^{\otimes 2} = \sum_{i\in[r]\colon \lambda_i=\lambda} \alpha_i \boldsymbol{u}_i^{\otimes 2},$$

where $\{\boldsymbol{u}_i\}$ is a set of orthonormal vectors. Notice that the matrix on the right-hand side has rank $|\{i \in [r]\colon \lambda_i = \lambda\}|$ while the matrix on the left-hand side has rank 1. Since the rank of a matrix is unambiguously determined, we must have $|\{i \in [r]\colon \lambda_i = \lambda\}| = 1$. Therefore, $\boldsymbol{a}^{\otimes 2} = \boldsymbol{u}_{i^*}^{\otimes 2}$ holds for some $i^* \in [r]$; that is, $\boldsymbol{a}$ is a robust eigenvector of $\mathcal{T}$. □

B Exact Recovery for SOD Tensors

B.1 Proof of Proposition 3.3

Proof of Proposition 3.3. Suppose $\boldsymbol{M}$ is a rank-1 matrix in $\mathcal{LS}_0 = \operatorname{Span}\{\boldsymbol{u}_1^{\otimes 2}, \ldots, \boldsymbol{u}_r^{\otimes 2}\}$, where each $\boldsymbol{u}_i$ is a robust eigenvector of $\mathcal{T}$. Thus, there exist coefficients $\{\alpha_i\}_{i\in[r]}$ such that

$$\boldsymbol{M} = \alpha_1 \boldsymbol{u}_1^{\otimes 2} + \cdots + \alpha_r \boldsymbol{u}_r^{\otimes 2}.$$

Notice that $\{\boldsymbol{u}_i\}$ is a set of orthonormal vectors and the rank of a matrix is unambiguously determined. We must have $|\{i \in [r]\colon \alpha_i \neq 0\}| = 1$. Hence, $\boldsymbol{M} = \alpha_{i^*} \boldsymbol{u}_{i^*}^{\otimes 2}$ holds for some $i^* \in [r]$. □

B.2 Proof of Theorem 3.4

Proof of Theorem 3.4. Note that every matrix $\boldsymbol{M} \in \mathcal{LS}_0$ can be written as $\boldsymbol{M} = \alpha_1 \boldsymbol{u}_1^{\otimes 2} + \cdots + \alpha_r \boldsymbol{u}_r^{\otimes 2}$, where $\{\alpha_i\}_{i\in[r]}$ is a set of scalars in $\mathbb{R}$. Thus, the optimization problem is equivalent to

$$\max_{\alpha_1^2+\cdots+\alpha_r^2=1} \left\|\alpha_1 \boldsymbol{u}_1^{\otimes 2} + \cdots + \alpha_r \boldsymbol{u}_r^{\otimes 2}\right\|_\sigma = \max_{\alpha_1^2+\cdots+\alpha_r^2=1} \max_{i\in[r]} |\alpha_i|. \tag{2}$$

Let $f(\boldsymbol{\alpha}) = \max_{i\in[r]} |\alpha_i|$ denote the objective function in (2), where $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_r)^T \in \boldsymbol{S}^{r-1}$. Notice that the objective is upper bounded by 1; i.e., $f(\boldsymbol{\alpha}) \leq 1$ for all $\boldsymbol{\alpha} \in \mathbf{S}^{r-1}$. Suppose $\boldsymbol{\alpha^*} = (\alpha_1^*, \ldots, \alpha_r^*)^T \in \boldsymbol{S}^{r-1}$ is a local maximizer of (2). We show below that $f(\boldsymbol{\alpha}^*) = 1$.

Suppose $f(\boldsymbol{\alpha}^*) \neq 1$. Then we must have $\max_{i\in[r]} |\alpha_i^*| < 1$. Without loss of generality, assume α_1^* is the element with the largest magnitude in the set $\{\alpha_i^*\}_{i\in[r]}$. Since $|\alpha_1^*| < 1$ and $(\alpha_1^*)^2 + \cdots + (\alpha_r^*)^2 = 1$, there must also exist some $j \geq 2$ such that $\alpha_j^* \neq 0$. Without loss of generality again, assume $\alpha_2^* \neq 0$. Now construct another vector $\widetilde{\boldsymbol{\alpha}} = (\widetilde{\alpha}_1, \ldots, \widetilde{\alpha}_r)^T \in \mathbb{R}^r$, where

$$\widetilde{\alpha}_i = \begin{cases} \alpha_1^* \eta, & i = 1, \\ \operatorname{sign}(\alpha_2^*)\sqrt{(\alpha_2^*)^2 - (\eta^2 - 1)(\alpha_1^*)^2}, & i = 2, \\ \alpha_i^*, & i = 3, \ldots, r, \end{cases}$$

and $\eta \in \mathbb{R}_+$ is any value in $\left(1, \frac{\sqrt{(\alpha_1^*)^2 + (\alpha_2^*)^2}}{\alpha_1^*}\right]$. It is easy to verify that $\widetilde{\boldsymbol{\alpha}} \in \boldsymbol{S}^{r-1}$ for all such η. Moreover,

$$\begin{aligned} \|\widetilde{\boldsymbol{\alpha}} - \boldsymbol{\alpha}^*\|_2^2 &= \sum_{i=1}^{r} (\widetilde{\alpha}_i - \alpha_i^*)^2 = (\alpha_1^*)^2(\eta - 1)^2 + (\alpha_2^* - \widetilde{\alpha}_2)^2 \\ &\leq (\alpha_1^*)^2(\eta - 1)^2 + (\alpha_2^*)^2 + (\widetilde{\alpha}_2)^2 - 2(\widetilde{\alpha}_2)^2 = 2(\alpha_1^*)^2\eta(\eta - 1). \end{aligned}$$

As we see in the right-hand side of the above inequality, the distance between $\widetilde{\boldsymbol{\alpha}}$ and $\boldsymbol{\alpha}^*$ can be arbitrarily small as $\eta \to 1^+$. However, $f(\widetilde{\boldsymbol{\alpha}}) = |\alpha_1^* \eta| > f(\boldsymbol{\alpha}^*)$, which contradicts the local optimality of $\boldsymbol{\alpha}^*$. Hence, we must have $f(\boldsymbol{\alpha}^*) = 1$, which completes the proof of (A1). As an aside, we have also proved that every local maximizer of (2) is a global maximizer.

To see that there are exactly r pairs of maximizers in $\mathcal{LS}_0$, just notice that $\|\boldsymbol{M}^*\|_\sigma / \|\boldsymbol{M}^*\|_F = 1$ is equivalent to saying $\boldsymbol{M}^*$ is a rank-1 matrix. Thus by Proposition 3.3, $\boldsymbol{M}^* = \pm \boldsymbol{u}_i^{\otimes 2}$ for some $i \in [r]$. Conversely, every matrix of the form $\pm \boldsymbol{u}_i^{\otimes 2}$ is a maximizer in $\mathcal{LS}_0$ since $\left\|\boldsymbol{u}_i^{\otimes 2}\right\|_\sigma = 1$. The conclusions (A2) and (A3) then follow from the property of $\{\boldsymbol{u}_i^{\otimes 2}\}_{i\in[r]}$. □

C Two-Mode HOSVD via Nearly Matrix Pursuit

C.1 Auxiliary Theorems

The following results pertain to standard perturbation theory for the singular value decomposition of matrices. For any matrix $\boldsymbol{X}$, we use $\boldsymbol{X}^\dagger$ to denote the Hermitian transpose of $\boldsymbol{X}$. Given a diagonal matrix $\boldsymbol{\Sigma}$ of singular values, let $\sigma_{\min}(\boldsymbol{\Sigma})$ and $\sigma_{\max}(\boldsymbol{\Sigma})$ denote, respectively, the minimum and the maximum singular values in $\boldsymbol{\Sigma}$.

Theorem C.1 (Wedin [3])**.** *Let $\boldsymbol{B}$ and $\widetilde{\boldsymbol{B}}$ be two $m \times n$ ($m \geq n$) real or complex matrices with SVDs*

$$\boldsymbol{B} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^\dagger \equiv (\boldsymbol{U}_1, \boldsymbol{U}_2) \begin{pmatrix} \boldsymbol{\Sigma}_1 & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{\Sigma}_2 \\ \boldsymbol{0} & \boldsymbol{0} \end{pmatrix} \begin{pmatrix} \boldsymbol{V}_1^\dagger \\ \boldsymbol{V}_2^\dagger \end{pmatrix}, \tag{3}$$

$$\widetilde{\boldsymbol{B}} = \widetilde{\boldsymbol{U}}\widetilde{\boldsymbol{\Sigma}}\widetilde{\boldsymbol{V}}^\dagger \equiv \left(\widetilde{\boldsymbol{U}}_1, \widetilde{\boldsymbol{U}}_2\right) \begin{pmatrix} \widetilde{\boldsymbol{\Sigma}}_1 & \boldsymbol{0} \\ \boldsymbol{0} & \widetilde{\boldsymbol{\Sigma}}_2 \\ \boldsymbol{0} & \boldsymbol{0} \end{pmatrix} \begin{pmatrix} \widetilde{\boldsymbol{V}}_1^\dagger \\ \widetilde{\boldsymbol{V}}_2^\dagger \end{pmatrix}, \tag{4}$$

and

$$\begin{aligned} \boldsymbol{\Sigma}_1 &= \operatorname{diag}(\sigma_1, \ldots, \sigma_k), \quad \boldsymbol{\Sigma}_2 = \operatorname{diag}(\sigma_{k+1}, \ldots, \sigma_n), \\ \widetilde{\boldsymbol{\Sigma}}_1 &= \operatorname{diag}(\widetilde{\sigma}_1, \ldots, \widetilde{\sigma}_k), \quad \widetilde{\boldsymbol{\Sigma}}_2 = \operatorname{diag}(\widetilde{\sigma}_{k+1}, \ldots, \widetilde{\sigma}_n), \end{aligned} \tag{5}$$

with $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_n$ and $\tilde{\sigma}_1 \geq \tilde{\sigma}_2 \geq \cdots \geq \tilde{\sigma}_n$ in descending order. If there exist an $\alpha \geq 0$ and a $\delta > 0$ such that

$$\sigma_{\min}(\boldsymbol{\Sigma}_1) = \sigma_k \geq \alpha + \delta \quad \textit{and} \quad \sigma_{\max}(\widetilde{\boldsymbol{\Sigma}}_2) = \tilde{\sigma}_{k+1} \leq \alpha, \tag{6}$$

then

$$\max\left\{\left\|\sin\Theta(\boldsymbol{U}_1, \widetilde{\boldsymbol{U}}_1)\right\|_\sigma, \left\|\sin\Theta(\boldsymbol{V}_1, \widetilde{\boldsymbol{V}}_1)\right\|_\sigma\right\} \leq \frac{\max\left\{\left\|\widetilde{\boldsymbol{B}}\boldsymbol{V}_1 - \boldsymbol{U}_1\boldsymbol{\Sigma}_1\right\|_\sigma, \left\|\widetilde{\boldsymbol{B}}^\dagger\boldsymbol{U}_1 - \boldsymbol{V}_1\boldsymbol{\Sigma}_1\right\|_\sigma\right\}}{\delta}.$$

Remark C.2. In the above theorem, $\boldsymbol{U}_1$, $\widetilde{\boldsymbol{U}}_1$ are d-by-k matrices and $\Theta(\boldsymbol{U}_1, \widetilde{\boldsymbol{U}}_1)$ denotes the matrix of canonical angles between the ranges of $\boldsymbol{U}_1$ and $\widetilde{\boldsymbol{U}}_1$. If we let $\mathcal{L}$ (standing for "left" singular vectors) and $\widetilde{\mathcal{L}}$ denote the column spaces of $\boldsymbol{U}_1$ and $\widetilde{\boldsymbol{U}}_1$ respectively, then by definition, $\left\|\sin\Theta(\boldsymbol{U}_1, \widetilde{\boldsymbol{U}}_1)\right\|_\sigma \stackrel{\text{def}}{=} \left\|\boldsymbol{U}_1^T \widetilde{\boldsymbol{U}}_1^\perp\right\|_\sigma = \max_{\boldsymbol{x}\in\mathcal{L}, \boldsymbol{y}\in\widetilde{\mathcal{L}}} \frac{\boldsymbol{x}^T\boldsymbol{y}}{\|\boldsymbol{x}\|_2\|\boldsymbol{y}\|_2}$. When no confusion arises, we will simply use $\sin\Theta(\mathcal{L}, \widetilde{\mathcal{L}})$ to denote $\left\|\sin\Theta(\boldsymbol{U}_1, \widetilde{\boldsymbol{U}}_1)\right\|_\sigma$.

Proposition C.3. *Let $\mathcal{L}_1$, $\mathcal{L}_2$ be two subspaces in $\mathbb{R}^d$. Then for any vector $\boldsymbol{u}_1 \in \mathcal{L}_1$,*

$$\sin\Theta\left(\boldsymbol{u}_1, \mathcal{L}_2\right) \leq \sin\Theta\left(\mathcal{L}_1, \mathcal{L}_2\right).$$

Proof. The conclusion follows readily from Remark C.2. □

Theorem C.4 (Weyl [4]). *Let $\boldsymbol{B}$ and $\widetilde{\boldsymbol{B}}$ be two matrices with SVDs* (3)*,* (4)*, and* (5)*, Then,*

$$|\widetilde{\sigma}_i - \sigma_i| \leq \left\|\widetilde{\boldsymbol{B}} - \boldsymbol{B}\right\|_\sigma \quad \text{for all } i = 1, \ldots, n.$$

In our proofs, we often make use of the following corollary based on Wedin's and Weyl's Theorems.

Corollary C.5. *Let $\boldsymbol{B}$ and $\widetilde{\boldsymbol{B}}$ be two matrices with SVDs* (3)*,* (4)*, and* (5)*. Let $\boldsymbol{E} \stackrel{\text{def}}{=} \widetilde{\boldsymbol{B}} - \boldsymbol{B}$, and $\mathcal{L}$, $\mathcal{R}$, $\widetilde{\mathcal{L}}$ and $\widetilde{\mathcal{R}}$ be the column spaces of $\boldsymbol{U}_1$, $\boldsymbol{V}_1$, $\widetilde{\boldsymbol{U}}_1$ and $\widetilde{\boldsymbol{V}}_1$, respectively. Define $\Delta = \min\left\{\sigma_{\min}(\boldsymbol{\Sigma}_1), \sigma_{\min}(\boldsymbol{\Sigma}_1) - \sigma_{\max}(\boldsymbol{\Sigma}_2)\right\}$. If $\Delta > \|\boldsymbol{E}\|_\sigma$, then*

$$\max\left\{\sin\Theta(\mathcal{L}, \widetilde{\mathcal{L}}), \sin\Theta(\mathcal{R}, \widetilde{\mathcal{R}})\right\} \leq \frac{\|\boldsymbol{E}\|_\sigma}{\Delta - \|\boldsymbol{E}\|_\sigma}. \tag{7}$$

Proof. By Weyl's theorem, $\sigma_{\max}(\boldsymbol{\Sigma}_2) - \sigma_{\max}(\widetilde{\boldsymbol{\Sigma}}_2) \geq -\|\boldsymbol{E}\|_\sigma$. Combining this with the assumption $\sigma_{\min}(\boldsymbol{\Sigma}_1) - \sigma_{\max}(\boldsymbol{\Sigma}_2) > \|\boldsymbol{E}\|_\sigma$, we have

$$\sigma_{\min}(\boldsymbol{\Sigma}_1) - \sigma_{\max}(\widetilde{\boldsymbol{\Sigma}}_2) = \sigma_{\min}(\boldsymbol{\Sigma}_1) - \sigma_{\max}(\boldsymbol{\Sigma}_2) + \sigma_{\max}(\boldsymbol{\Sigma}_2) - \sigma_{\max}(\widetilde{\boldsymbol{\Sigma}}_2) > \|\boldsymbol{E}\|_\sigma - \|\boldsymbol{E}\|_\sigma = 0.$$

This implies that the spectrum of $\boldsymbol{\Sigma}_1$ is well-separated from that of $\widetilde{\boldsymbol{\Sigma}}_2$, and thus (6) holds with $\alpha = \max\{0, \sigma_{\max}(\widetilde{\boldsymbol{\Sigma}}_2)\} \geq 0$ and $\delta = \sigma_{\min}(\boldsymbol{\Sigma}_1) - \alpha > 0$. By Wedin's theorem, we get

$$\max\left\{\sin\Theta(\mathcal{L}, \widetilde{\mathcal{L}}), \sin\Theta(\mathcal{R}, \widetilde{\mathcal{R}})\right\} \leq \frac{\left\{\left\|\widetilde{\boldsymbol{B}}\boldsymbol{V}_1 - \boldsymbol{U}_1\boldsymbol{\Sigma}_1\right\|_\sigma, \left\|\widetilde{\boldsymbol{B}}^\dagger\boldsymbol{U}_1 - \boldsymbol{V}_1\boldsymbol{\Sigma}_1\right\|_\sigma\right\}}{\delta}.$$

Then, noting

$$\left\|\widetilde{\boldsymbol{B}}\boldsymbol{V}_1 - \boldsymbol{U}_1\boldsymbol{\Sigma}_1\right\|_\sigma = \left\|\widetilde{\boldsymbol{B}}\boldsymbol{V}_1 - \boldsymbol{B}\boldsymbol{V}_1\right\|_\sigma = \left\|\widetilde{\boldsymbol{B}} - \boldsymbol{B}\right\|_\sigma = \|\boldsymbol{E}\|_\sigma,$$

$$\left\|\widetilde{\boldsymbol{B}}^\dagger\boldsymbol{U}_1 - \boldsymbol{V}_1\boldsymbol{\Sigma}_1\right\|_\sigma = \left\|\widetilde{\boldsymbol{B}}^\dagger\boldsymbol{U}_1 - \boldsymbol{B}^\dagger\boldsymbol{U}_1\right\|_\sigma = \left\|\widetilde{\boldsymbol{B}}^\dagger - \boldsymbol{B}^\dagger\right\|_\sigma = \|\boldsymbol{E}\|_\sigma,$$

and

$$\delta = \sigma_{\min}(\boldsymbol{\Sigma}_1) - \max\{0, \sigma_{\max}(\widetilde{\boldsymbol{\Sigma}}_2)\} \geq \sigma_{\min}(\boldsymbol{\Sigma}_1) - \max\{0, \sigma_{\max}(\boldsymbol{\Sigma}_2)\} - \|\boldsymbol{E}\|_\sigma = \Delta - \|\boldsymbol{E}\|_\sigma,$$

we obtain (7). □

Lemma C.6 (Taylor Expansion). *If $\varepsilon = o(1)$, then*

- $(1+\varepsilon)^\alpha = 1 + \alpha\varepsilon + o(\varepsilon), \quad \forall \alpha \in \mathbb{R}$;
- $\sin\varepsilon = \varepsilon + o(\varepsilon^2)$;
- $\cos\varepsilon = 1 - \frac{1}{2}\varepsilon^2 + o(\varepsilon^2)$.

C.2 Proof of Proposition 4.2 (Uniqueness of $\mathcal{LS}^{(r)}$)

Proof of Proposition 4.2. Let $\widetilde{\mathcal{T}}_{(12)(3\dots k)} = \sum_i \mu_i \boldsymbol{a}_i \boldsymbol{b}_i^T$ be the two-mode HOSVD with $\{\mu_i\}$ in descending order, and $\mathcal{LS}^{(r)} = \operatorname{Span}\{\boldsymbol{a}_1, \dots, \boldsymbol{a}_r\}$ is the r-truncated two-mode singular space. In order to show that $\mathcal{LS}^{(r)}$ is uniquely determined, it suffices to show that μ_r is strictly larger than μ_{r+1}.

Note that the tensor perturbation model $\widetilde{\mathcal{T}} = \sum_{i=1}^r \lambda_i \boldsymbol{u}_i^{\otimes k} + \mathcal{E}$ implies the matrix perturbation model

$$\widetilde{\mathcal{T}}_{(12)(3\dots k)} = \sum_{i=1}^r \lambda_i \operatorname{Vec}(\boldsymbol{u}_i^{\otimes 2}) \operatorname{Vec}(\boldsymbol{u}_i^{\otimes (k-2)})^T + \mathcal{E}_{(12)(3\dots k)}, \tag{8}$$

where by [2]

$$\left\| \mathcal{E}_{(12)(3\dots k)} \right\|_\sigma \le d^{(k-2)/2} \left\| \mathcal{E} \right\|_\sigma \le d^{(k-2)/2} \varepsilon. \tag{9}$$

Now apply Corollary C.5 to (8) with $\widetilde{\boldsymbol{B}} = \widetilde{\mathcal{T}}_{(12)(3\dots k)}$, $\boldsymbol{B} = \sum_{i=1}^r \lambda_i \operatorname{Vec}(\boldsymbol{u}_i^{\otimes 2}) \operatorname{Vec}(\boldsymbol{u}_i^{\otimes (k-2)})^T$, and $\widetilde{\boldsymbol{B}} - \boldsymbol{B} = \mathcal{E}_{(12)(3\dots k)}$. Considering the corresponding rth and $(r+1)$th singular values of $\widetilde{\boldsymbol{B}}$ and $\boldsymbol{B}$, we obtain

$$|\mu_r - \lambda_r| \le \left\| \mathcal{E}_{(12)(3\dots k)} \right\|_\sigma, \quad \text{and} \quad |\mu_{r+1} - 0| \le \left\| \mathcal{E}_{(12)(3\dots k)} \right\|_\sigma,$$

which implies

$$\mu_r - \mu_{r+1} = \lambda_r + (\mu_r - \lambda_r) - (\mu_{r+1} - 0) \ge \lambda_r - 2 \left\| \mathcal{E}_{(12)(3\dots k)} \right\|_\sigma.$$

By (9) and Assumption 4.1,

$$\lambda_r - 2 \left\| \mathcal{E}_{(12)(3\dots k)} \right\|_\sigma \ge \lambda_{\min} - 2 d^{(k-2)/2} \varepsilon > 0.$$

Therefore $\mu_r > \mu_{r+1}$, which ensures the uniqueness of $\mathcal{LS}^{(r)}$. □

C.3 Proof of Theorem 4.4 (Perturbation of $\mathcal{LS}_0$)

Definition C.7 (Singular Space)**.** Let $\widetilde{\mathcal{T}}_{(12)(3\dots k)} \in \mathbb{R}^{d^2 \times d^{k-2}}$ be the two-mode unfolding of $\widetilde{\mathcal{T}}$, and $\widetilde{\mathcal{T}}_{(12)(3\dots k)} = \sum_i \mu_i \boldsymbol{a}_i \boldsymbol{b}_i^T$ be the two-mode HOSVD with $\mu_1 \ge \mu_2 \ge \cdots \ge \mu_r$ in descending order. We define the r-truncated left (respectively, right) singular space by

$$\mathcal{LS}^{(r)} = \operatorname{Span}\left\{ \operatorname{Mat}(\boldsymbol{a}_i) \in \mathbb{R}^{d \times d} \colon \boldsymbol{a}_i \text{ is the } i\text{th left singular vector of } \widetilde{\mathcal{T}}_{(12)(3\dots k)}, i \in [r] \right\},$$
$$\mathcal{RS}^{(r)} = \operatorname{Span}\left\{ \boldsymbol{b}_i \in \mathbb{R}^{d^{k-2}} \colon \boldsymbol{b}_i \text{ is the } i\text{th right singular vector of } \widetilde{\mathcal{T}}_{(12)(3\dots k)}, i \in [r] \right\}.$$

The noise-free version ($\varepsilon = 0$) reduces to

$$\mathcal{LS}_0 = \operatorname{Span}\left\{ \boldsymbol{u}_i^{\otimes 2} \colon i \in [r] \right\}, \quad \text{and} \quad \mathcal{RS}_0 = \operatorname{Span}\left\{ \operatorname{Vec}(\boldsymbol{u}_i^{\otimes (k-2)}) \colon i \in [r] \right\}.$$

Remark C.8. We make the convention that the elements in $\mathcal{LS}^{(r)}$ (respectively, $\mathcal{LS}_0$) are viewed as d-by-d matrices, while the elements in $\mathcal{RS}^{(r)}$ (respectively, $\mathcal{RS}_0$) are viewed as length-d^{k-2} vectors. For g of notation, we drop the subscript r from $\mathcal{LS}^{(r)}$ (respectively, $\mathcal{RS}^{(r)}$) and simply write $\mathcal{LS}$ (respectively, $\mathcal{RS}$) hereafter.

Definition C.9 (Inner-Product)**.** For any two tensors $\mathcal{A} = [\![a_{i_1 \dots i_k}]\!]$, $\mathcal{B} = [\![b_{i_1 \dots i_k}]\!] \in \mathbb{R}^{d_1 \times \cdots \times d_k}$ of identical order and dimensions, their inner product is defined as

$$\langle \mathcal{A}, \mathcal{B} \rangle = \sum_{i_1, \dots, i_k} a_{i_1 \dots i_k} b_{i_1 \dots i_k},$$

while the tensor Frobenius norm of $\mathcal{A}$ is defined as

$$\|\mathcal{A}\|_F = \sqrt{\langle \mathcal{A}, \mathcal{A} \rangle} = \sqrt{\sum_{i_1, \dots, i_k} |a_{i_1 \dots i_k}|^2},$$

both of which are analogues of standard definitions for vectors and matrices.

Lemma C.10. *For every matrix $\boldsymbol{M} \in \mathcal{LS}$ satisfying $\|\boldsymbol{M}\|_F = 1$, there exists a unit vector $\boldsymbol{b}_{\boldsymbol{M}} \in \mathcal{RS}$ such that*

$$\boldsymbol{M} = c\widetilde{\mathcal{T}}_{(1)(2)(3\dots k)}(\boldsymbol{I}, \boldsymbol{I}, \boldsymbol{b}_{\boldsymbol{M}}), \tag{10}$$

where $c = 1/\left\|\widetilde{\mathcal{T}}_{(1)(2)(3\dots k)}(\boldsymbol{I}, \boldsymbol{I}, \boldsymbol{b}_{\boldsymbol{M}})\right\|_F$ is a normalizing constant.

Proof. Let $\widetilde{\mathcal{T}}_{(12)(3\dots k)} = \sum_i \mu_i \boldsymbol{a}_i \boldsymbol{b}_i^T$ denote the two-mode HOSVD. Following a similar line of argument as in the proof of Proposition 4.2, we have $\mu_r \geq \lambda_{\min} - \left\|\mathcal{E}_{(12)(3\dots k)}\right\|_\sigma > 0$. By the property of matrix SVD,

$$\boldsymbol{a}_i = \frac{1}{\mu_i}\widetilde{\mathcal{T}}_{(12)(3\dots k)}\boldsymbol{b}_i, \quad \text{for all } i \in [r],$$

which implies

$$\text{Mat}(\boldsymbol{a}_i) = \frac{1}{\mu_i}\widetilde{\mathcal{T}}_{(1)(2)(3\dots k)}(\boldsymbol{I}, \boldsymbol{I}, \boldsymbol{b}_i), \quad \text{for all } i \in [r].$$

Recall that $\mathcal{LS} = \text{Span}\{\text{Mat}(\boldsymbol{a}_i)\colon i \in [r]\}$. Thus, for any $\boldsymbol{M} \in \mathcal{LS}$, there exist coefficients $\{\alpha_i\}_{i\in[r]}$ such that

$$\begin{aligned}
\boldsymbol{M} &= \alpha_1 \text{Mat}(\boldsymbol{a}_1) + \dots + \alpha_r \text{Mat}(\boldsymbol{a}_r) \\
&= \frac{\alpha_1}{\mu_1}\widetilde{\mathcal{T}}_{(1)(2)(3\dots k)}(\boldsymbol{I}, \boldsymbol{I}, \boldsymbol{b}_1) + \dots + \frac{\alpha_r}{\mu_r}\widetilde{\mathcal{T}}_{(1)(2)(3\dots k)}(\boldsymbol{I}, \boldsymbol{I}, \boldsymbol{b}_r) \\
&= \widetilde{\mathcal{T}}_{(1)(2)(3\dots k)}\left(\boldsymbol{I}, \boldsymbol{I}, \frac{\alpha_1}{\mu_1}\boldsymbol{b}_1 + \dots + \frac{\alpha_r}{\mu_r}\boldsymbol{b}_r\right),
\end{aligned}$$

where the last line follows from the multilinearity of $\mathcal{T}_{(1)(2)(3\dots k)}$. Now define $\boldsymbol{b}'_{\boldsymbol{M}} = \frac{\alpha_1}{\mu_1}\boldsymbol{b}_1 + \dots + \frac{\alpha_r}{\mu_r}\boldsymbol{b}_r$. The conclusion (10) then follows by setting $\boldsymbol{b}_{\boldsymbol{M}} = \boldsymbol{b}'_{\boldsymbol{M}}/\left\|\boldsymbol{b}'_{\boldsymbol{M}}\right\|_2 \in \mathcal{RS}$. □

Lemma C.11 (Perturbation of $\mathcal{RS}_0$)**.** *Under Assumption 4.1,*

$$\min_{\boldsymbol{b}\in\mathcal{RS}, \|\boldsymbol{b}\|_2=1} \left\|\boldsymbol{b}\big|_{\mathcal{RS}_0}\right\|_2 \geq 1 - \frac{d^{k-2}}{2\lambda_{\min}^2}\varepsilon^2 + o(\varepsilon^2).$$

where $\boldsymbol{b}\big|_{\mathcal{RS}_0}$ denotes the vector projection of $\boldsymbol{b} \in \mathcal{RS}$ onto the space $\mathcal{RS}_0$.

Proof. As seen in the proof of Proposition 4.2, $\widetilde{\mathcal{T}}_{(12)(3\dots k)}$ can be written as

$$\widetilde{\mathcal{T}}_{(12)(3\dots k)} = \sum_{i=1}^r \lambda_i \operatorname{Vec}(\boldsymbol{u}_i^{\otimes 2}) \operatorname{Vec}(\boldsymbol{u}_i^{\otimes(k-2)})^T + \mathcal{E}_{(12)(3\dots k)}, \quad \text{where} \quad \left\|\mathcal{E}_{(12)(3\dots k)}\right\|_\sigma \leq d^{(k-2)/2}\varepsilon. \tag{11}$$

The noise-free version of (11) reduces to

$$\mathcal{T}_{(12)(3\dots k)} = \sum_{i=1}^r \lambda_i \operatorname{Vec}(\boldsymbol{u}_i^{\otimes 2}) \operatorname{Vec}(\boldsymbol{u}_i^{\otimes(k-2)})^T.$$

Following the notation of Corollary C.5, we set $\widetilde{\boldsymbol{B}} = \widetilde{\mathcal{T}}_{(12)(3\dots k)}$, $\boldsymbol{B} = \mathcal{T}_{(12)(3\dots k)}$, $\boldsymbol{\Sigma}_1 = \text{diag}\{\lambda_1, \dots, \lambda_r\}$, $\boldsymbol{\Sigma}_2 = \text{diag}\{0, \dots, 0\}$, and $\Delta = \min\{\sigma_{\min}(\boldsymbol{\Sigma}_1), \sigma_{\min}(\boldsymbol{\Sigma}_1) - \sigma_{\max}(\boldsymbol{\Sigma}_2)\} = \min_{i\in[r]}\lambda_i$. Then, $\left\|\widetilde{\boldsymbol{B}} - \boldsymbol{B}\right\|_\sigma = \left\|\mathcal{E}_{(12)(3\dots k)}\right\|_\sigma$. By Assumption 4.1, $\Delta = \lambda_{\min} > 2d^{(k-2)/2}\varepsilon > \left\|\mathcal{E}_{(12)(3\dots k)}\right\|_\sigma$. Hence the condition of Corollary C.5 holds. Applying Corollary C.5 then yields

$$\begin{aligned}
\sin\Theta\left(\mathcal{RS}_0, \mathcal{RS}\right) &\leq \frac{\left\|\mathcal{E}_{(12)(3\dots k)}\right\|_\sigma}{\lambda_{\min} - \left\|\mathcal{E}_{(12)(3\dots k)}\right\|_\sigma} = \frac{\left\|\mathcal{E}_{(12)(3\dots k)}\right\|_\sigma}{\lambda_{\min}}\left[1 - \frac{\left\|\mathcal{E}_{(12)(3\dots k)}\right\|_\sigma}{\lambda_{\min}}\right]^{-1} \\
&\leq \frac{d^{(k-2)/2}\varepsilon}{\lambda_{\min}}\left[1 - \frac{d^{(k-2)/2}\varepsilon}{\lambda_{\min}}\right]^{-1} = \frac{d^{(k-2)/2}}{\lambda_{\min}}\varepsilon + o(\varepsilon).
\end{aligned} \tag{12}$$

Now let $\boldsymbol{b} \in \mathcal{RS}$ be a unit vector. Decompose $\boldsymbol{b}$ into

$$\boldsymbol{b} = \boldsymbol{b}\big|_{\mathcal{RS}_0} + \boldsymbol{b}\big|_{\mathcal{RS}_0{}^{\perp}},$$

where $\boldsymbol{b}\big|_{\mathcal{RS}_0}$ and $\boldsymbol{b}\big|_{\mathcal{RS}_0^{\perp}}$ are vector projections of $\boldsymbol{b}$ onto the spaces $\mathcal{RS}_0$ and $\mathcal{RS}_0^{\perp}$, respectively. By (12) and Taylor expansion,

$$\left\|\boldsymbol{b}\big|_{\mathcal{RS}_0}\right\|_2 = \cos\Theta\left(\boldsymbol{b}, \mathcal{RS}_0\right) = \left[1 - \sin^2\Theta\left(\boldsymbol{b}, \mathcal{RS}_0\right)\right]^{1/2} \geq 1 - \frac{d^{k-2}}{2\lambda_{\min}^2}\varepsilon^2 + o(\varepsilon^2).$$

Since the above holds for every unit vector $\boldsymbol{b} \in \mathcal{RS}$, we conclude

$$\min_{\boldsymbol{b}\in\mathcal{RS}, \|\boldsymbol{b}\|_2=1} \left\|\boldsymbol{b}\big|_{\mathcal{RS}_0}\right\|_2 \geq 1 - \frac{d^{k-2}}{2\lambda_{\min}^2}\varepsilon^2 + o(\varepsilon^2).$$

□

Corollary C.12. *Under Assumption 4.1,*

$$\min_{\boldsymbol{b}\in\mathcal{RS}, \|\boldsymbol{b}\|_2=1} \left\|\boldsymbol{b}\big|_{\mathcal{RS}_0}\right\|_2 \geq 1 - \frac{1}{(c_0-1)^2},$$

which is ≥ 0.98 *for* $c_0 \geq 10$.

Proof. Note that $\frac{\left\|\mathcal{E}_{(12)(3\ldots k)}\right\|_\sigma}{\lambda_{\min}} \leq \frac{1}{c_0}$ by Assumption 4.1. The right-hand side of (12) can be bounded as follows,

$$\frac{\left\|\mathcal{E}_{(12)(3\ldots k)}\right\|_\sigma}{\lambda_{\min} - \left\|\mathcal{E}_{(12)(3\ldots k)}\right\|_\sigma} \leq \frac{1}{c_0 - 1}.$$

By a similar argument as in the proof of Lemma C.11, we obtain

$$\min_{\boldsymbol{b}\in\mathcal{RS}, \|\boldsymbol{b}\|_2=1} \left\|\boldsymbol{b}\big|_{\mathcal{RS}_0}\right\|_2 = \cos\Theta(\boldsymbol{b}, \mathcal{RS}_0) \geq \cos^2\Theta(\boldsymbol{b}, \mathcal{RS}_0) \geq \cos^2\Theta(\mathcal{RS}, \mathcal{RS}_0) \geq 1 - \frac{1}{(c_0-1)^2}, \tag{13}$$

which is the desired result. □

Proof of Theorem 4.4. To prove the upper bound in Theorem 4.4, it suffices to show that for every matrix $\boldsymbol{M} \in \mathcal{LS}$ satisfying $\|\boldsymbol{M}\|_F = 1$, there exist coefficients $\{\alpha_i \in \mathbb{R}\}_{i=1}^r$ such that

$$\boldsymbol{M} = \sum_{i=1}^r \alpha_i \boldsymbol{u}_i^{\otimes 2} + \boldsymbol{E}, \quad \text{where } \|\boldsymbol{E}\|_\sigma \leq \frac{d^{(k-3)/2}}{\lambda_{\min}}\varepsilon + o(\varepsilon). \tag{14}$$

Let $\boldsymbol{M}$ be a d-by-d matrix satisfying $\boldsymbol{M} \in \mathcal{LS}$ and $\|\boldsymbol{M}\|_F = 1$. By Lemma C.10, there exists $\boldsymbol{b}_{\boldsymbol{M}} \in \mathcal{RS}$ such that

$$\begin{aligned}\boldsymbol{M} &= \frac{\widetilde{\mathcal{T}}_{(1)(2)(3\ldots k)}(\boldsymbol{I}, \boldsymbol{I}, \boldsymbol{b}_{\boldsymbol{M}})}{\left\|\widetilde{\mathcal{T}}_{(1)(2)(3\ldots k)}(\boldsymbol{I}, \boldsymbol{I}, \boldsymbol{b}_{\boldsymbol{M}})\right\|_F} \\ &= \sum_{i=1}^r \frac{\lambda_i \langle \mathrm{Vec}(\boldsymbol{u}_i^{\otimes(k-2)}), \boldsymbol{b}_{\boldsymbol{M}}\rangle}{\left\|\widetilde{\mathcal{T}}_{(1)(2)(3\ldots k)}(\boldsymbol{I}, \boldsymbol{I}, \boldsymbol{b}_{\boldsymbol{M}})\right\|_F} \boldsymbol{u}_i^{\otimes 2} + \frac{\mathcal{E}_{(1)(2)(3\ldots k)}(\boldsymbol{I}, \boldsymbol{I}, \boldsymbol{b}_{\boldsymbol{M}})}{\left\|\widetilde{\mathcal{T}}_{(1)(2)(3\ldots k)}(\boldsymbol{I}, \boldsymbol{I}, \boldsymbol{b}_{\boldsymbol{M}})\right\|_F}.\end{aligned} \tag{15}$$

We now claim that (15) is a desired decomposition that satisfies (14). Namely, we seek to prove

$$\frac{\left\|\mathcal{E}_{(1)(2)(3\ldots k)}(\boldsymbol{I}, \boldsymbol{I}, \boldsymbol{b}_{\boldsymbol{M}})\right\|_\sigma}{\left\|\widetilde{\mathcal{T}}_{(1)(2)(3\ldots k)}(\boldsymbol{I}, \boldsymbol{I}, \boldsymbol{b}_{\boldsymbol{M}})\right\|_F} \leq \frac{d^{(k-3)/2}}{\lambda_{\min}}\varepsilon + o(\varepsilon). \tag{16}$$

Observe that by the triangle inequality,

$$\begin{aligned}\left\|\widetilde{\mathcal{T}}_{(1)(2)(3\ldots k)}(\boldsymbol{I},\boldsymbol{I},\boldsymbol{b}_{\boldsymbol{M}})\right\|_F &= \left\|\sum_{i=1}^r \lambda_i \langle \mathrm{Vec}(\boldsymbol{u}_i^{\otimes(k-2)}),\boldsymbol{b}_{\boldsymbol{M}}\rangle \boldsymbol{u}_i^{\otimes 2} + \mathcal{E}_{(1)(2)(3\ldots k)}(\boldsymbol{I},\boldsymbol{I},\boldsymbol{b}_{\boldsymbol{M}})\right\|_F \\ &\geq \underbrace{\left\|\sum_{i=1}^r \lambda_i \langle \mathrm{Vec}(\boldsymbol{u}_i^{\otimes(k-2)}),\boldsymbol{b}_{\boldsymbol{M}}\rangle \boldsymbol{u}_i^{\otimes 2}\right\|_F}_{\text{Part I}} - \underbrace{\left\|\mathcal{E}_{(1)(2)(3\ldots k)}(\boldsymbol{I},\boldsymbol{I},\boldsymbol{b}_{\boldsymbol{M}})\right\|_F}_{\text{Part II}}.\end{aligned} \tag{17}$$

By the orthogonality of $\{\boldsymbol{u}_i\}_{i\in[r]}$, Part I has a lower bound,

$$\begin{aligned}\left\|\sum_{i=1}^r \lambda_i \langle \mathrm{Vec}(\boldsymbol{u}_i^{\otimes(k-2)}),\boldsymbol{b}_{\boldsymbol{M}}\rangle \boldsymbol{u}_i^{\otimes 2}\right\|_F &\geq \lambda_{\min}\sqrt{\sum_{i=1}^r \langle \mathrm{Vec}(\boldsymbol{u}_i^{\otimes(k-2)}),\boldsymbol{b}_{\boldsymbol{M}}\rangle^2} \\ &= \lambda_{\min}\left\|\boldsymbol{b}_{\boldsymbol{M}}|_{\mathcal{RS}_0}\right\|_2.\end{aligned} \tag{18}$$

By the inequality between the Frobenius norm and the spectral norm for matrices, Part II has an upper bound,

$$\left\|\mathcal{E}_{(1)(2)(3\ldots k)}(\boldsymbol{I},\boldsymbol{I},\boldsymbol{b}_{\boldsymbol{M}})\right\|_F \leq \sqrt{d}\left\|\mathcal{E}_{(1)(2)(3\ldots k)}(\boldsymbol{I},\boldsymbol{I},\boldsymbol{b}_{\boldsymbol{M}})\right\|_\sigma \leq \sqrt{d}\left\|\mathcal{E}_{(1)(2)(3\ldots k)}\right\|_\sigma \leq d^{(k-2)/2}\varepsilon, \tag{19}$$

where we have used the inequality [2] that

$$\left\|\mathcal{E}_{(1)(2)(3\ldots k)}\right\|_\sigma \leq d^{(k-3)/2}\left\|\mathcal{E}\right\|_\sigma. \tag{20}$$

Combining (17), (18) and (19) gives

$$\left\|\widetilde{\mathcal{T}}_{(1)(2)(3\ldots k)}(\boldsymbol{I},\boldsymbol{I},\boldsymbol{b}_{\boldsymbol{M}})\right\|_F \geq \lambda_{\min}\left[\left\|\boldsymbol{b}_{\boldsymbol{M}}|_{\mathcal{RS}_0}\right\|_2 - \frac{d^{(k-2)/2}\varepsilon}{\lambda_{\min}}\right]. \tag{21}$$

By Corollary C.12 and Assumption 4.1 with $c_0 \geq 10$, $\left\|\boldsymbol{b}_{\boldsymbol{M}}|_{\mathcal{RS}_0}\right\|_2 - \frac{d^{(k-2)/2}\varepsilon}{\lambda_{\min}} \geq 0.98 - 0.1 > 0$. So the right-hand side of (21) is strictly positive. Taking the reciprocal of (21) and combining it with (20), we obtain

$$\begin{aligned}\frac{\left\|\mathcal{E}_{(1)(2)(3\ldots k)}(\boldsymbol{I},\boldsymbol{I},\boldsymbol{b}_{\boldsymbol{M}})\right\|_\sigma}{\left\|\widetilde{\mathcal{T}}_{(1)(2)(3\ldots k)}(\boldsymbol{I},\boldsymbol{I},\boldsymbol{b}_{\boldsymbol{M}})\right\|_F} &\leq \frac{d^{(k-3)/2}\varepsilon}{\lambda_{\min}}\left[\left\|\boldsymbol{b}_{\boldsymbol{M}}|_{\mathcal{RS}_0}\right\|_\sigma - \frac{d^{(k-2)/2}\varepsilon}{\lambda_{\min}}\right]^{-1} \\ &\leq \frac{d^{(k-3)/2}\varepsilon}{\lambda_{\min}}\left[1 - o(\varepsilon) - \frac{d^{(k-2)/2}\varepsilon}{\lambda_{\min}}\right]^{-1} = \frac{d^{(k-3)/2}}{\lambda_{\min}}\varepsilon + o(\varepsilon),\end{aligned} \tag{22}$$

where the second line follows from Lemma C.11. This completes the proof of (16) and therefore (14). Since (14) holds for every $\boldsymbol{M} \in \mathcal{LS}$ that satisfies $\|\boldsymbol{M}\|_F = 1$, and $\sum_{i=1}^r \alpha_i \boldsymbol{u}_i^{\otimes 2} \in \mathcal{LS}_0$, we immediately have

$$\max_{\boldsymbol{M}\in\mathcal{LS},\|\boldsymbol{M}\|_F=1}\ \min_{\boldsymbol{M}^*\in\mathcal{LS}_0}\|\boldsymbol{M}-\boldsymbol{M}^*\|_\sigma \leq \frac{d^{(k-3)/2}}{\lambda_{\min}}\varepsilon + o(\varepsilon).$$

□

Remark C.13. In addition to (14), $\boldsymbol{M}$ can also be decomposed into

$$\boldsymbol{M} = \sum_{i=1}^r \alpha_i \boldsymbol{u}_i^{\otimes 2} + \boldsymbol{E}', \quad \text{where} \quad \|\boldsymbol{E}'\|_\sigma \leq \frac{2d^{(k-3)/2}}{\lambda_{\min}}\varepsilon + o(\varepsilon),$$

where $\boldsymbol{E}'$ satisfies

$$\langle \boldsymbol{E}', \boldsymbol{u}_i^{\otimes 2}\rangle = 0 \quad \text{for all } i \in [r].$$

To see this, rewrite (14) as

$$\boldsymbol{M} = \sum_{i=1}^{r} \alpha_i \boldsymbol{u}_i^{\otimes 2} + \boldsymbol{E} = \sum_{i=1}^{r} \alpha_i \boldsymbol{u}_i^{\otimes 2} + \sum_{i=1}^{r} \langle \boldsymbol{E}, \boldsymbol{u}_i^{\otimes 2} \rangle \boldsymbol{u}_i^{\otimes 2} + \boldsymbol{E} - \sum_{i=1}^{r} \langle \boldsymbol{E}, \boldsymbol{u}_i^{\otimes 2} \rangle \boldsymbol{u}_i^{\otimes 2}$$
$$= \underbrace{\sum_{i=1}^{r} \left(\alpha_i + \langle \boldsymbol{E}, \boldsymbol{u}_i^{\otimes 2} \rangle\right) \boldsymbol{u}_i^{\otimes 2}}_{\in \mathcal{LS}_0} + \underbrace{\boldsymbol{E} - \sum_{i=1}^{r} \langle \boldsymbol{E}, \boldsymbol{u}_i^{\otimes 2} \rangle \boldsymbol{u}_i^{\otimes 2}}_{=:\boldsymbol{E}'}.$$

By construction, $\boldsymbol{E}'$ satisfies

$$\begin{aligned}
\langle \boldsymbol{E}', \boldsymbol{u}_i^{\otimes 2} \rangle &= \langle \boldsymbol{E} - \sum_{j=1}^{r} \langle \boldsymbol{E}, \boldsymbol{u}_j^{\otimes 2} \rangle \boldsymbol{u}_j^{\otimes 2},\ \boldsymbol{u}_i^{\otimes 2} \rangle \\
&= \langle \boldsymbol{E}, \boldsymbol{u}_i^{\otimes 2} \rangle - \sum_{j=1}^{r} \langle \boldsymbol{E}, \boldsymbol{u}_j^{\otimes 2} \rangle \langle \boldsymbol{u}_j^{\otimes 2}, \boldsymbol{u}_i^{\otimes 2} \rangle \\
&= \langle \boldsymbol{E}, \boldsymbol{u}_i^{\otimes 2} \rangle - \sum_{j=1}^{r} \langle \boldsymbol{E}, \boldsymbol{u}_j^{\otimes 2} \rangle \delta_{ij} \\
&= 0.
\end{aligned}$$

Moreover,

$$\begin{aligned}
\|\boldsymbol{E}'\|_\sigma &\le \|\boldsymbol{E}\|_\sigma + \left\| \sum_{i=1}^{r} \langle \boldsymbol{E}, \boldsymbol{u}_i^{\otimes 2} \rangle \boldsymbol{u}_i^{\otimes 2} \right\|_\sigma \\
&\le \|\boldsymbol{E}\|_\sigma + \max_i |\langle \boldsymbol{E}, \boldsymbol{u}_i^{\otimes 2} \rangle| \\
&\le 2 \|\boldsymbol{E}\|_\sigma \\
&\le \frac{2d^{(k-3)/2}}{\lambda_{\min}} \varepsilon + o(\varepsilon),
\end{aligned}$$

where the first line follows from the triangle inequality and the second lines follows from the orthogonality of $\{\boldsymbol{u}_i\}_{i\in[r]}$.

Corollary C.14. *Under Assumption 4.1,*

$$\max_{\boldsymbol{M} \in \mathcal{LS}, \|\boldsymbol{M}\|_F = 1} \min_{\boldsymbol{M}^* \in \mathcal{LS}_0} \|\boldsymbol{M} - \boldsymbol{M}^*\|_\sigma \le \frac{1.13}{c_0},$$

which is ≤ 0.12 *for* $c_0 \ge 10$.

Proof. By Corollary C.12, the right-hand side of (22) has the following upper bound,

$$\frac{d^{(k-3)/2}\varepsilon}{\lambda_{\min}} \left[\left\| \boldsymbol{b}_{\boldsymbol{M}}|_{\mathcal{RS}_0} \right\|_\sigma - \frac{d^{(k-2)/2}\varepsilon}{\lambda_{\min}} \right]^{-1} \le \frac{1}{\sqrt{d} c_0} \left[1 - \frac{1}{(c_0-1)^2} - \frac{1}{c_0} \right]^{-1} \le \frac{1.13}{c_0} \le 0.12.$$

The claim then follows from the same argument as in the proof of Theorem 4.4. □

Corollary C.15. *Suppose* $c_0 \ge 10$ *in Assumption 4.1. In the notation of* (14)*, we have*

$$\max_{i\in[r]} |\alpha_i| \le 1 + \frac{1.13}{c_0} \le 1.12.$$

Proof. By the triangle inequality and Corollary C.14,

$$\max_{i\in[r]} |\alpha_i| \le \sqrt{\sum_{i=1}^{r} |\alpha_i|^2} = \|\boldsymbol{M} - \boldsymbol{E}\|_F \le \|\boldsymbol{M}\|_F + \|\boldsymbol{E}\|_F \le 1 + \frac{1.13}{c_0} = 1.12.$$

□

C.4 Perturbation Bounds

C.4.1 Proof of Lemma 4.5

Proof of Lemma 4.5. We prove by construction. Define $\boldsymbol{M}_i = \boldsymbol{u}_i^{\otimes 2} \in \mathcal{LS}_0$ for $i \in [r]$, and project $\boldsymbol{M}_i$ onto the space $\mathcal{LS}$,

$$\boldsymbol{M}_i = \boldsymbol{M}_i\big|_{\mathcal{LS}} + \boldsymbol{M}_i\big|_{\mathcal{LS}^\perp}, \tag{23}$$

where $\boldsymbol{M}_i\big|_{\mathcal{LS}}$ and $\boldsymbol{M}_i\big|_{\mathcal{LS}^\perp}$ denote the projections of $\boldsymbol{M}_i \in \mathcal{LS}_0$ onto the vector space $\mathcal{LS}$ and $\mathcal{LS}^\perp$, respectively. We seek to show that the set of matrices $\big\{\boldsymbol{M}_i\big|_{\mathcal{LS}} : i \in [r]\big\}$ satisfies

$$\frac{\big\|\boldsymbol{M}_i\big|_{\mathcal{LS}}\big\|_\sigma}{\big\|\boldsymbol{M}_i\big|_{\mathcal{LS}}\big\|_F} \geq 1 - \frac{d^{(k-2)/2}}{\lambda_{\min}}\varepsilon + o(\varepsilon), \quad \text{for all } i \in [r]. \tag{24}$$

Applying the subadditivity of spectral norm to (23) gives

$$\begin{aligned} \big\|\boldsymbol{M}_i\big|_{\mathcal{LS}}\big\|_\sigma &\geq \|\boldsymbol{M}_i\|_\sigma - \big\|\boldsymbol{M}_i\big|_{\mathcal{LS}^\perp}\big\|_\sigma \\ &\geq 1 - \big\|\boldsymbol{M}_i\big|_{\mathcal{LS}^\perp}\big\|_F = 1 - \sin\Theta(\boldsymbol{M}_i, \mathcal{LS})\,\|\boldsymbol{M}_i\|_F \\ &\geq 1 - \sin\Theta(\mathcal{LS}_0, \mathcal{LS}), \end{aligned} \tag{25}$$

where the second line comes from $\|\boldsymbol{M}_i\|_\sigma = \|\boldsymbol{M}_i\|_F = 1$, $\big\|\boldsymbol{M}_i\big|_{\mathcal{LS}^\perp}\big\|_\sigma \leq \big\|\boldsymbol{M}_i\big|_{\mathcal{LS}^\perp}\big\|_F$, and the last line comes from Proposition C.3. By following the same line of argument in Lemma C.11, we have

$$\sin\Theta\left(\mathcal{LS}_0, \mathcal{LS}\right) \leq \frac{\big\|\mathcal{E}_{(12)(3\dots k)}\big\|_\sigma}{\lambda_{\min} - \big\|\mathcal{E}_{(12)(3\dots k)}\big\|_\sigma} \leq \frac{d^{(k-2)/2}}{\lambda_{\min}}\varepsilon + o(\varepsilon). \tag{26}$$

Combining (25) and (26) leads to

$$\big\|\boldsymbol{M}_i\big|_{\mathcal{LS}}\big\|_\sigma \geq 1 - \frac{d^{(k-2)/2}}{\lambda_{\min}}\varepsilon + o(\varepsilon).$$

By construction, $\big\|\boldsymbol{M}_i\big|_{\mathcal{LS}}\big\|_F \leq \|\boldsymbol{M}_i\|_F = 1$, and therefore (24) is proved. Note that $\boldsymbol{M}_i\big|_{\mathcal{LS}} \in \mathcal{LS}$ for all $i \in [r]$. Hence,

$$\max_{\boldsymbol{M} \in \mathcal{LS}} \frac{\|\boldsymbol{M}\|_\sigma}{\|\boldsymbol{M}\|_F} \geq \frac{\big\|\boldsymbol{M}_i\big|_{\mathcal{LS}}\big\|_\sigma}{\big\|\boldsymbol{M}_i\big|_{\mathcal{LS}}\big\|_F} \geq 1 - \frac{d^{(k-2)/2}}{\lambda_{\min}}\varepsilon + o(\varepsilon). \tag{27}$$

The conclusion then follows by the equivalence

$$\max_{\boldsymbol{M} \in \mathcal{LS}} \frac{\|\boldsymbol{M}\|_\sigma}{\|\boldsymbol{M}\|_F} = \max_{\boldsymbol{M} \in \mathcal{LS}, \|\boldsymbol{M}\|_F = 1} \|\boldsymbol{M}\|_\sigma.$$

□

Remark C.16. The above proof reveals that there are at least r elements $\boldsymbol{M}_i\big|_{\mathcal{LS}}$ in $\mathcal{LS}$ that satisfy the right-hand side of (27). These r elements are linearly independent, and in fact, $\{\boldsymbol{M}_i\big|_{\mathcal{LS}}\}_{i\in[r]}$ are approximately orthogonal to each other. To see this, we bound $\cos\Theta(\boldsymbol{M}_i\big|_{\mathcal{LS}}, \boldsymbol{M}_j\big|_{\mathcal{LS}})$ for all $i, j \in [r]$, with $i \neq j$. Recall that $\{\boldsymbol{M} \overset{\text{def}}{=} \boldsymbol{u}_i^{\otimes 2}\}_{i\in[r]}$ is a set of mutually orthogonal vectors in $\mathcal{LS}_0$. Then for all $i \neq j$,

$$\begin{aligned} 0 = \langle \mathbf{M}_i, \mathbf{M}_j \rangle &= \langle \mathbf{M}_i\big|_{\mathcal{LS}} + \mathbf{M}_i\big|_{\mathcal{LS}^\perp}, \mathbf{M}_j\big|_{\mathcal{LS}} + \mathbf{M}_j\big|_{\mathcal{LS}^\perp} \rangle \\ &= \langle \mathbf{M}_i\big|_{\mathcal{LS}}, \mathbf{M}_j\big|_{\mathcal{LS}} \rangle + \langle \mathbf{M}_i\big|_{\mathcal{LS}^\perp}, \mathbf{M}_j\big|_{\mathcal{LS}^\perp} \rangle, \end{aligned} \tag{28}$$

which implies $\langle \mathbf{M}_i\big|_{\mathcal{LS}}, \mathbf{M}_j\big|_{\mathcal{LS}} \rangle = -\langle \mathbf{M}_i\big|_{\mathcal{LS}^\perp}, \mathbf{M}_j\big|_{\mathcal{LS}^\perp} \rangle$. Hence,

$$\big|\cos\Theta(\mathbf{M}_i\big|_{\mathcal{LS}}, \mathbf{M}_j\big|_{\mathcal{LS}})\big| = \frac{\big|\langle \mathbf{M}_i\big|_{\mathcal{LS}}, \mathbf{M}_j\big|_{\mathcal{LS}} \rangle\big|}{\big\|\mathbf{M}_i\big|_{\mathcal{LS}}\big\|_F \big\|\mathbf{M}_j\big|_{\mathcal{LS}}\big\|_F}$$

$$
\begin{aligned}
&= \frac{\left|\langle \mathbf{M}_i|_{\mathcal{LS}^\perp}, \mathbf{M}_j|_{\mathcal{LS}^\perp}\rangle\right|}{\left\|\mathbf{M}_i|_{\mathcal{LS}}\right\|_F \left\|\mathbf{M}_j|_{\mathcal{LS}}\right\|_F} \\
&\le \frac{\left\|\mathbf{M}_i|_{\mathcal{LS}^\perp}\right\|_F}{\left\|\mathbf{M}_i|_{\mathcal{LS}}\right\|_F} \times \frac{\left\|\mathbf{M}_j|_{\mathcal{LS}^\perp}\right\|_F}{\left\|\mathbf{M}_j|_{\mathcal{LS}}\right\|_F} \\
&\le \tan^2 \Theta(\mathcal{LS}_0, \mathcal{LS}),
\end{aligned}
$$

where the second line comes from (28), the third line comes from Cauchy-Schwarz inequality and the last line uses the fact that $\boldsymbol{M}_i, \boldsymbol{M}_j \in \mathcal{LS}_0$. Following the similar argument as in Corollary C.12 (in particular, the last inequality in (13)), we have $|\sin\Theta(\mathcal{LS}_0, \mathcal{LS})| \le \frac{1}{c_0-1} \le 0.12$ under the assumption $c_0 \ge 10$. Thus,

$$
|\cos\Theta(\mathbf{M}_i|_{\mathcal{LS}}, \mathbf{M}_j|_{\mathcal{LS}})| \le \tan^2\Theta(\mathcal{LS}_0, \mathcal{LS}) \le 0.015.
$$

This implies $89.2^\circ \le \Theta(\boldsymbol{M}_i|_{\mathcal{LS}}, \boldsymbol{M}_j|_{\mathcal{LS}}) \le 90.8^\circ$; that is, $\left\{\boldsymbol{M}_i|_{\mathcal{LS}}\right\}_{i\in[r]}$ are approximately orthogonal to each other.

Corollary C.17. *Suppose $c_0 \ge 10$ in Assumption 4.1. Then*

$$
\max_{\boldsymbol{M}\in\mathcal{LS}, \|\boldsymbol{M}\|_F=1} \|\boldsymbol{M}\|_\sigma \ge 1 - \frac{1}{c_0-1} \ge 0.88.
$$

Proof. As seen in Corollary C.12,

$$
\sin\Theta(\mathcal{LS}_0, \mathcal{LS}) \le \frac{\left\|\mathcal{E}_{(12)(3\ldots k)}\right\|_\sigma}{\lambda_{\min} - \left\|\mathcal{E}_{(12)(3\ldots k)}\right\|_\sigma} \le \frac{1}{c_0-1}.
$$

Combining this with (25) and (26) gives

$$
\left\|\boldsymbol{M}_i|_{\mathcal{LS}}\right\|_\sigma \ge 1 - \sin\Theta(\mathcal{LS}_0, \mathcal{LS}) \ge 1 - \frac{1}{c_0-1} \ge 0.88.
$$

The remaining argument is exactly the same as the above proof of Lemma 4.5. □

C.4.2 Proof of Lemma 4.6

Proof of Lemma 4.6. Because of the symmetry of $\widetilde{\mathcal{T}}$ and Lemma C.10, $\widehat{\boldsymbol{M}}_1$ must be a symmetric matrix. Now let $\widehat{\boldsymbol{M}}_1 = \sum_{i=1}^d \gamma_i \boldsymbol{x}_i \boldsymbol{x}_i^T$ denote the eigen-decomposition of $\widehat{\boldsymbol{M}}_1$, where γ_i is sorted in decreasing order and $\boldsymbol{x}_i \in \mathbb{R}^d$ is the eigenvector corresponding to γ_i for all $i \in [d]$. Without loss of generality, we assume $\gamma_1 > 0$. By construction, $\widehat{\boldsymbol{M}}_1 = \arg\max_{\boldsymbol{M}\in\mathcal{LS}, \|\boldsymbol{M}\|_F=1} \|\boldsymbol{M}\|_\sigma$. By Lemma 4.5,

$$
\gamma_1 = \left\|\widehat{\boldsymbol{M}}_1\right\|_\sigma \ge 1 - \frac{d^{(k-2)/2}}{\lambda_{\min}}\varepsilon + o(\varepsilon).
$$

Since $\sum_i \gamma_i^2 = \left\|\widehat{\boldsymbol{M}}_1\right\|_F^2 = 1$, $|\gamma_2| \le \left(1-\gamma_1^2\right)^{1/2} \le \frac{\sqrt{2}d^{(k-2)/4}}{\sqrt{\lambda_{\min}}}\sqrt{\varepsilon} + o(\sqrt{\varepsilon})$. Define $\Delta := \min\{\gamma_1, \gamma_1 - \gamma_2\}$. Then,

$$
\Delta \ge \gamma_1 - |\gamma_2| \ge 1 - \frac{\sqrt{2}d^{(k-2)/4}}{\sqrt{\lambda_{\min}}}\sqrt{\varepsilon} + o(\sqrt{\varepsilon}).
$$

Under the assumption $c_0 \ge 10$, $\gamma_1 \ge 0.88$ by Corollary C.17. Hence, $\Delta \ge \gamma_1 - |\gamma_2| \ge 0.88 - \sqrt{1-0.88^2} \approx 0.41 > 0$.

By Theorem 4.4, there exists $\boldsymbol{M}^* = \sum_{i=1}^r \alpha_i \boldsymbol{u}_i^{\otimes 2} \in \mathcal{LS}_0$ such that

$$
\left\|\widehat{\boldsymbol{M}}_1 - \boldsymbol{M}^*\right\|_\sigma \le \frac{d^{(k-3)/2}}{\lambda_{\min}}\varepsilon + o(\varepsilon).
$$

Without loss of generality, suppose the dominant eigenvector of $\boldsymbol{M}^*$ is $\boldsymbol{u}_1$. Following the notation of Corollary C.5, we set $\boldsymbol{B} = \widehat{\boldsymbol{M}}_1$, $\widetilde{\boldsymbol{B}} = \boldsymbol{M}^*$, $\boldsymbol{E} = \widehat{\boldsymbol{M}}_1 - \boldsymbol{M}^*$, $\boldsymbol{\Sigma}_1 = \{\gamma_1\}$ and $\boldsymbol{\Sigma}_2 = \text{diag}\{\gamma_2, \dots, \gamma_d\}$. From Corollary C.14, $\|\boldsymbol{E}\|_\sigma \leq 0.12$. Combining this with earlier calculation, we have $\Delta - \|\boldsymbol{E}\|_\sigma \geq 0.41 - 0.12 = 0.29 > 0$. Hence, the condition in Corollary C.5 holds.

Applying Corollary C.5 to the specified setting yields

$$|\sin\Theta(\widehat{\boldsymbol{u}}_1, \boldsymbol{u}_1)| \leq \frac{\|\boldsymbol{E}\|_\sigma}{\Delta - \|\boldsymbol{E}\|_\sigma} \leq \frac{d^{(k-3)/2}}{\lambda_{\min}}\varepsilon \left[1 - \frac{\sqrt{2}d^{(k-2)/4}}{\sqrt{\lambda_{\min}}}\sqrt{\varepsilon} + o(\sqrt{\varepsilon})\right]^{-1} = \frac{d^{(k-3)/2}}{\lambda_{\min}}\varepsilon + o(\varepsilon). \tag{29}$$

To bound $\text{Loss}(\widehat{\boldsymbol{u}}_1, \boldsymbol{u}_1)$, we notice that

$$\text{Loss}(\widehat{\boldsymbol{u}}_1, \boldsymbol{u}_1) = [2 - 2\,|\cos\Theta(\widehat{\boldsymbol{u}}_1, \boldsymbol{u}_1)|]^{1/2} = \left[2 - 2\sqrt{1 - \sin^2\Theta(\widehat{\boldsymbol{u}}_1, \boldsymbol{u}_1)}\right]^{1/2}.$$

By Taylor expansion and (29), we conclude

$$\text{Loss}(\widehat{\boldsymbol{u}}_1, \boldsymbol{u}_1) \leq \frac{d^{(k-3)/2}}{\lambda_{\min}}\varepsilon + o(\varepsilon).$$

□

Corollary C.18. *Under Assumption 4.1,*

$$\text{Loss}(\widehat{\boldsymbol{u}}_1, \boldsymbol{u}_1) \leq \frac{5}{c_0},$$

which is ≤ 0.5 *for* $c_0 \geq 10$.

Proof. In the proof of Lemma 4.6, we have shown that $\Delta - \|\boldsymbol{E}\|_\sigma \geq 0.29$. By Corollary C.14, $\|\boldsymbol{E}\|_\sigma \leq 1.13/c_0$. Therefore, (29) has the following upper bound,

$$|\sin\Theta(\widehat{\boldsymbol{u}}_1, \boldsymbol{u}_1)| \leq \frac{\|\boldsymbol{E}\|_\sigma}{\Delta - \|\boldsymbol{E}\|_\sigma} \leq \frac{4}{c_0}.$$

Following the same argument as in the proof of Lemma 4.6, we obtain

$$\text{Loss}(\widehat{\boldsymbol{u}}_1, \boldsymbol{u}_1) = [2 - 2|\cos\Theta(\widehat{\boldsymbol{u}}_1, \boldsymbol{u}_1)|]^{1/2} \leq \frac{5}{c_0} \leq 0.5.$$

□

C.4.3 Proof of Lemma 4.7

Proof of Lemma 4.7. For clarity, we use $\widehat{\boldsymbol{M}}_1$ and $\widehat{\boldsymbol{u}}_1$ to denote the estimators in line 5 of Algorithm 1, and use $\widehat{\boldsymbol{M}}_1^*$ and $\widehat{\boldsymbol{u}}_1^*$ to denote the estimators in line 6 of Algorithm 1. Namely,

$$\widehat{\boldsymbol{M}}_1^* = \widetilde{\mathcal{T}}(\boldsymbol{I}, \boldsymbol{I}, \widehat{\boldsymbol{u}}_1, \dots, \widehat{\boldsymbol{u}}_1), \quad \text{and} \quad \widehat{\boldsymbol{u}}_1^* = \arg\max_{\boldsymbol{x} \in \boldsymbol{S}^{d-1}} |\boldsymbol{x}^T \widehat{\boldsymbol{M}}_1^* \boldsymbol{x}|.$$

By construction, the perturbation model of $\widetilde{\mathcal{T}}$ implies the perturbation model of $\widehat{\boldsymbol{M}}_1^*$,

$$\widehat{\boldsymbol{M}}_1^* = \sum_{i=1}^r \lambda_i \langle \widehat{\boldsymbol{u}}_1, \boldsymbol{u}_i \rangle^{(k-2)} \boldsymbol{u}_i^{\otimes 2} + \mathcal{E}(\boldsymbol{I}, \boldsymbol{I}, \widehat{\boldsymbol{u}}_1, \dots, \widehat{\boldsymbol{u}}_1),$$

where $\|\mathcal{E}(\boldsymbol{I}, \boldsymbol{I}, \widehat{\boldsymbol{u}}_1, \dots, \widehat{\boldsymbol{u}}_1)\|_\sigma \leq \|\mathcal{E}\|_\sigma \leq \varepsilon$.

Without loss of generality, assume $\widehat{\boldsymbol{u}}_1$ is the estimator of $\boldsymbol{u}_1$ and $\langle \widehat{\boldsymbol{u}}_1, \boldsymbol{u}_1 \rangle > 0$; otherwise, we take $-\widehat{\boldsymbol{u}}_1$ to be the estimator. Let $\eta_i := \lambda_i \langle \widehat{\boldsymbol{u}}_1, \boldsymbol{u}_i \rangle^{(k-2)}$ for all $i \in [r]$. In the context of Corollary C.5, we set $\boldsymbol{B} =$

$\sum_{i\in[r]} \eta_i \boldsymbol{u}_i^{\otimes 2}$, $\widetilde{\boldsymbol{B}} = \widehat{\boldsymbol{M}}^*$, $\boldsymbol{E} = \widetilde{\boldsymbol{B}} - \boldsymbol{B}$, $\boldsymbol{\Sigma}_1 = \{\eta_1\}$, $\boldsymbol{\Sigma}_2 = \mathrm{diag}\{\eta_2, \ldots, \eta_r\}$, and $\Delta = \min\{\eta_1, \eta_1 - \max_{i\neq 1} \eta_i\}$. Then,

$$\Delta \geq \eta_1 - \max_{i\neq 1} |\eta_i| = \lambda_1 \langle \widehat{\boldsymbol{u}}_1, \boldsymbol{u}_1\rangle^{(k-2)} - \max_{i\neq 1} |\lambda_i \langle \widehat{\boldsymbol{u}}_1, \boldsymbol{u}_1\rangle|^{(k-2)}. \tag{30}$$

Note that $\|\boldsymbol{E}\|_\sigma \leq \|\mathcal{E}\|_\sigma \leq \varepsilon$. In order to apply Corollary C.5, we seek to show $\Delta > \varepsilon$.

By Definition 4.3, we have

$$\langle \widehat{\boldsymbol{u}}_1, \boldsymbol{u}_1\rangle = \cos\Theta(\widehat{\boldsymbol{u}}_1, \boldsymbol{u}_1) = 1 - \frac{1}{2}\,\mathrm{Loss}^2(\widehat{\boldsymbol{u}}_1, \boldsymbol{u}_1), \tag{31}$$

and by the orthogonality of $\{\boldsymbol{u}_i\}_{i\in[r]}$,

$$|\langle \widehat{\boldsymbol{u}}_1, \boldsymbol{u}_i\rangle|^2 \leq \sum_{j=2}^{r} |\langle \widehat{\boldsymbol{u}}_1, \boldsymbol{u}_j\rangle|^2 \leq 1 - \cos^2\Theta(\widehat{\boldsymbol{u}}_1, \boldsymbol{u}_1) = \mathrm{Loss}^2(\widehat{\boldsymbol{u}}_1, \boldsymbol{u}_1)\left[1 - \frac{1}{4}\,\mathrm{Loss}^2(\widehat{\boldsymbol{u}}_1, \boldsymbol{u}_1)\right], \tag{32}$$

for all $i = 2, \ldots, r$.

Combining (31), (32), $0 \leq \mathrm{Loss}(\widehat{\boldsymbol{u}}_1, \boldsymbol{u}_1) \leq 1/2$ (by Corollary C.18), and the fact that $(1-x)^{(k-2)} \geq 1-(k-2)x$ for all $0 \leq x \leq 1$ and $k \geq 3$, we further have

$$\langle \widehat{\boldsymbol{u}}_1, \boldsymbol{u}_1\rangle^{(k-2)} = \left[1 - \frac{1}{2}\,\mathrm{Loss}^2(\widehat{\boldsymbol{u}}_1, \boldsymbol{u}_1)\right]^{(k-2)} \geq 1 - \frac{k-2}{2}\,\mathrm{Loss}^2(\widehat{\boldsymbol{u}}_1, \boldsymbol{u}_1) \geq 1 - \frac{k-2}{4}\,\mathrm{Loss}(\widehat{\boldsymbol{u}}_1, \boldsymbol{u}_1), \tag{33}$$

and

$$|\langle \widehat{\boldsymbol{u}}_1, \boldsymbol{u}_i\rangle|^{(k-2)} \leq \left[\mathrm{Loss}^2(\widehat{\boldsymbol{u}}_1, \boldsymbol{u}_1)\right]^{(k-2)/2} = \mathrm{Loss}^{k-2}(\widehat{\boldsymbol{u}}_1, \boldsymbol{u}_1) \leq \mathrm{Loss}(\widehat{\boldsymbol{u}}_1, \boldsymbol{u}_1), \tag{34}$$

for all $i = 2, \ldots, r$. Putting (33) and (34) back in (30), we obtain

$$\begin{aligned}\Delta &\geq \lambda_1\left[1 - \frac{k-2}{4}\,\mathrm{Loss}(\widehat{\boldsymbol{u}}_1, \boldsymbol{u}_1)\right] - \lambda_{\max}\,\mathrm{Loss}(\widehat{\boldsymbol{u}}_1, \boldsymbol{u}_1)\\ &\geq \lambda_1\left[1 - \left(\frac{k-2}{4} + \frac{\lambda_{\max}}{\lambda_{\min}}\right)\mathrm{Loss}(\widehat{\boldsymbol{u}}_1, \boldsymbol{u}_1)\right].\end{aligned}$$

By Corollary C.18, $\mathrm{Loss}(\widehat{\boldsymbol{u}}_1, \boldsymbol{u}_1) \leq 5/c_0$. Write $c := \frac{k-2}{4} + \frac{\lambda_{\max}}{\lambda_{\min}}$. Under the assumption $c_0 \geq \max\{10, 6c\}$, we have $\Delta \geq \lambda_1/6$ and hence

$$\Delta - \varepsilon \geq \frac{\lambda_1}{6} - \frac{\lambda_{\min}}{c_0 d^{(k-2)/2}} > \frac{\lambda_1}{6} - \frac{\lambda_{\min}}{10} > 0.$$

This implies that the condition in Corollary C.5 holds. Now applying Corollary C.5 to the specified setting gives

$$\begin{aligned}|\sin\Theta(\widehat{\boldsymbol{u}}_1, \boldsymbol{u}_1)| &\leq \frac{\varepsilon}{\Delta - \varepsilon}\\ &\leq \frac{\varepsilon}{\lambda_1}\left[1 - c\,\mathrm{Loss}(\widehat{\boldsymbol{u}}_1, \boldsymbol{u}_1) - \frac{\varepsilon}{\lambda_1}\right]^{-1}\\ &\leq \frac{\varepsilon}{\lambda_1}\left[1 - \frac{c d^{(k-3)/2}\varepsilon}{\lambda_{\min}} - \frac{\varepsilon}{\lambda_1} + o(\varepsilon)\right]^{-1} = \frac{\varepsilon}{\lambda_1} + o(\varepsilon),\end{aligned}$$

where the third line follows from Lemma 4.6. Using the fact that $\mathrm{Loss}(\widehat{\boldsymbol{u}}_1, \boldsymbol{u}_1) = [2 - 2|\cos\Theta(\widehat{\boldsymbol{u}}_1, \boldsymbol{u}_1)|]^{1/2} = \left[2 - 2\sqrt{1 - \sin^2\Theta(\widehat{\boldsymbol{u}}_1, \boldsymbol{u}_1)}\right]^{1/2}$ and Taylor expansion, we conclude

$$\mathrm{Loss}(\widehat{\boldsymbol{u}}_1, \boldsymbol{u}_1) \leq \frac{\varepsilon}{\lambda_1} + o(\varepsilon).$$

To obtain $\mathrm{Loss}(\widehat{\lambda}_1, \lambda_1)$, recall that under the assumption $\langle \widehat{\boldsymbol{u}}_1, \boldsymbol{u}_1 \rangle > 0$, $\mathrm{Loss}(\widehat{\lambda}_1, \lambda_1) = |\widehat{\lambda}_1 - \lambda_1|$. (Otherwise, we need to consider $|\widehat{\lambda}_1 + \lambda_1|$ instead). Observe that by the triangle inequality,

$$\begin{aligned}|\widehat{\lambda}_1 - \lambda_1| &= |\mathcal{T}(\widehat{\boldsymbol{u}}_1, \dots, \widehat{\boldsymbol{u}}_1) - \lambda_1| = \left| \sum_{i=1}^{r} \lambda_i \langle \widehat{\boldsymbol{u}}_1, \boldsymbol{u}_i \rangle^k + \mathcal{E}(\widehat{\boldsymbol{u}}_1, \dots, \widehat{\boldsymbol{u}}_1) - \lambda_1 \right| \\ &\le \lambda_1 \left| 1 - \langle \widehat{\boldsymbol{u}}_1, \boldsymbol{u}_1 \rangle^k \right| + \sum_{i=2}^{r} \lambda_i \left| \langle \widehat{\boldsymbol{u}}_1, \boldsymbol{u}_i \rangle \right|^k + \left| \mathcal{E}(\widehat{\boldsymbol{u}}_1, \dots, \widehat{\boldsymbol{u}}_1) \right|.\end{aligned}$$

Using similar techniques as in (31), (32), (33) and (34), as well as the fact $(1-x)^k \ge 1 - kx$ for all $0 \le x \le 1$ and $k \ge 3$, we conclude

$$\begin{aligned}|\widehat{\lambda}_1 - \lambda_1| &\le \frac{\lambda_1 k}{2} \mathrm{Loss}^2(\widehat{\boldsymbol{u}}_1, \boldsymbol{u}_1) + \lambda_{\max} \mathrm{Loss}^2(\widehat{\boldsymbol{u}}_1, \boldsymbol{u}_1) + \varepsilon \\ &\le \left(\frac{\lambda_1 k}{2} + \lambda_{\max} \right) \left[\frac{\varepsilon}{\lambda_1} + o(\varepsilon) \right]^2 + \varepsilon \\ &= \varepsilon + o(\varepsilon).\end{aligned}$$

□

C.4.4 Proof of Lemma 4.8

Proof of Lemma 4.8. Let $\boldsymbol{M}$ be a d-by-d matrix in the space $\mathcal{LS}(X) \overset{\text{def}}{=} \mathcal{LS} \cap \mathrm{Span}\{\widehat{\boldsymbol{u}}_i^{\otimes 2} : i \in X\}^{\perp}$ and suppose $\boldsymbol{M}$ satisfies $\|\boldsymbol{M}\|_F = 1$. Since $\mathcal{LS}(X) \subset \mathcal{LS}$, from Remark C.13, $\boldsymbol{M}$ can be decomposed into

$$\boldsymbol{M} = \sum_{i=1}^{r} \alpha_i \boldsymbol{u}_i^{\otimes 2} + \boldsymbol{E}, \tag{35}$$

where

$$\langle \boldsymbol{E}, \boldsymbol{u}_i^{\otimes 2} \rangle = 0 \quad \text{for all } i \in [r], \quad \text{and} \quad \|\boldsymbol{E}\|_{\sigma} \le \frac{2d^{(k-3)/2}\varepsilon}{\lambda_{\min}} + o(\varepsilon). \tag{36}$$

By definition, every element in $\mathcal{LS}(X)$ is orthogonal to $\mathrm{Vec}(\widehat{\boldsymbol{u}}_i^{\otimes 2})$ for all $i \in X$. We claim that under this condition, one must have $\alpha_i = o(\varepsilon)$ for all $i \in X$. To show this, we project $\widehat{\boldsymbol{u}}_i$ onto the space $\mathrm{Span}\{\boldsymbol{u}_i\}$ and write

$$\widehat{\boldsymbol{u}}_i = \xi_i \boldsymbol{u}_i + \eta_i \boldsymbol{u}_i^{\perp},$$

where $\xi_i^2 + \eta_i^2 = 1$ and $\boldsymbol{u}_i^{\perp} \in \mathbf{S}^{d-1}$ denotes the normalized (i.e., unit) vector projection of $\widehat{\boldsymbol{u}}_i$ onto the space $\mathrm{Span}\{\boldsymbol{u}_i\}^{\perp}$. Then for all $i \in X$,

$$\begin{aligned}0 &= \langle \boldsymbol{M}, \widehat{\boldsymbol{u}}_i^{\otimes 2} \rangle \qquad (37) \\ &= \left\langle \sum_{j \in [r]} \alpha_j \boldsymbol{u}_j^{\otimes 2} + \boldsymbol{E},\ \left(\xi_i \boldsymbol{u}_i + \eta_i \boldsymbol{u}_i^{\perp} \right)^{\otimes 2} \right\rangle \\ &= \left\langle \alpha_i \boldsymbol{u}_i^{\otimes 2} + \sum_{j \ne i,\ j \in [r]} \alpha_j \boldsymbol{u}_j^{\otimes 2} + \boldsymbol{E},\ \xi_i^2 \boldsymbol{u}_i^{\otimes 2} + 2\xi_i \eta_i \boldsymbol{u}_i \otimes \boldsymbol{u}_i^{\perp} + \eta_i^2 (\boldsymbol{u}_i^{\perp})^{\otimes 2} \right\rangle \\ &= \alpha_i \xi_i^2 + 2\xi_i \eta_i \left\langle \boldsymbol{E}, \boldsymbol{u}_i \otimes \boldsymbol{u}_i^{\perp} \right\rangle + \eta_i^2 \left\langle \sum_{j \ne i,\ j \in [r]} \alpha_j \boldsymbol{u}_j^{\otimes 2} + \boldsymbol{E},\ (\boldsymbol{u}_i^{\perp})^{\otimes 2} \right\rangle,\end{aligned}$$

where the last line uses the fact that $\langle \boldsymbol{E}, \boldsymbol{u}_i^{\otimes 2} \rangle = 0$, $\langle \boldsymbol{u}_i, \boldsymbol{u}_i^{\perp} \rangle = 0$ and $\langle \boldsymbol{u}_i, \boldsymbol{u}_j \rangle = 0$ for all $j \ne i$. By assumption, $\mathrm{Loss}(\widehat{\boldsymbol{u}}_i, \boldsymbol{u}_i) \le 2\varepsilon/\lambda_i + o(\varepsilon)$. This implies $|\eta_i| = |\langle \widehat{\boldsymbol{u}}_i, \boldsymbol{u}_i^{\perp} \rangle| = [1 - \cos^2 \Theta(\widehat{\boldsymbol{u}}_i, \boldsymbol{u}_i)]^{1/2} \le \mathrm{Loss}(\widehat{\boldsymbol{u}}_i, \boldsymbol{u}_i)[1 - \frac{1}{4} \mathrm{Loss}^2(\widehat{\boldsymbol{u}}_i, \boldsymbol{u}_i)]^{1/2} \le \mathrm{Loss}(\widehat{\boldsymbol{u}}_i, \boldsymbol{u}_i) = O(\varepsilon)$, and $|\xi_i| = (1 - \eta_i^2)^{1/2} \ge 1 - O(\varepsilon)$. It then follows from (37) that

$$\xi_i^2 |\alpha_i| = \left| 2\xi_i \eta_i \left\langle \boldsymbol{E}, \boldsymbol{u}_i \otimes \boldsymbol{u}_i^{\perp} \right\rangle + \eta_i^2 \left\langle \sum_{j \ne i,\ j \in [r]} \alpha_j \boldsymbol{u}_j^{\otimes 2} + \boldsymbol{E},\ (\boldsymbol{u}_i^{\perp})^{\otimes 2} \right\rangle \right|$$

$$\leq 2|\xi_i\eta_i|\left|\left\langle \boldsymbol{E}, \boldsymbol{u}_i \otimes \boldsymbol{u}_i^{\perp}\right\rangle\right| + \eta_i^2 \left(\sum_{j\neq i,\ j\in[r]} \left|\alpha_j \left\langle \boldsymbol{u}_j^{\otimes 2},\ (\boldsymbol{u}_i^{\perp})^{\otimes 2}\right\rangle\right| + \left|\left\langle \boldsymbol{E},\ (\boldsymbol{u}_i^{\perp})^{\otimes 2}\right\rangle\right|\right)$$

$$\leq 2|\xi_i\eta_i| \left\|\boldsymbol{E}\right\|_\sigma + \eta_i^2 \left(\sum_{j\neq i,\ j\in[r]} |\alpha_j| + \left\|\boldsymbol{E}\right\|_\sigma\right)$$

$$\leq O(\varepsilon)\left(\frac{2d^{(k-3)/2}}{\lambda_{\min}}\varepsilon + o(\varepsilon)\right) + O(\varepsilon^2)\left(1.12r + \frac{2d^{(k-3)/2}}{\lambda_{\min}}\varepsilon + o(\varepsilon)\right) = o(\varepsilon),$$

where the last line follows from $|\eta_i| \leq O(\varepsilon)$, $|\xi_i| \leq 1$, $\left\|\boldsymbol{E}\right\|_\sigma \leq \frac{2d^{(k-3)/2}\varepsilon}{\lambda_{\min}} + o(\varepsilon)$ (cf.(36)) and $\max_{i\in[r]} |\alpha_i| \leq 1.12$ (cf. Corollary C.15). Therefore, since $|\xi_i| \geq 1 - O(\varepsilon)$, we conclude that $|\alpha_i| = o(\varepsilon)$ for all $i \in X$.

Now write (35) as

$$\boldsymbol{M} = \sum_{i\in[r]\backslash X} \alpha_i \boldsymbol{u}_i^{\otimes 2} + \sum_{i\in X} \alpha_i \boldsymbol{u}_i^{\otimes 2} + \boldsymbol{E},$$

Note that $\sum_{i\in[r]\backslash X} \alpha_i \boldsymbol{u}_i^{\otimes 2} \in \mathcal{LS}_0(X) \stackrel{\text{def}}{=} \text{Span}\{\boldsymbol{u}_i^{\otimes 2} \colon i \in [r]\backslash X\}$. Hence,

$$\min_{\boldsymbol{M}^*\in\mathcal{LS}_0(X)} \left\|\boldsymbol{M} - \boldsymbol{M}^*\right\|_\sigma \leq \left\|\sum_{i\in X} \alpha_i \boldsymbol{u}_i^{\otimes 2} + \boldsymbol{E}\right\|_\sigma \leq \max_{i\in X} |\alpha_i| + \left\|\boldsymbol{E}\right\|_\sigma \leq \frac{2d^{(k-3)/2}\varepsilon}{\lambda_{\min}} + o(\varepsilon).$$

Since the above holds for all $\boldsymbol{M} \in \mathcal{LS}(X)$ that satisfies $\left\|\boldsymbol{M}\right\|_F = 1$, taking maximum over $\boldsymbol{M}$ yields the desired result. □

C.4.5 Proof of Theorem 4.9

We use the following lemma [1] in our proof of Theorem 4.9.

Lemma C.19. *Fix a subset $X \subset [r]$ and assume that $0 \leq \varepsilon \leq \lambda_i/2$ for each $i \in X$. Choose any $\{\widehat{\boldsymbol{u}}_i, \widehat{\lambda}_i\}_{i\in X} \subset \mathbb{R}^d \times \mathbb{R}$ such that*

$$|\lambda_i - \widehat{\lambda}_i| \leq \varepsilon, \quad \left\|\widehat{\boldsymbol{u}}_i\right\|_2 = 1, \quad \text{and} \quad \langle \boldsymbol{u}_i, \widehat{\boldsymbol{u}}_i\rangle \geq 1 - 2(\varepsilon/\lambda_i)^2 > 0,$$

and define tensor $\Delta_i := \lambda_i \boldsymbol{u}_i^{\otimes k} - \widehat{\lambda}_i \widehat{\boldsymbol{u}}_i^{\otimes k}$ for $i \in X$. Pick any unit vector $\boldsymbol{a} = \sum_{i=1}^d a_i \boldsymbol{u}_i$. Then, there exist positive constants $C_1, C_2 > 0$, depending only on k, such that

$$\left\|\sum_{i\in X} \Delta_i \boldsymbol{a}^{\otimes k-1}\right\|_\sigma \leq C_1 \left(\sum_{i\in X} |a_i|^{k-1}\varepsilon\right) + C_2 \left(|X| \left(\frac{\varepsilon}{\lambda_{\min}}\right)^{k-1}\right), \tag{38}$$

where $\Delta_i \boldsymbol{a}^{\otimes k-1} := \Delta_i(\boldsymbol{a}, \ldots, \boldsymbol{a}, \boldsymbol{I}) \in \mathbb{R}^d$.

Proof of Theorem 4.9. We prove the conclusion

$$\text{Loss}(\widehat{\boldsymbol{u}}_i, \boldsymbol{u}_{\pi(i)}) \leq \frac{2\varepsilon}{\lambda_{\pi(i)}} + o(\varepsilon), \quad \text{Loss}(\widehat{\lambda}_i, \lambda_{\pi(i)}) \leq 2\varepsilon + o(\varepsilon), \tag{39}$$

by induction on i. For $i = 1$, the error bound of $\{(\widehat{\boldsymbol{u}}_1, \widehat{\lambda}_1) \in \mathbb{R}^d \times \mathbb{R}\}$ follows readily from Lemmas 4.5–4.7. Now suppose (39) holds for $i \leq s$. Taking $X = [s]$ in Lemma 4.8 yields the deviation of $\mathcal{LS}(X)$ from $\mathcal{LS}_0(X)$,

$$\max_{\boldsymbol{M}\in\mathcal{LS}(X), \left\|\boldsymbol{M}\right\|_F=1} \ \min_{\boldsymbol{M}^*\in\mathcal{LS}_0(X)} \left\|\boldsymbol{M} - \boldsymbol{M}^*\right\|_\sigma \leq \frac{2d^{(k-3)/2}\varepsilon}{\lambda_{\min}} + o(\varepsilon). \tag{40}$$

Applying Theorem 4.4 and Lemmas 4.5–4.7 to $i = s + 1$ with ε replaced by 2ε (because of the additional factor "2" in (40) compared to Theorem 4.4), we obtain

$$\text{Loss}(\widehat{\boldsymbol{u}}_{s+1}, \boldsymbol{u}_{\pi(s+1)}) \leq \frac{2\varepsilon}{\lambda_{\pi(s+1)}} + o(\varepsilon), \quad \text{Loss}(\widehat{\lambda}_{s+1}, \lambda_{\pi(s+1)}) \leq 2\varepsilon + o(\varepsilon).$$

So (39) also holds for $i = s+1$.

It remains to bound the residual tensor $\Delta\widetilde{\mathcal{T}} \stackrel{\text{def}}{=} \widetilde{\mathcal{T}} - \sum_{i\in[r]} \widehat{\lambda}_i \widehat{\boldsymbol{u}}_i^{\otimes k}$. Note that $\operatorname{Loss}(\widehat{\boldsymbol{u}}_i, \boldsymbol{u}_{\pi(i)}) \le 2\varepsilon/\lambda_{\pi(i)} + o(\varepsilon)$ implies $\langle \widehat{\boldsymbol{u}}_i, \boldsymbol{u}_{\pi(i)} \rangle = 1 - \frac{1}{2}\operatorname{Loss}^2(\widehat{\boldsymbol{u}}_i, \boldsymbol{u}_{\pi(i)}) \ge 1 - 2(\varepsilon/\lambda_{\pi(i)})^2 + o(\varepsilon^2)$. When c_0 is sufficiently large (i.e., ε is sufficiently small), $\widehat{\boldsymbol{u}}_i$ is approximately parallel to $\boldsymbol{u}_{\pi(i)}$ and orthogonal to $\boldsymbol{u}_j$ for all $j \ne \pi(i)$. For ease of notation, we renumber the indices and assume $\pi(i) = i$ for all $i \in [r]$. Following the definition of Δ_i in Lemma C.19,

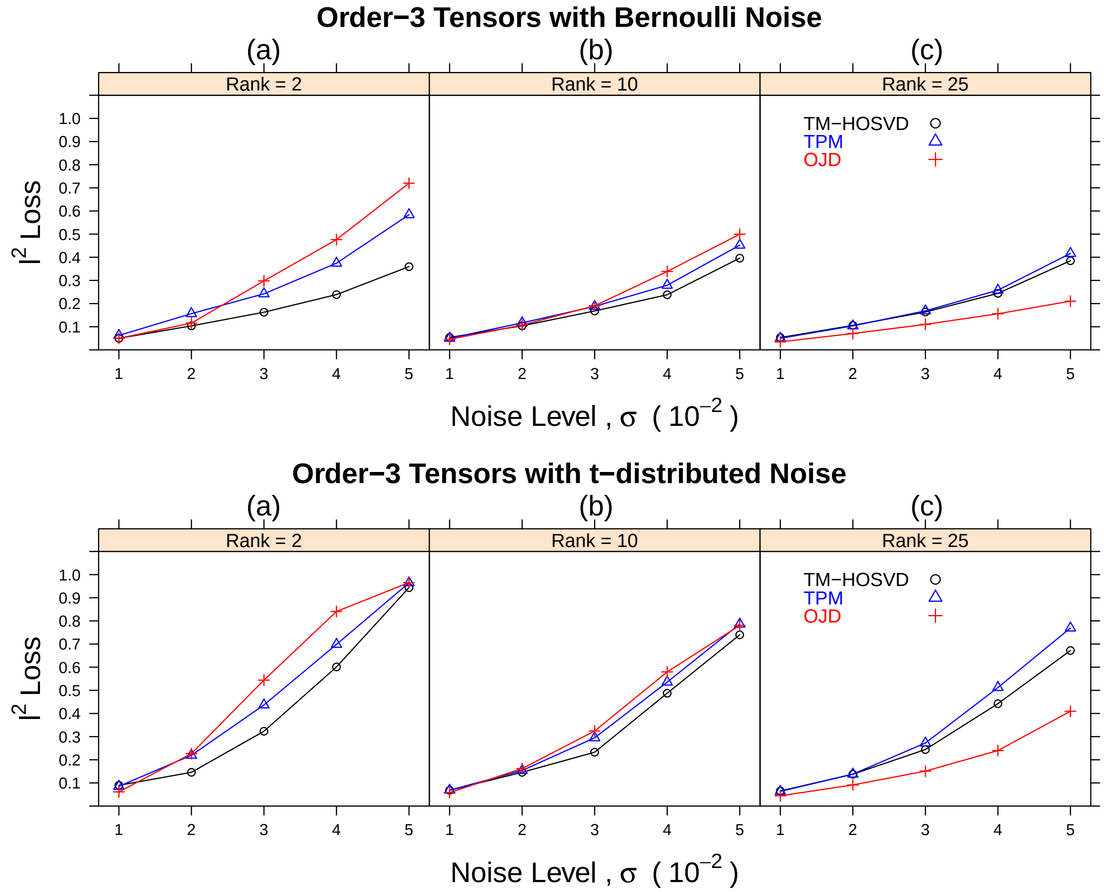
$$\left\|\Delta\widetilde{\mathcal{T}}\right\|_\sigma = \left\|\sum_{i\in[r]} \lambda_i \boldsymbol{u}_i^{\otimes k} + \mathcal{E} - \sum_{i\in[r]} \widehat{\lambda}_i \widehat{\boldsymbol{u}}_i^{\otimes k}\right\|_\sigma = \left\|\sum_{i\in[r]} \Delta_i + \mathcal{E}\right\|_\sigma .$$

Now taking $X = [r]$ in (38) gives

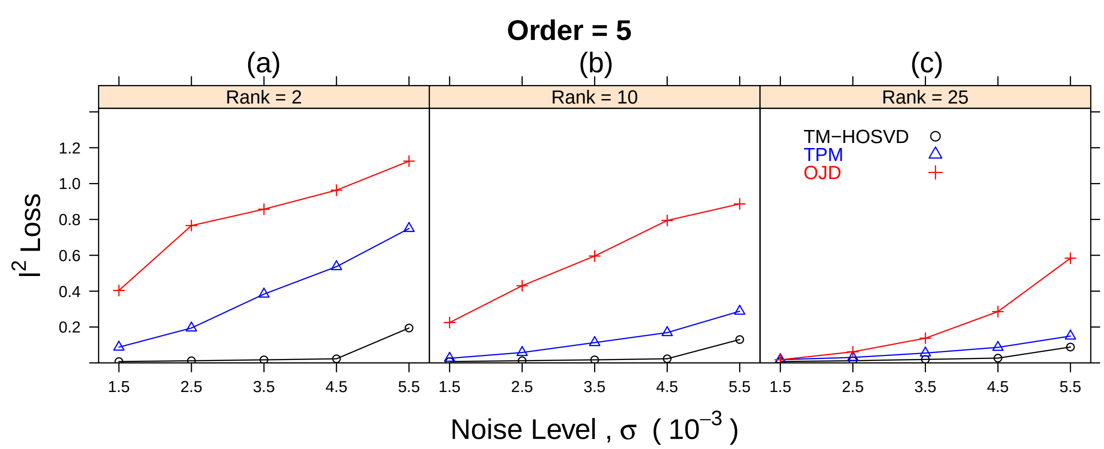
$$\begin{aligned}
\left\|\Delta\widetilde{\mathcal{T}}\right\|_\sigma &\le \max_{\boldsymbol{a}\in\mathbf{S}^{d-1}} \left\|\sum_{i\in[r]} \Delta_i \boldsymbol{a}^{\otimes(k-1)}\right\|_\sigma + \varepsilon \\
&\le \max_{\boldsymbol{a}\in\mathbf{S}^{d-1}} C_1 \sum_{i\in[r]} |a_i|^{k-1}\varepsilon + C_2 r\left(\frac{\varepsilon}{\lambda_{\min}}\right)^{k-1} + \varepsilon \\
&\le \max_{\boldsymbol{a}\in\mathbf{S}^{d-1}} C_1\varepsilon \sum_{i\in[r]} |a_i|^2 + C_2 r\left(\frac{\varepsilon}{\lambda_{\min}}\right)^{2} + \varepsilon \\
&\le C\varepsilon + o(\varepsilon),
\end{aligned}$$

where the third line comes from the fact that $k \ge 3$, $|a_i| \le 1$, and $\varepsilon/\lambda_{\min} \le 1$ from Assumption 4.1. □

D Supplementary Figures and Table



Supplementary Figure S1: Average l^2 Loss for decomposing order-3 nearly SOD tensors with Bernoulli/T-distributed noise, $d = 25$.



Supplementary Figure S2: Average l^2 Loss for decomposing order-5 nearly SOD tensors with Gaussian noise, $d = 25$.

Supplementary Table S1: Runtime for decomposing nearly-SOD tensors with Gaussian noise, $d = 25$.

Order	Rank	Noise Level (σ)	Time (sec.)		
			TM-HOSVD	TPM	OJD
3	2	5×10^{-2}	0.08	0.01	0.13
3	10	5×10^{-2}	0.20	0.03	0.80
3	25	5×10^{-2}	0.47	0.07	0.92
4	2	1.5×10^{-2}	0.13	0.06	0.12
4	10	1.5×10^{-2}	0.29	0.14	1.06
4	25	1.5×10^{-2}	0.57	0.25	1.58
5	2	5.5×10^{-3}	0.25	0.51	0.14
5	10	5.5×10^{-3}	0.45	1.98	1.01
5	25	5.5×10^{-3}	0.87	4.27	2.66

References

[1] Mu, C., Hsu, D., and Goldfarb, D. Successive rank-one approximations for nearly orthogonally decomposable symmetric tensors. *SIAM Journal on Matrix Analysis and Applications 36*, 4 (2015), 1638–1659.

[2] Wang, M., Dao Duc, K., Fischer, J., and Song, Y. S. Operator norm inequalities between tensor unfoldings on the partition lattice. *Linear Algebra and its Applications 520* (2017), 44–66.

[3] Wedin, P.-Å. Perturbation bounds in connection with singular value decomposition. *BIT Numerical Mathematics 12*, 1 (1972), 99–111.

[4] Weyl, H. Inequalities between the two kinds of eigenvalues of a linear transformation. *Proceedings of the National Academy of Sciences 35*, 7 (1949), 408–411.