Fast and Provable Nonconvex Tensor RPCA

Haiquan Qiu [1 2 3] Yao Wang [1] Shaojie Tang [4] Deyu Meng [1 5] Quanming Yao [2]

Abstract

In this paper, we study nonconvex tensor robust principal component analysis (RPCA) based on the t-SVD. We first propose an alternating projection method, i.e., APT, which converges linearly to the ground-truth under the incoherence conditions of tensors. However, as the projection to the low-rank tensor space in APT can be slow, we further propose to speedup such a process by utilizing the property of the tangent space of low-rank. The resulting algorithm, i.e., EAPT, is not only more efficient than APT but also keeps the linear convergence. Compared with existing tensor RPCA works, the proposed method, especially EAPT, is not only more effective due to the recovery guarantee and adaption in the transformed (frequency) domain but also more efficient due to faster convergence rate and lower iteration complexity. These benefits are also empirically verified both on synthetic data, and real applications, e.g., hyperspectral image denoising and video background subtraction.

1. Introduction

A tensor is a multidimensional array that can model the linear and multilinear relationships in the data. Tensor related methods have been widely used in many areas such as recommender systems (Candès & Recht, 2009), computer vision (Zhang et al., 2014), and signal processing (Cichocki et al., 2015). For example, a hyperspectral image with multiple bands can be naturally represented as a three-way tensor with the column, row, and spectral bands; a grayscale video is indexed by two spatial variables and one temporal variable. All these examples are the three-way tensors, which are the focus of this paper.

Analogous to robust PCA (Candès et al., 2011), tensor robust PCA (Huang et al., 2015; Lu et al., 2016; Anandkumar et al., 2016; Lu et al., 2019) attempts to recover a low-rank tensor that best approximates grossly corrupted observations. While there are many works for matrix RPCA (Wright et al., 2009; Netrapalli et al., 2014; Yi et al., 2016; Xu et al., 2010), these matrix techniques cannot be directly adopted for tensors as the tensor rank is more complicated than matrices. Typically, to impose a low-rank structure in the tensor RPCA, CP and Tucker decompositions (Kolda & Bader, 2009) factorize a tensor into low-rank matrices and are used in methods such as RTD (Anandkumar et al., 2016). Besides, Tensor-Train decomposition (Oseledets, 2011), which approximates a higher order tensor with a collection of small three-way tensors, has been applied to tensor RPCA in TTNN (Yang et al., 2020). Furthermore, a convex-optimization-based tensor decomposition approach, called overlapped approach (Gandy et al., 2011; Liu et al., 2012; Tomioka et al., 2010), which penalizes each unfolding matrices from the original tensor using the nuclear norm, is considered in SNN (Huang et al., 2015).

All the above decompositions consider low-rank structure in the time domain. However, many applications (Chang, 1966; Rao et al., 2010) show that the structures in the frequency domain is important as well. For example, both SVD truncation and high-frequency filtering can remove noise from an image. For tensor decomposition, there exists a method called t-SVD (Kilmer & Martin, 2011) that can take advantage of structures in both the time domain and frequency domain. t-SVD first conducts fast Fourier transformation (FFT) on the tube fibers of a tensor. This operation converts the tensor from the time domain to the frequency domain. In this way, the high- and low-frequency information of a tensor is separated, while the low-rank structure of a tensor is preserved in this process. Recently, transformed t-SVD (Kernfeld et al., 2015; Kilmer et al., 2021) extends the FFT in t-SVD to a general unitary transformation, where the main superiority is that the transformed tensor may have lower multi-rank by using suitable unitary transformation. Extensive numerical examples have shown its effectiveness in many applications (Zhang & Ng, 2021; Lu, 2021; Song et al., 2020a).

Because of the aforementioned benefits, there emerges

[1]Xi'an Jiaotong University, Xi'an, China [2]Tsinghua University, Beijing, China [3]4Paradigm Inc., Beijing, China [4]The University of Texas at Dallas, Texas, USA [5]Macau University of Science and Technology, Macau, China. Correspondence to: Yao Wang <yao.s.wang@gmail.com>, Quanming Yao <qyaoaa@connect.ust.hk>.

Proceedings of the 39^{th} International Conference on Machine Learning, Baltimore, Maryland, USA, PMLR 162, 2022.

Table 1. Comparison of various tensor RPCA methods based on (transformed) t-SVD. "transformed": whether base on transformed t-SVD; "✓": the corresponding method has such a property or uses such a technique; "✗": the corresponding method does not have such a property or use such a technique. The computational complexity is calculated on a tensor in $\mathbb{R}^{n\times n\times q}$ with multi-rank $\boldsymbol{r}$ ($s_r = \sum_{i=1}^q r_i$).

t-SVD methods	effectiveness (recovery performance)			efficiency (optimization time)		
	transformed	rank	recovery guarantee	convergence rate	iteration complexity	
					FFT	DCT
TRPCA (Lu et al., 2019)	✗	tubal	✓	$\mathcal{O}(1/\epsilon)$	$\mathcal{O}\left(n^2 q\log q + n^3 q\right)$	—
ETRPCA (Gao et al., 2020)	✗	tubal	✗	✗	$\mathcal{O}\left(n^2 q\log q + n^3 q\right)$	—
T-TRPCA (Lu, 2021)	✓	tubal	✓	$\mathcal{O}(1/\epsilon)$	—	$\mathcal{O}(n^2q^2 + n^3 q)$
APT	✓	multi	✓	$\mathcal{O}\left(\log(1/\epsilon)\right)$	$\mathcal{O}(n^2 q\log q + n^3 q)$	$\mathcal{O}(n^2q^2 + n^3 q)$
EAPT	✓	multi	✓	$\mathcal{O}\left(\log(1/\epsilon)\right)$	$\mathcal{O}(n^2 q\log q + n^2 s_r)$	$\mathcal{O}(n^2q^2 + n^2 s_r)$

works trying to solve tensor RPCA based on t-SVD (Lu et al., 2019; Gao et al., 2020; Lu, 2021). TRPCA (Lu et al., 2019) proposes a new tensor nuclear norm and solves a convex optimization objective for tensor RPCA with alternating direction method of multipliers (ADMM) (Gabay & Mercier, 1976) and possesses statistical recovery guarantee. However, TRPCA penalizes all singular values equally, which does not fully utilize the information in large singular values. So ETRPCA (Gao et al., 2020) introduces weighted tensor Schatten p-norm to make large singular values shrink less. Transformed TRPCA (T-TRPCA) (Lu, 2021) extends TRPCA with a new nuclear norm derived from *transformed* t-SVD. Same as TRPCA, T-TRPCA also has statistical recovery guarantee. All these works, summarized in Table 1, are based on the tubal rank, which cannot comprehensively capture the difference in low- and high-frequency information in transformed tensors. Also, these methods cannot be efficient and effective at the same time.

In this paper, we propose two alternating projection algorithms for tensor RPCA, i.e., APT and EAPT, based on transformed t-SVD. Both our methods have recovery guarantee and linear convergence rate. Moreover, by utilizing properties of the tangent of space of low-rank tensors, EAPT avoids t-SVD on a full-sized tensor. When considering multi-rank, our methods can adaptively keep important information in the frequency domain. The main contributions of our method are as follows.

- We study nonconvex tensor robust PCA based on transformed t-SVD and propose two alternating projection algorithms, i.e., APT and EAPT. Specifically, EAPT is more efficient since it utilizes the tangent space of low-rank tensor to reduce iteration complexity.
- As far as we know, our work is the first to prove the exact recovery guarantee for nonconvex tensor RPCA with t-SVD decomposition. Specially, we show that linear convergence to the ground-truth can be guaranteed under suitable tensor incoherence conditions.
- Experiments on synthetic data and real data applications, e.g., hyperspectral image denoising and video background subtraction, demonstrate both efficiency and effectiveness of our methods.

Notations Scalars and vectors are denoted as lowercase letters and bold lowercase letters respectively, e.g., x and $\mathbf{x}$. Bold capital letters and calligraphic letters denote the matrices and tensors, respectively, e.g., $\mathbf{X}$ and $\mathcal{X}$. The real and complex Euclidean spaces are denoted as $\mathbb{R}$ and $\mathbb{C}$, respectively. Superscript H and T denote conjugate transpose and transpose respectively. $A \lesssim B$ means $A \leq cB$ for some positive number c.

For a three-way tensor $\mathcal{X} \in \mathbb{C}^{m\times n\times q}$, its (i,j,k)-th entry is denoted as $\mathcal{X}_{ijk}$ and the MATLAB notations $\mathcal{X}(i,:,:)$, $\mathcal{X}(:,i,:)$ and $\mathcal{X}(:,:,i)$ are used to denote the i-th horizontal, lateral, and frontal slice of $\mathcal{X}$, respectively. A tube fiber of a three-way tensor is defined as $\mathcal{X}(i,j,:)$ where the first two indices of $\mathcal{X}$ are fixed. The inner product between two tensors $\mathcal{X}$ and $\mathcal{Y}$ in $\mathbb{C}^{m\times n\times q}$ is defined as $\langle \mathcal{X}, \mathcal{Y}\rangle = \sum_{ijt} \mathcal{X}_{ijt}\mathcal{Y}_{ijt}$. We denote the Frobenius norm of a tensor as $\|\mathcal{X}\|_F = \sqrt{\langle \mathcal{X}, \mathcal{X}\rangle}$. The ℓ_1-norm is defined as $\|\mathcal{X}\|_1 = \sum_{ijt} |\mathcal{X}_{ijt}|$ and the infinity norm $\|\mathcal{X}\|_\infty$ is denoted as $\|\mathcal{X}\|_\infty = \max_{ijt} |\mathcal{X}_{ijt}|$. The Euclidean projection operator P is defined as $P_\Omega(\mathcal{X}) = \arg\min_{\mathcal{Y}\in\Omega} \|\mathcal{Y} - \mathcal{X}\|_F$.

2. Preliminaries

In this section, we introduce basic facts about t-SVD (Kernfeld et al., 2015; Kilmer et al., 2021), which will be used in the subsequent sections. Specifically, we focus on three-way tensor in this paper. More related facts about tensor-tensor product are given in the Appendix B.

For any three-way tensor $\mathcal{X} \in \mathbb{C}^{m\times n\times q}$ and any unitary matrix $\Phi \in \mathbb{C}^{q\times q}$, $\hat{\mathcal{X}}_\Phi$ denotes a three-way tensor with each tube fiber multiplied by Φ, i.e.,

$$\hat{\mathcal{X}}_\Phi(i,j,:) = \Phi \cdot \mathcal{X}(i,j,:), i = 1\ldots m, j = 1\ldots n. \quad (1)$$

Also, $\hat{\mathcal{X}}_\Phi$ is often denoted as $\Phi[\mathcal{X}]$ for simplicity. The block-diagonal matrix representation of a three-way tensor

$\mathcal{X}$ defined as

$$\bar{\mathcal{X}} := \begin{bmatrix} \hat{\mathcal{X}}_\Phi(:,:,1) & & \\ & \ddots & \\ & & \hat{\mathcal{X}}_\Phi(:,:,q) \end{bmatrix}.$$

The conjugate transpose of a tensor $\mathcal{X} \in \mathbb{C}^{m\times n\times q}$ is denoted as $\mathcal{X}^H$ with $\overline{\mathcal{X}^H} = \bar{\mathcal{X}}^H$. The spectral norm of $\mathcal{X}$ is defined as $\|\mathcal{X}\|_\Phi = \|\bar{\mathcal{X}}\|$ and denoted as $\|\mathcal{X}\|$ for simplicity.

Similar to the t-product in Fourier domain (Kilmer & Martin, 2011), the result $\mathcal{Z} \in \mathbb{C}^{n_1\times n_3\times q}$ of tensor-tensor product (Kilmer et al., 2021; Kernfeld et al., 2015) between two tensors $\mathcal{X} \in \mathbb{C}^{n_1\times n_2\times q}$ and $\mathcal{Y} \in \mathbb{C}^{n_2\times n_3\times q}$ with any unitary transformation Φ is given by $\mathcal{Z} = \mathcal{X} \diamond_\Phi \mathcal{Y} = \Phi^H\left[\text{fold}\left(\bar{\mathcal{X}}\cdot\bar{\mathcal{Y}}\right)\right]$ where fold $\left(\bar{\mathcal{X}}\right) = \hat{\mathcal{X}}_\Phi = \Phi[\mathcal{X}]$.

Transformed t-SVD and tensor rank are described in the following Definitions 2.1 and 2.2 respectively. Note that t-SVD is a special case of transformed t-SVD by setting $\Phi = \mathcal{F}/\sqrt{q}$ where $\mathcal{F}$ is a DFT matrix.

Definition 2.1 (Kernfeld et al. 2015)**.** For any $\mathcal{X} \in \mathbb{C}^{m\times n\times q}$, the transformed t-SVD is given by $\mathcal{X} = \mathcal{U}\diamond_\Phi\mathcal{S}\diamond_\Phi\mathcal{V}^H$, where $\mathcal{U} \in \mathbb{C}^{m\times m\times q}$, $\mathcal{V} \in \mathbb{C}^{n\times n\times q}$ are unitary tensors with respect to the transformation Φ, and $\mathcal{S} \in \mathbb{C}^{m\times n\times q}$ is a diagonal tensor. And the singular value of $\mathcal{X}$ is defined as $\hat{\mathcal{S}}_\Phi(i,i,c)$, where $1 \le i \le \min(m,n)$ and $1 \le c \le q$.

Definition 2.2 (tensor rank (Kilmer et al., 2021))**.** The multi-rank of a tensor $\mathcal{X} \in \mathbb{C}^{m\times n\times q}$ is a vector $\boldsymbol{r} \in \mathbb{R}^q$, i.e., $\text{rank}_m(\mathcal{X}) = \boldsymbol{r}$, with its i-th entry being the rank of the i-th frontal slice $\hat{\mathbf{X}}_\Phi(:,:,i)$ of $\hat{\mathcal{X}}_\Phi$, i.e., $r_i = \text{rank}(\hat{\mathbf{X}}_\Phi(:,:,i))$. Let $\mathcal{X} = \mathcal{U} \diamond_\Phi \mathcal{S} \diamond_\Phi \mathcal{V}^H$. The tubal rank of $\mathcal{X}$, denoted by $\text{rank}_{tt}(\mathcal{X})$, is defined as the number of nonzero singular value tubes of $\mathcal{S}$, i.e., $\text{rank}_{tt}(\mathcal{X}) = \#\{i|\mathcal{S}(i,i,:) \neq \mathbf{0}\}$.

The relationship between multi-rank and tubal rank is $\text{rank}_{tt}(\mathcal{X}) = \max\{r_i\}$, i.e., tubal rank is equal to be maximum element in multi-rank. Previous works such as TRPCA and T-TRPCA are based on tubal rank while our methods are based on multi-rank. In real applications (Hao et al., 2013; Kilmer et al., 2021), using multi-rank can adaptively utilize the high- and low-frequency information in the frequency (transformed) domain. Because for a tensor converted into frequency domain, the high-frequency slices are more likely to be noise, then applying lower rank for the high-frequency slices can effectively remove the noise on the original tensor.

3. Related Work

Given the observed tensor $\mathcal{D} \in \mathbb{R}^{n\times n\times q}$, tensor RPCA attempts to separate $\mathcal{D}$ into a low-rank tensor part $\mathcal{L}$ and a sparse tensor part $\mathcal{S}$. However, different from matrices, there are many rank definitions for tensors because a tensor can be factorized in many ways. Common tensor decompositions include CP, Tucker (Kolda & Bader, 2009), Tensor-Train (TT) (Oseledets, 2011), overlapped/latent tensor nuclear norm (Gandy et al., 2011; Tomioka et al., 2010), and t-SVD (Kilmer & Martin, 2011; Kilmer et al., 2021). With these, there are many methods (Lu et al., 2019; Lu, 2021; Gao et al., 2020; Cai et al., 2021; Yang et al., 2020; Huang et al., 2015; Driggs et al., 2019; Anandkumar et al., 2016) for tensor RPCA (see Appendix A for details).

Tensor RPCA (Lu et al., 2019; Lu, 2021; Gao et al., 2020) based on tensor-tensor product can utilize information in both time and frequency domains simultaneously. Among them, T-TRPCA (Lu, 2021) constructs two convex surrogates, i.e., $\|\mathcal{L}\|_{\Phi^*}$ and $\|\mathcal{S}\|_1$, of the low-rank and sparse constraints. The optimization objective of T-TRPCA is

$$\min_{\mathcal{L},\mathcal{S}} \|\mathcal{L}\|_{\Phi^*} + \lambda\|\mathcal{S}\|_1, \text{subject to } \mathcal{D} = \mathcal{L} + \mathcal{S}.$$

When $\Phi = \mathcal{F}/\sqrt{q}$, TRPCA (Lu et al., 2019) becomes a special case of T-TRPCA. Both TRPCA and T-TRPCA converge sub-linearly with recovery guarantee. The computational complexities of TRPCA and T-TRPCA are $\mathcal{O}(n^2q\log q + n^3q)$ and $\mathcal{O}(n^2q^2 + n^3q)$ respectively. Different from TRPCA, T-TRPCA replaces the DFT matrix with a DCT matrix, which achieves better performance in real applications (Lu, 2021).

Besides, ETRPCA (Gao et al., 2020) is a nonconvex method based on weighted tensor Schatten p-norm with only algorithmic convergence guarantee. ETRPCA improves TRPCA by making large singular values shrink less while TRPCA and T-TRPCA penalize all singular values equally. The computational complexity of ETRPCA is $\mathcal{O}(n^2q\log q + n^3q)$. However, all these works, summarized in Table 1, use tubal rank that ignores the difference between the high- and low-frequency information in transformed tensor while multi-rank can specify different ranks for the frontal slices at different frequencies.

Another line of tensor RPCA methods is based on alternating projection (Cai et al., 2021; Anandkumar et al., 2016). These works alternatively project the tensor onto the low-rank and sparse spaces. Among them, RTCUR (Cai et al., 2021) is based on a new decomposition called Fiber CUR without a recovery guarantee. Anandkumar et al. 2016 proposes an alternating projection algorithm based on CP decomposition with a recovery guarantee. To better utilize tensor-tensor product and frequency information, we will propose two methods based on alternating projection and transformed t-SVD.

4. Fast Nonconvex Tensor RPCA

Here, we first introduce APT which is an alternating projection algorithm with exact recovery guarantee and linear convergence rate (Section 4.1). Then, EAPT improves the low-rank projection in APT while keeping exact recovery guarantee and linear convergence rate (Section 4.2).

For simplicity, our methods are proposed based on symmetric tensors, i.e., $\bar{\mathcal{X}} = \bar{\mathcal{X}}^H \in \mathbb{C}^{n\times n\times q}$. The extension of our methods to asymmetric tensors is given in the Appendix B. All proofs are presented in Appendix D.

4.1. Alternative Projection with Exact Recovery

To better utilize the information in both time and frequency domains, the rank constraint of our objective is based on multi-rank (see Definition 2.2) because it can better utilize information in frequency domain. Our optimization objective follows as:

$$\min_{\mathcal{L}\in\mathbb{L},\mathcal{S}\in\mathbb{S}} \|\mathcal{D}-\mathcal{L}-\mathcal{S}\|_F^2, \tag{2}$$

$$\text{subject to } \begin{cases}\mathbb{L}\equiv\{\mathcal{X}\,|\,\text{rank}_m(\mathcal{X})\le \boldsymbol{r}\},\\ \mathbb{S}\equiv\{\mathcal{X}\,|\,\|\mathcal{X}\|_0\le K\},\end{cases} \tag{3}$$

where we assume its optimal solution $\mathcal{L}^*$ and $\mathcal{S}^*$ satisfying the incoherence and sparse conditions in Section 5. More importantly, if Φ is a normalized DFT matrix or DCT matrix, we can set multi-rank $\boldsymbol{r}$ as a descending sequence so that our methods can adaptively take advantage of the information in the frequency (transformed) domain which is not possible with TRPCA and T-TRPCA as shown in Table 1.

The basic idea for solving (2) is to alternatively project between the low multi-rank space $\mathbb{L}$ and sparse space $\mathbb{S}$. For example, at k-th iteration, $\mathcal{L}_{k+1}$ is the projection of $\mathcal{D}-\mathcal{S}_k$ onto the low-rank space $\mathbb{L}$. Let $\mathcal{X}=\mathcal{U}\diamond_\Phi\mathcal{S}\diamond_\Phi\mathcal{V}^H$ be the transformed t-SVD of $\mathcal{X}$. With the Eckart-Young theorem (Theorem B.1), the projection onto $\mathbb{L}$ can be conducted with truncated transformed t-SVD, i.e., $\mathcal{L}_{k+1}=P_{\mathbb{L}}(\mathcal{D}-\mathcal{S}_k)$ where $P_{\mathbb{L}}(\mathcal{X})=\Phi^H[\hat{\mathcal{X}}_{\boldsymbol{r}}]$ with

$$\hat{\mathcal{X}}_{\boldsymbol{r}}(:,:,i)=\hat{\mathcal{U}}(:,1:r_i,i)\hat{\mathcal{S}}(1:r_i,1:r_i,i)\hat{\mathcal{V}}^H(:,1:r_i,i). \tag{4}$$

Computing $\mathcal{L}_{k+1}$ is time-consuming which costs $\mathcal{O}(n^2q^2+n^3q)$. A direct implementation obtains $\mathcal{S}_{k+1}$ by projecting $\mathcal{D}-\mathcal{L}_{k+1}$ onto sparse space $\mathbb{S}$. The projection onto $\mathbb{S}$ with $\mathcal{S}_{k+1}=P_{\mathbb{S}}(\mathcal{D}-\mathcal{L}_{k+1})$ is equivalent to keep the top K elements in $\mathcal{D}-\mathcal{L}_{k+1}$. However, top K operation is unstable for the change of the magnitude of sparse tensor $\mathcal{S}$. Thus, this method cannot guarantee exact recovery.

Instead, we pick an appropriate threshold keeping the elements which are likely to be in the support set of $\mathcal{S}^*$. At the k-th iteration, the projection onto $\mathbb{S}$ is $\mathcal{S}_{k+1}=T_{\zeta_{k+1}}(\mathcal{D}-\mathcal{L}_{k+1})$ where

$$T_\zeta(\mathcal{Z})_{ijt}=\begin{cases}\mathcal{Z}_{ijt} & |\mathcal{Z}_{ijt}|>\zeta\\ 0 & \text{otherwise}\end{cases},$$

is the hard thresholding operator, and $\{\zeta_k\}$ is a descending sequence to ensure the convergence and $\zeta_{k+1}=\beta\gamma^k\bar{\sigma}_1(\mathcal{D}-\mathcal{S}_k)$ [1].

The complete procedure of APT is presented in Algorithm 2, which can be divided into two phases: initialization and alternating projection. The first phase of our algorithm is to find a good initialization for alternating projection. The same as (Netrapalli et al., 2014; Cai et al., 2019), the initialization phase (Algorithm 1) constructs an initial guess that can be close to the ground-truth and our recovery theory will give the values of β_0 and β_{init} in Algorithm 1. Next, our method conducts alternating projection. At the k-th iteration, the first step projects $\mathcal{D}-\mathcal{S}_k$ onto $\mathbb{L}$, i.e., $\mathcal{L}_{k+1}=P_{\mathbb{L}}(\mathcal{D}-\mathcal{S}_k)$. The second step projects $\mathcal{D}-\mathcal{L}_{k+1}$ onto $\mathbb{S}$, i.e., $\mathcal{S}_{k+1}=T_{\zeta_{k+1}}(\mathcal{D}-\mathcal{L}_{k+1})$.

Algorithm 1 Initialization

1: $\mathcal{S}_{-1}=T_{\zeta_{-1}}(\mathcal{D})$ where $\zeta_{-1}=\beta_{\text{init}}\cdot\bar{\sigma}_1(\mathcal{D})$
2: $\mathcal{L}_0=P_{\mathbb{L}}(\mathcal{D}-\mathcal{S}_{-1})$
3: $\mathcal{S}_0=T_{\zeta_0}(\mathcal{D}-\mathcal{L}_0)$ where $\zeta_0=\beta_0\cdot\bar{\sigma}_1(\mathcal{D}-\mathcal{S}_{-1})$
4: **Return:** $\mathcal{L}_0$ and $\mathcal{S}_0$

Algorithm 2 APT: Alternating Projection Algorithm for Tensor RPCA

1: Run Algorithm 1 for initialization
2: **for** $k=0$ to $T-1$ **do**
3: $\quad\mathcal{L}_{k+1}=P_{\mathbb{L}}(\mathcal{D}-\mathcal{S}_k)$
4: $\quad\mathcal{S}_{k+1}=T_{\zeta_{k+1}}(\mathcal{D}-\mathcal{L}_{k+1})$ where $\zeta_{k+1}=\beta\gamma^k\bar{\sigma}_1(\mathcal{D}-\mathcal{S}_k)$
5: **end for**
6: **Return:** $\mathcal{L}_T$ and $\mathcal{S}_T$

In Section 5, we can see that APT converges linearly to the ground-truth under suitable tensor incoherence conditions. However, the computational complexity of APT is $\mathcal{O}(n^2q^2+n^3q)$, which mainly comes from the transformed t-SVD on a full-sized tensor. Compared with previous works, APT converges faster but does not reduce iteration complexity (Table 1).

4.2. Speedup with Efficient Projection

For low-rank projection in matrix, SVD on a full-sized matrix can be avoided by projection onto the tangent space of a low-rank matrix (Vandereycken, 2013; Cai et al., 2019; Wei et al., 2016). This idea motivates us to make use of the property of tangent space of low-rank tensors to reduce the computational complexity in low-rank projection of tensors.

4.2.1. Efficient Projection $\mathcal{D}-\mathcal{S}_k$ onto $\mathbb{L}$

We begin with the projection onto the low multi-rank tensor space. In step 3 of Algorithm 2, $\mathcal{L}_{k+1}$ is obtained by truncated transformed t-SVD on a full-sized tensor. To reduce the computational complexity of transformed t-SVD, we

[1] $\bar{\sigma}_1(\mathcal{X}):=\max\{\hat{\mathcal{S}}(i,i,c)|\hat{\mathcal{S}}(i,i,c)>0, i\le r_i\}$ and $\beta=2\mu s_r/nq$ where $\mathcal{X}=\mathcal{U}\diamond_\Phi\mathcal{S}\diamond_\Phi\mathcal{V}^H$

incorporate the property of the tangent of low multi-rank tensor space. We divide this subsection into two parts: efficient projection and selection of tangent space.

Efficient projection. The efficient projection consists of two steps: the first step is to project $\mathcal{D}-\mathcal{S}_k$ onto a tangent space of the low multi-rank space, the second step is to project the tensor in the tangent space onto the low multi-rank space. Next proposition gives the tangent space of multi-rank $\boldsymbol{r}$ space at any point $\mathcal{L}$ and how to project to this tangent space.

Proposition 4.1 (Proposition 2.3 in Song et al. 2020b)**.** *Let $\mathcal{L}\in\mathbb{R}^{n\times n\times q}$ be a multi-rank $\boldsymbol{r}$ tensor. The tangent of multi-rank $\boldsymbol{r}$ tensor space at $\mathcal{L}$ is*

$$\mathbb{T}=\left\{\mathcal{U}\diamond_\Phi\mathcal{X}^H+\mathcal{Y}\diamond_\Phi\mathcal{V}^H|\mathcal{X},\mathcal{Y}\in\mathbb{R}^{n\times r\times q}\right\},$$

where $\mathcal{L}=\mathcal{U}\diamond_\Phi\Sigma\diamond_\Phi\mathcal{V}^H$ is the transformed t-SVD of $\mathcal{L}$. The Euclidean projection of $\mathcal{Z}\in\mathbb{R}^{n\times n\times q}$ onto $\mathbb{T}$ is

$$\begin{aligned}P_{\mathbb{T}}(\mathcal{Z})=&\mathcal{U}\diamond_\Phi\mathcal{V}^H\diamond_\Phi\mathcal{Z}+\mathcal{Z}\diamond_\Phi\mathcal{V}\diamond_\Phi\mathcal{V}^H\\&-\mathcal{U}\diamond_\Phi\mathcal{U}^H\diamond_\Phi\mathcal{Z}\diamond_\Phi\mathcal{V}\diamond_\Phi\mathcal{V}^H.\end{aligned}\tag{5}$$

Based on this proposition, the projection of $P_{\mathbb{T}}(\mathcal{Z})$ onto the low multi-rank space can be done as follows.

Proposition 4.2. *Projection of $P_{\mathbb{T}}(\mathcal{Z})$ onto low multi-rank space $\mathbb{L}$ can be executed by*

$$P_{\mathbb{L}}\circ P_{\mathbb{T}}(\mathcal{Z})=\begin{bmatrix}\mathcal{U} & \mathcal{Q}_1\end{bmatrix}\diamond_\Phi P_{\mathbb{L}}(\mathcal{M})\diamond_\Phi\begin{bmatrix}\mathcal{V}^H\\ \mathcal{Q}_2^H\end{bmatrix},\tag{6}$$

where

$$\mathcal{M}=\begin{bmatrix}\mathcal{U}^H\diamond_\Phi\mathcal{Z}\diamond_\Phi\mathcal{V} & \mathcal{R}_2^H\\ \mathcal{R}_1 & \mathbf{0}\end{bmatrix}\in\mathbb{R}^{2r\times 2r\times q},$$

$\mathcal{Q}_1\diamond_\Phi\mathcal{R}_1=(\mathcal{I}_\Phi-\mathcal{U}\diamond_\Phi\mathcal{U}^H)\diamond_\Phi\mathcal{Z}\diamond_\Phi\mathcal{V}$ and $\mathcal{Q}_2\diamond_\Phi\mathcal{R}_2=(\mathcal{I}_\Phi-\mathcal{V}\diamond_\Phi\mathcal{V}^H)\diamond_\Phi\mathcal{Z}\diamond_\Phi\mathcal{U}$ are t-QR (Theorem B.2).

Thus, instead of performing directly transformed t-SVD on $P_{\mathbb{T}}(\mathcal{D}-\mathcal{S}_k)$, this proposition allows performing t-SVD on a smaller-sized tensor $\mathcal{M}$. Hence, we have obtained a way for efficiently projecting onto the low multi-rank space.

Selection of tangent space. Next, we show which tangent space to be used in Proposition 4.1, so that exact recovery of the whole algorithm can still be ensured. Similar to APT, we will project $\mathcal{D}-\mathcal{S}_k$ to $\mathbb{L}$. So we need to give the tensor where the tangent space is defined. Such a tensor $\tilde{\mathcal{L}}_k$ is obtained by Algorithm 3. This algorithm trims $\mathcal{L}_k$ to a specific incoherence level μ (Assumption 5.1) with $\tilde{\mathcal{L}}_k$ remaining a multi-rank $\mathbf{r}$ tensor.

In order to get the t-SVD formulation of $\tilde{\mathcal{L}}_k$, we conduct t-QR on $\tilde{\mathcal{A}}_k$ and $\tilde{\mathcal{B}}_k$, i.e., $\tilde{\mathcal{A}}_k=\tilde{\mathcal{Q}}_1\diamond_\Phi\tilde{\mathcal{R}}_1$ and $\tilde{\mathcal{B}}_k=\tilde{\mathcal{Q}}_2\diamond_\Phi\tilde{\mathcal{R}}_2$. Hence, the t-SVD formulation of $\tilde{\mathcal{L}}_k$ is $\tilde{\mathcal{L}}_k=\mathcal{A}_k\diamond_\Phi\tilde{\Sigma}\diamond_\Phi\mathcal{B}_k^H$

Algorithm 3 Trim for tensor

Require: $\mathcal{L}_k=\mathcal{U}\diamond_\Phi\Sigma\diamond_\Phi\mathcal{V}^H\in\mathbb{R}^{n\times n\times q}$: tensor to be trimmed; μ: target incoherence level.
1: $a_\mu=\sqrt{\frac{\mu s_r}{nq}}, b_\mu=\sqrt{\frac{\mu s_r}{nq}}$
2: **for** $i=1$ to n **do**
3: $\quad\tilde{\mathcal{A}}_k^H(:,i,:)=\min\{1,\frac{a_\mu}{\|\mathcal{U}^H(:,i,:)\|_F}\}\mathcal{U}^H(:,i,:)$
4: $\quad\tilde{\mathcal{B}}_k^H(:,i,:)=\min\{1,\frac{b_\mu}{\|\mathcal{V}^H(:,i,:)\|_F}\}\mathcal{V}^H(:,i,:)$
5: **end for**
6: **Return:** $\tilde{\mathcal{L}}_k=\tilde{\mathcal{A}}_k\diamond_\Phi\Sigma\diamond_\Phi\tilde{\mathcal{B}}_k^H$

where $\mathcal{A}_k=\tilde{\mathcal{A}}_k\diamond_\Phi\tilde{\mathcal{U}}$, $\mathcal{B}_k=\tilde{\mathcal{B}}_k\diamond_\Phi\tilde{\mathcal{V}}$ with $\tilde{\mathcal{R}}_1\diamond_\Phi\Sigma\diamond_\Phi\tilde{\mathcal{R}}_2^H=\tilde{\mathcal{U}}\diamond_\Phi\tilde{\Sigma}\diamond_\Phi\tilde{\mathcal{V}}^H$ being the transformed t-SVD. So we only need to conduct t-SVD on a tensor with size $r\times r\times q$. According to Proposition 4.1, the projection onto the tangent space defined on $\tilde{\mathcal{L}}_k$ is

$$\begin{aligned}P_{\mathbb{T}_k}(\mathcal{Z})=&\mathcal{A}_k\diamond_\Phi\mathcal{A}_k^H\diamond_\Phi\mathcal{Z}+\mathcal{Z}\diamond_\Phi\mathcal{B}_k\diamond_\Phi\mathcal{B}_k^H\\&-\mathcal{A}_k\diamond_\Phi\mathcal{A}_k^H\diamond_\Phi\mathcal{Z}\diamond_\Phi\mathcal{B}_k\diamond_\Phi\mathcal{B}_k^H,\end{aligned}$$

where $\mathbb{T}_k$ is the tangent space at $\tilde{\mathcal{L}}_k$. Finally, we have $\mathcal{L}_{k+1}=P_{\mathbb{L}}\circ P_{\mathbb{T}_k}(\mathcal{D}-\mathcal{S}_k)$.

Because all operations of tensor here can be converted into the operations of block diagonal matrix, we do not need multiple Φ to transform tensor in each operation and the complexity of tensor transformation in low-rank projection counts once. In t-QR decomposition, they have six times multiplications and twice t-QR decompositions, which costs $\sum_{i=1}^q\left(6n^2r_i+\mathcal{O}(nr_i^2)\right)$. The projection onto tangent $\mathbb{T}$ has four times as many multiplications and once transformed t-SVD, which costs $\sum_{i=1}^q\left(n^2r_i+9nr_i^2+\mathcal{O}(r_i^3)\right)$. Adding the computational complexity of transformation into the block diagonal formulation, the overall computational complexity is $\mathcal{O}(n^2q^2+n^2s_r)$ with $s_r=\sum_{i=1}^q r_i$, which is less than the computational complexity of the T-TRPCA: $\mathcal{O}(n^2q^2+n^3q)$.

4.2.2. The Complete Algorithm

The complete procedure of EAPT is described in Algorithm 4. The difference from APT is that the projection onto $\mathbb{L}$ can be efficiently executed. Like APT, EAPT can be divided into two parts: initialization and alternating projection. Here, we use the same initialization method as APT. After initialization, EAPT performs alternating projection. At the k-th iteration, the first step projecting $\mathcal{D}-\mathcal{S}_k$ onto $\mathbb{L}$ is presented in Section 4.2.1. The second step follows the idea in Section 4.1 which projects onto $\mathbb{S}$ with $\mathcal{S}_{k+1}=T_{\zeta_{k+1}}(\mathcal{D}-\mathcal{L}_{k+1})$ and

$$\zeta_{k+1}=\beta\left(\bar{\sigma}_{s_r+1}(\mathcal{X})+\gamma^{k+1}\bar{\sigma}_1(\mathcal{X})\right),\tag{7}$$

where $\mathcal{X}=P_{\mathbb{T}_k}(\mathcal{D}-\mathcal{S}_{k+1})$, $\beta=\mu s_r/2nq$, $\bar{\sigma}_{s_r+1}(\mathcal{X}):=\max\{\hat{\mathcal{S}}(i,i,c)|\hat{\mathcal{S}}(i,i,c)>0,i>r_i\}$ with $\mathcal{X}=\mathcal{U}\diamond_\Phi\mathcal{S}\diamond_\Phi$

$\mathcal{V}^H$ being the transformed t-SVD.

Algorithm 4 EAPT: Efficient Alternating Projection Algorithm for Tensor RPCA

```
Run Algorithm 1 for initialization
for k = 0 to T − 1 do
    L̃_k = Trim(L_k, μ)
    L_{k+1} = P_𝕃 ∘ P_{𝕋_k} (D − S_k)
    S_{k+1} = T_{ζ_{k+1}}(D − L_{k+1}) where ζ_{k+1} is in (7)
end for
Return: L_T and S_T
```

In conclusion, EAPT utilizes the low multi-rank tangent space property to reduce computational complexity. Compared to other methods, EAPT has computational complexity as $\mathcal{O}(n^2 q \log q + n^2 s_r)$ when using FFT transformation and $\mathcal{O}(n^2 q^2 + n^2 s_r)$ when using common unitary transformation, while keeping the exact recovery guarantee (Theorem 5.5) and linear convergence rate as shown in Table 1.

5. Theoretical Guarantees

Following (Zhang & Aeron, 2016), we first have the below tensor incoherence condition:

Assumption 5.1. Given the transformed t-SVD of a tensor $\mathcal{L} = \mathcal{U} \diamond_\Phi \Sigma \diamond_\Phi \mathcal{V}^H \in \mathbb{R}^{n\times n\times q}$ with multi-rank $\boldsymbol{r}$, $\mathcal{L}$ is said to satisfy the tensor incoherence condition, if there exists $\mu > 0$ such that

$$\textit{Tensor-column:} \quad \frac{nq}{s_r} \max_{i\in[n]} \|\mathcal{U}^H \diamond_\Phi \dot{e}_i\|_F^2 \le \mu;$$
$$\textit{Tensor-row:} \quad \frac{nq}{s_r} \max_{j\in[n]} \|\mathcal{V}^H \diamond_\Phi \dot{e}_j\|_F^2 \le \mu.$$

Here $\dot{e}_i$ is defined as the $n \times 1 \times q$ column basis with $\Phi[\dot{e}_i](i, 1, :) = 1$.

Basically, Assumption 5.1 shows that the tensor columns $\mathcal{U}(:, i, :)$'s and $\mathcal{V}(:, i, :)$'s need to be uncorrelated with the standard tensor basis. The second assumption is the sparse condition for $\mathcal{S}$ shown in below.

Assumption 5.2. A sparse tensor $\mathcal{S} \in \mathbb{R}^{n\times n\times q}$ is α-sparse, i.e., $\|\mathcal{S}(:, i, :)\|_0 \le \alpha nq$ and $\|\mathcal{S}(i, :, :)\|_0 \le \alpha nq$ for $i \in [n]$.

This assumption ensures the sparse elements are uniformly distributed across different slices so that $\mathcal{S}$ and $\mathcal{L}$ will not be low-rank tensors simultaneously. We do not require sparsity on $S(:, :, i), i \in [q]$ here because such condition cannot guarantee $\mathcal{S}$ is not low-rank.

Let $\mathcal{D} = \mathcal{L}^* + \mathcal{S}^*$ be the observed tensor. Proposition 5.3 guarantees that the initialization obtained by Algorithm 1 is close to the ground-truth.

Proposition 5.3 (Algorithm 1 for initialization)**.** *Assume that a low multi-rank $\boldsymbol{r}$ tensor $\mathcal{L}^*$ satisfies Assumption 5.1 and a sparse tensor $\mathcal{S}^*$ satisfies Assumption 5.2 with $\alpha\mu \lesssim \frac{1}{s_r \kappa \sqrt{q}}$. For hyperparameters obeying $\frac{\mu s_r \bar{\sigma}_1(\mathcal{L}^*)}{nq\bar{\sigma}_1(\mathcal{D})} \le \beta_{init} \le \frac{3\mu s_r \bar{\sigma}_1(\mathcal{L}^*)}{nq\bar{\sigma}_1(\mathcal{D})}$ and $\beta_0 = \frac{\mu s_r}{2nq}$, the outputs of Algorithm 1 satisfy $\|\mathcal{L} - \mathcal{L}_0\| \le 8\alpha\mu s_r \bar{\sigma}_1(\mathcal{L}^*)$ and $\|\mathcal{S} - \mathcal{S}_0\|_\infty \le \frac{\mu s_r}{nq} \bar{\sigma}_1(\mathcal{L}^*)$.*

Based on above assumptions, linear convergence to the ground-truth of Algorithm 2 is ensured in Theorem 5.4.

Theorem 5.4 (Exact recovery of Algorithm 2)**.** *Under the assumption of Proposition 5.3, for any $\epsilon > 0$, we have $\|\mathcal{L}_T - \mathcal{L}^*\| \le 8\alpha\epsilon$ and $\|\mathcal{S}_T - \mathcal{S}^*\|_\infty \le {}^{4\epsilon}/_{nq}$ with $T = \mathcal{O}\left(\log\left({}^1/_\epsilon\right)\right), \beta = {}^{2\mu s_r}/_{nq}$.*

Different from Theorem 5.4, proof of the recovery guarantee of EAPT is more difficult because its low-rank projection is more complicated than APT. Specifically, we need to construct an inequality that reflects the relationship between $\tilde{\mathcal{L}}_k$ and $\mathcal{L}_k$ while APT has no such intermediate variable. However, linear convergence to the ground-truth of Algorithm 4 is still ensured as stated in Theorem 5.5.

Theorem 5.5 (Exact recovery of Algorithm 4)**.** *Under the assumption of Proposition 5.3 except that $\alpha \lesssim \min\{{}^1/_{\mu s_r^2 \kappa^3}, {}^{q^{0.5}}/_{\mu^{1.5} s_r^2 \kappa^2}, {}^{q^{0.5}}/_{\mu^2 s_r^2 \kappa}\}$, for any $\epsilon > 0$, we have $\|\mathcal{L}_T - \mathcal{L}^*\|_\infty \le 8\alpha\epsilon$ and $\|\mathcal{S}_T - \mathcal{S}^*\|_\infty \le {}^\epsilon/_{nq}$ with $T = \mathcal{O}\left(\log\left({}^1/_\epsilon\right)\right), \beta = {}^{\mu s_r}/_{2nq}$.*

By choosing appropriate unitary transformation matrix, a transformed tensor will have a lower multi-rank and its prominent information will be concentrated in some specific slices (Zhang & Ng, 2021). For example, by using DCT matrix for HSI denoising, the prominent information will be concentrated in the low-frequency slices. While the goodness of multi-rank truncation cannot be directly seen from these two theorems, empirically it leads to better performance than tubal rank in Section 6.2.

Existing works that are most technically similar to ours are AltProj (Netrapalli et al., 2014) and AccAltProj (Cai et al., 2019). However, they are designed for matrices. As explained in Zhang & Aeron 2016 and Lu et al. 2019, the extension from matrices to tensors are not trivial as different mathematical tools are required. For example, sparse tensor $\mathcal{S}$ is not sprase any more in the frequency domain and some bounds of norms have to use the property of tensor-tensor product and interconvert between time and frequency domains in proofs. The detailed proofs is presented in Appendix D.

Theorems 5.4 and 5.5 indicate that β_{init} and β depend on unknown parameters about the ground truth tensor $\mathcal{S}^*$. Similar RPCA methods namely AltProj (Netrapalli et al., 2014) and AccAltProj (Cai et al., 2019) also rely on these unknown parameters about their ground truth matrix. How to remove such dependency is still an open issue. Also, unknown μ, s_r, κ and $\bar{\sigma}_1(\mathcal{L}^*)$ is not a practical issue. As we can see in Tabel 2, good empirical performance can be still obtained

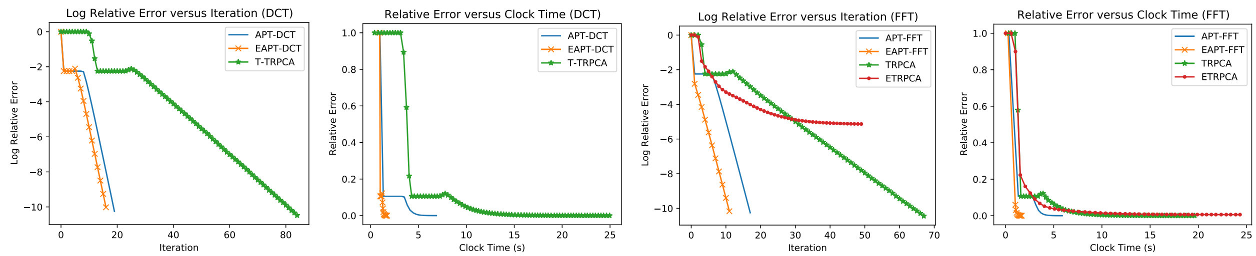

Figure 1. Comparison between different t-SVD based tensor RPCA methods.

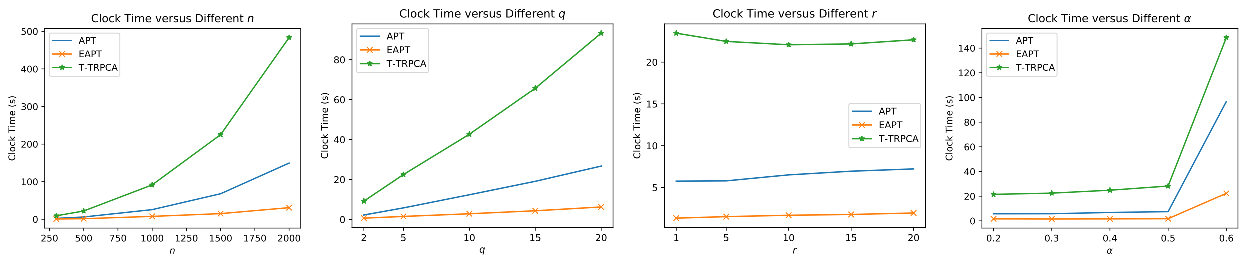

Figure 2. Consistent advantages over T-TRPCA with various parameters. $\alpha = 0.6$ fails to successfully recovery.

by setting β_{init} and β with inaccurate $\mu, s_r, \kappa, \bar{\sigma}_1(\mathcal{L}^*)$.

6. Experiments

In this section, experiments are conducted on both synthetic and real data. All experiments are conducted on a PC with two Intel Xeon Silver 4215R 3.20GHz CPUs and 187GB memory with MATLAB R2021b.

6.1. Synthetic Data

We generate a symmetric low-rank tensor $\mathcal{L} = \mathcal{Q} \diamond_\Phi \mathcal{Q}^H$ in $\mathbb{R}^{n\times n\times q}$ where $\mathcal{Q} \in \mathbb{R}^{n\times r\times q}$ is a random tensor with its entry sampled i.i.d. from the standard normal distribution. We use normalized DFT and DCT as the unitary transformation for tensor-tensor product because these two transformations have clear physical meanings as the noise in the time domain will be transformed into the high-frequency terms in the frequency domain (Rao et al., 2010). Considering DFT and DCT cannot concentrate information in synthetic data, we do not consider multi-rank in this subsection. Whether the elements in the sparse tensor $\mathcal{S}$ are 0 is determined by the Bernoulli distribution with the parameter α. The value of the non-zero entries $\mathcal{S}_{ijt}$ in $\mathcal{S}$ is drawn uniformly from the interval $[-c \cdot \mathbb{E}[|\mathcal{L}|_{ijt}], c \cdot \mathbb{E}[|\mathcal{L}|_{ijt}]]$ for some constant $c > 0$. We set $c = 3$ in our synthetic experiments. The observed tensor is $\mathcal{D} = \mathcal{L} + \mathcal{S}$. The stop criterion of all methods is $\|\mathcal{D} - \mathcal{L}_k - \mathcal{S}_k\|_F / \|\mathcal{D}\|_F \leq 10^{-4}$. Relative error ("Rel Err" for short) is defined as $\|\mathcal{L}_k - \mathcal{L}\|_F / \|\mathcal{L}\|_F$.

The convergence and efficiency of our methods are shown in Figure 1. We conduct experiments on tensor-tensor product with both FFT and DCT, and compare with other methods based on tensor-tensor product such as TRPCA (Lu et al., 2019), T-TRPCA (Lu, 2021), and ETRPCA (Gao et al., 2020). We set $n = 500, q = 5, r = 5, \alpha = 0.3$ in Figure 1. The hyperparameters of our methods are set as the same suggested by Proposition 5.3, Theorems 5.4 and 5.5. The hyperparameters of other methods are listed in the Appendix C.1. The experimental results in Figure 1 show that both APT and EAPT converge linearly. Also, EAPT can save plenty of time cost compared to others.

Then, we conduct experiments to show the performance of our methods with various parameters. Figure 2 shows the running time of different methods with different parameters. Parameters are changed based on $n = 500, q = 5, r = 5, \alpha = 0.3$, that is, the experiment in each figure changes one parameter and leaves the other unchanged. The experimental results show that our methods have consistent advantages over T-TRPCA with various parameters. We also conduct experiments to show the robustness of our methods in Appendix C.1.2.

6.2. Real Data

In this subsection, we compare the performance of TRPCA (Lu et al., 2019), T-TRPCA (Lu, 2021), ETRPCA (Gao et al., 2020), TTNN (Yang et al., 2020), SNN (Huang et al., 2015), Atomic Norm (Driggs et al., 2019), and RTD (Anandkumar et al., 2016) on real datasets. The difference of these methods is presented in Table 3.

6.2.1. HSI DENOISING

Here, we compare the performance of different methods on hyperspectral image denoising. We use the CAVE dataset (Yasuma et al., 2008) for experiments. For a hyperspectral image $\mathcal{H} \in \mathbb{R}^{m\times n\times q}$, αmnq pixels from $\mathcal{H}$ are randomly selected as noises by Bernoulli distribution. The values of these αmnq pixels are sampled from uniform

Table 2. Performance and clock time in seconds of HSI denoising. The best result in "**bold**" and the second best in "underline".

	TT		Tucker		CP		CP		t-SVD									
	TTNN		SNN		Atomic Norm		RTD		TRPCA		T-TRPCA		ETRPCA		EAPT-FFT		EAPT-DCT	
	PSNR	Time	PSNR	Time	PSNR	Time	PSNR	Time	PSNR	Time	PSNR	Time	PSNR	Time	PSNR	Time	PSNR	Time
toys	29.39	85.4	28.47	297.3	16.96	223.1	24.18	368.6	34.04	63.6	34.09	75.5	38.47	121.7	39.95	36.5	**41.87**	**35.8**
feathers	29.00	84.5	28.00	297.8	17.89	331.4	24.42	486.8	31.62	59.9	31.36	70.7	36.72	120.8	38.01	33.4	**39.61**	**32.3**
sponges	36.90	83.3	37.38	305.1	19.48	337.4	28.24	249.8	31.52	59.9	30.28	70.8	34.81	122.2	38.88	33.6	**40.05**	**22.5**
watercolors	28.74	85.9	28.31	284.7	18.33	377.2	23.49	353.1	36.28	61.6	36.3	71.3	40.59	121.6	41.43	**34.2**	**41.66**	35.9
paints	30.35	83.1	30.33	291.5	18.98	336.0	25.16	457.5	33.83	61.1	33.72	71.5	38.15	123.7	39.45	**34.3**	**39.53**	35.6
sushi	31.60	84.0	31.57	312.2	17.42	320.9	29.96	492.3	33.40	62.5	33.50	72.5	35.67	120.3	36.03	36.3	**39.30**	**32.3**

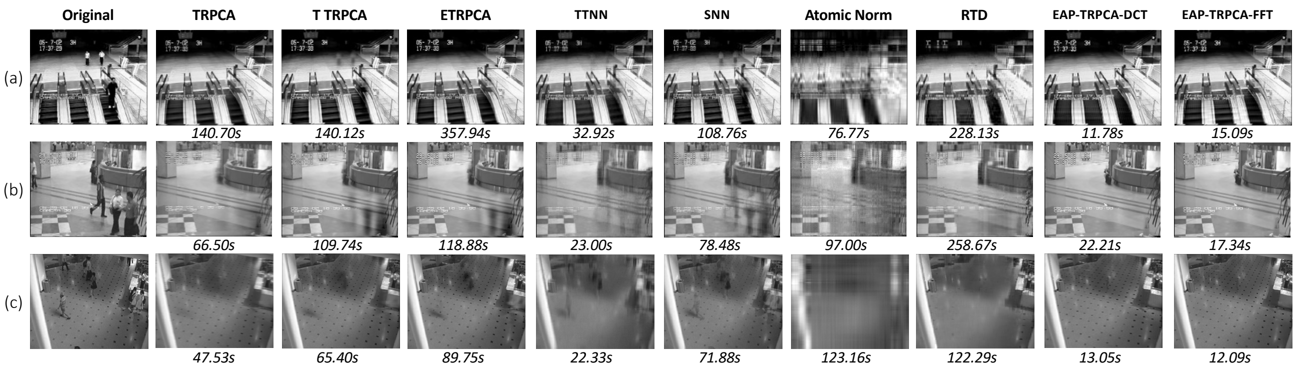

Figure 3. Video background subtraction results of different methods and their corresponding reconstruction clock time. (a) Escalator with 200 frames; (b) Hall with 100 frames; (c) ShoppingMall with 50 frames.

distribution between interval $[0,1]$. Since EAPT has comparable performance with APT but is more efficient than APT, we only involve EAPT for comparison. For detailed hyperparameter settings, please refer to the Appendix C.2.

We randomly select 6 hyperspectral images from CAVE dataset for comparison. The performance evaluated with final PSNR and clock time is presented in Table 2. "EAPT-FFT" and "EAPT-DCT" in the table mean that tensor-tensor product with respect to a normalized DFT matrix and a DCT matrix are used for experiments, respectively. In Table 2, methods based on t-SVD have better performance than that based on others because t-SVD can utlize information in both time and frequency (transformed) domain. Also, the performance of our methods is better than other methods, while ours can save a large amount of time cost compared to others. Besides, EAPT-DCT has better performance than EAPT-FFT. The reason is probably the "spectral compaction" property of DCT, i.e., a transformed tensor in frequency domain tends to have its value concentrated in a small number of slices when compared to other transformation like FFT. And multi-rank truncation can better utilize this property by setting the large rank for the low-frequency slices and small rank for the high-frequency slices. Also, we want to clarify that the improvement in terms of the time cost here is consistent with our theoretical result. Because the hyperspectral images are reshaped into a different size (see Figure 4) and the number of the bands of CAVE is not too larger than the rank of the tensor so that the improvement of our method on HSI denoising becomes less significant.

6.2.2. VIDEO BACKGROUND SUBTRACTION

In this section, we compare the performance of various methods on video background subtraction. The task is to separate the moving foreground objects from a static background. Three videos are used for comparison, i.e., Escalator, Hall, and ShoppingMall. We choose 200, 100 and 50 frames from Escalator, Hall and ShoppingMall respectively for comparison. The visual comparison in Figure 3 shows that our methods can successfully extract the background from these videos with less reconstruction time.

7. Conclusion

In this paper, we study nonconvex tensor robust PCA based on transformed t-SVD and propose two algorithms, i.e., APT and EAPT, based on alternating projection. Our methods converge linearly and guarantee exact recovery. Also, EAPT utilizes the property of tangent space to reduce the computational complexity. Experimental results show that our proposed methods are more effective and efficient than existing methods. As a future work, it will be interesting to extend our work to higher order tensors and consider the noise in theoretical analysis.

Acknowledgements

This research was supported in part by National Key R&D Program of China (2020YFA0713900), the China NSFC projects under contract 11971374, 61721002, U1811461, and the Macao Science and Technology Development Fund under Grant 061/2020/A2.

References

Anandkumar, A., Jain, P., Shi, Y., and Niranjan, U. N. Tensor vs. matrix methods: Robust tensor decomposition under block sparse perturbations. In *Artificial Intelligence and Statistics*, pp. 268–276. PMLR, 2016.

Cai, H., Cai, J.-F., and Wei, K. Accelerated alternating projections for robust principal component analysis. *Journal of Machine Learning Research*, 20(1):685–717, 2019.

Cai, H., Chao, Z., Huang, L., and Needell, D. Fast robust tensor principal component analysis via fiber cur decomposition. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 189–197, 2021.

Candès, E. J. and Recht, B. Exact matrix completion via convex optimization. *Foundations of Computational mathematics*, 9(6):717–772, 2009.

Candès, E. J., Li, X., Ma, Y., and Wright, J. Robust principal component analysis? *Journal of the ACM (JACM)*, 58(3): 1–37, 2011.

Chang, R. W. Synthesis of band-limited orthogonal signals for multichannel data transmission. *Bell System Technical Journal*, 45(10):1775–1796, 1966.

Cichocki, A., Mandic, D., De Lathauwer, L., Zhou, G., Zhao, Q., Caiafa, C., and Phan, H. A. Tensor decompositions for signal processing applications: From two-way to multiway component analysis. *IEEE Signal Processing Magazine*, 32(2):145–163, 2015.

Driggs, D., Becker, S., and Boyd-Graber, J. Tensor robust principal component analysis: Better recovery with atomic norm regularization. *arXiv preprint arXiv:1901.10991*, 2019.

Gabay, D. and Mercier, B. A dual algorithm for the solution of nonlinear variational problems via finite element approximation. *Computers & Mathematics with Applications*, 2(1):17–40, 1976.

Gandy, S., Recht, B., and Yamada, I. Tensor completion and low-n-rank tensor recovery via convex optimization. *Inverse Problems*, 27(2):025010, 2011.

Gao, Q., Zhang, P., Xia, W., Xie, D., Gao, X., and Tao, D. Enhanced tensor rpca and its application. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 43 (6):2133–2140, 2020.

Hao, N., Kilmer, M. E., Braman, K., and Hoover, R. C. Facial recognition using tensor-tensor decompositions. *SIAM Journal on Imaging Sciences*, 6(1):437–463, 2013.

Huang, B., Mu, C., Goldfarb, D., and Wright, J. Provable models for robust low-rank tensor completion. *Pacific Journal of Optimization*, 11(2):339–364, 2015.

Kernfeld, E., Kilmer, M., and Aeron, S. Tensor–tensor products with invertible linear transforms. *Linear Algebra and its Applications*, 485:545–570, 2015.

Kilmer, M. E. and Martin, C. D. Factorization strategies for third-order tensors. *Linear Algebra and its Applications*, 435(3):641–658, 2011.

Kilmer, M. E., Horesh, L., Avron, H., and Newman, E. Tensor-tensor algebra for optimal representation and compression of multiway data. *Proceedings of the National Academy of Sciences*, 118(28), 2021.

Kolda, T. G. and Bader, B. W. Tensor decompositions and applications. *SIAM review*, 51(3):455–500, 2009.

Liu, J., Musialski, P., Wonka, P., and Ye, J. Tensor completion for estimating missing values in visual data. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(1):208–220, 2012.

Lu, C. Transforms based tensor robust pca: Corrupted low-rank tensors recovery via convex optimization. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 1145–1152, 2021.

Lu, C., Feng, J., Chen, Y., Liu, W., Lin, Z., and Yan, S. Tensor robust principal component analysis: Exact recovery of corrupted low-rank tensors via convex optimization. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 5249–5257, 2016.

Lu, C., Feng, J., Chen, Y., Liu, W., Lin, Z., and Yan, S. Tensor robust principal component analysis with a new tensor nuclear norm. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 42(4):925–938, 2019.

Netrapalli, P., Niranjan, U., Sanghavi, S., Anandkumar, A., and Jain, P. Provable non-convex robust pca. In *Proceedings of the 27th International Conference on Neural Information Processing Systems-Volume 1*, pp. 1107–1115, 2014.

Oseledets, I. V. Tensor-train decomposition. *SIAM Journal on Scientific Computing*, 33(5):2295–2317, 2011.

Rao, K. R., Kim, D. N., and Hwang, J. J. Fast fourier transform: algorithms and applications. 2010.

Song, G., Ng, M. K., and Zhang, X. Robust tensor completion using transformed tensor singular value decomposition. *Numerical Linear Algebra with Applications*, 27(3): e2299, 2020a.

Song, G.-J., Wang, X.-Z., and Ng, M. K. Riemannian conjugate gradient descent method for third-order tensor completion. *arXiv preprint arXiv:2011.11417*, 2020b.

Tomioka, R., Hayashi, K., and Kashima, H. Estimation of low-rank tensors via convex optimization. *arXiv preprint arXiv:1010.0789*, 2010.

Vandereycken, B. Low-rank matrix completion by riemannian optimization. *SIAM Journal on Optimization*, 23(2): 1214–1236, 2013.

Wei, K., Cai, J.-F., Chan, T. F., and Leung, S. Guarantees of riemannian optimization for low rank matrix recovery. *SIAM Journal on Matrix Analysis and Applications*, 37 (3):1198–1222, 2016.

Wright, J., Ganesh, A., Rao, S. R., Peng, Y., and Ma, Y. Robust principal component analysis: Exact recovery of corrupted low-rank matrices via convex optimization. In *NIPS*, volume 58, pp. 289–298, 2009.

Xu, H., Caramanis, C., and Sanghavi, S. Robust pca via outlier pursuit. *arXiv preprint arXiv:1010.4237*, 2010.

Yang, J.-H., Zhao, X.-L., Ji, T.-Y., Ma, T.-H., and Huang, T.-Z. Low-rank tensor train for tensor robust principal component analysis. *Applied Mathematics and Computation*, 367:124783, 2020.

Yasuma, F., Mitsunaga, T., Iso, D., and Nayar, S. Generalized Assorted Pixel Camera: Post-Capture Control of Resolution, Dynamic Range and Spectrum. Technical report, Nov 2008.

Yi, X., Park, D., Chen, Y., and Caramanis, C. Fast algorithms for robust pca via gradient descent. In *Advances in Neural Information Processing Systems*, pp. 4152–4160, 2016.

Zhang, X. and Ng, M. K.-P. Low rank tensor completion with poisson observations. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2021.

Zhang, Z. and Aeron, S. Exact tensor completion using t-svd. *IEEE Transactions on Signal Processing*, 65(6): 1511–1526, 2016.

Zhang, Z., Ely, G., Aeron, S., Hao, N., and Kilmer, M. Novel methods for multilinear data completion and denoising based on tensor-svd. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 3842–3849, 2014.

A. Different Tensor RPCA Methods

Here, we present the comparison of tensor RPCA methods based on different tensor decomposition in Table 3.

Table 3. Comparison of various tensor RPCA methods. "✓": exist; "✗": doesn't exist; "AP": alternating projection; "ALM": augmented Lagrange Multiplier; "T t-SVD": Transformed t-SVD. The computational complexities are calculated on tensors in $\mathbb{R}^{n\times n\times n}$.

	decomposition	Convergence	Iteration complexity	Algorithm	Recovery
TRPCA (Lu et al., 2019)	t-SVD	$\mathcal{O}(1/\epsilon)$	$\mathcal{O}\left(n^3\log n+n^4\right)$	ADMM	✓
Transformed TRPCA (Lu, 2021)	T t-SVD	$\mathcal{O}(1/\epsilon)$	$\mathcal{O}\left(n^4\right)$	ADMM	✓
ETRPCA (Gao et al., 2020)	t-SVD	✓	$\mathcal{O}\left(n^3\log n+n^4\right)$	ALM	✗
RTCUR (Cai et al., 2021)	Fiber CUR	✗	$\mathcal{O}\left(nr^2\log^2(n)+r^4\log^4(n)\right)$	AP	✗
TTNN (Yang et al., 2020)	TT	✓	$\mathcal{O}\left(n^4\right)$	ADMM	✗
SNN (Huang et al., 2015)	Tucker	$\mathcal{O}(1/\epsilon)$	$\mathcal{O}\left(n^4\right)$	ADMM	✓
Atomic Norm (Driggs et al., 2019)	CP	✗	$\mathcal{O}(n^4r)$	LBFGS	✗
RTD (Anandkumar et al., 2016)	CP	$\mathcal{O}\left(\log(1/\epsilon)\right)$	$\mathcal{O}\left(n^{4+c}r^2\right)$	AP	✓
APT	T t-SVD	$\mathcal{O}\left(\log(1/\epsilon)\right)$	$\mathcal{O}(n^4)$	AP	✓
EAPT	T t-SVD	$\mathcal{O}\left(\log(1/\epsilon)\right)$	$\mathcal{O}(n^4+n^2s_r)$	AP	✓

B. Extension of "Preliminaries"

With the tensor-tensor product and the block-diagonal matrix representation, the following relation is obvious for any tensor $\mathcal{X}\in\mathbb{C}^{n_1\times n_2\times q}$ and $\mathcal{Y}\in\mathbb{C}^{n_2\times n_3\times q}$:

$$\mathcal{X}\diamond_\Phi\mathcal{Y}=\mathcal{C}\Leftrightarrow\bar{\mathcal{X}}\cdot\bar{\mathcal{Y}}=\bar{\mathcal{C}}.$$

Next, we recall the definitions of the identity tensor, unitary tensor, diagonal tensor, and conjugate transpose of a tensor. The identity tensor $\mathcal{I}\in\mathbb{C}^{n\times n\times q}$ respect to Φ is denoted as $\mathcal{I}=\Phi^H[\mathcal{I}_\Phi]$, where each frontal slice of $\mathcal{I}_\Phi\in\mathbb{C}^{n\times n\times q}$ is the $n\times n$ identity matrix. Based on the identity tensor, the unitary tensor with respect to Φ is defined as $\mathcal{U}\diamond_\Phi\mathcal{U}^H=\mathcal{U}^H\diamond_\Phi\mathcal{U}=\mathcal{I}$. A tensor with each frontal slice being diagonal is called to be diagonal. The conjugate transpose of any tensor $\mathcal{X}\in\mathbb{C}^{m\times n\times q}$ with respect to a unitary matrix $\Phi\in\mathbb{C}^{q\times q}$, denoted by $\mathcal{X}^H\in\mathbb{C}^{n\times m\times q}$, is defined as $\mathcal{X}^H=\Phi^H\left[\text{fold}\left(\bar{\mathcal{X}}^H\right)\right]$.

Now we define the basis for tensor-tensor product. $\dot{e}_i$ is defined as the $m\times 1\times q$ column basis with $\Phi[\dot{e}_i](i,1,:)=1$ and $\dot{e}_j$ is defined as the $n\times 1\times q$ column basis with $\Phi[\dot{e}_j](j,1,:)=1$. Another tensor basis is called tubal basis e_k of size $1\times 1\times q$ with $e_i(1,1,k)=1$ and the rest entries equal 0. Tensor $\mathcal{E}_{ijk}$ with its (i,j,k)-th element equal 1 and others equal 0 can be denoted by $\mathcal{E}_{ijk}=\dot{e}_i\diamond_\Phi e_k\diamond_\Phi\dot{e}_j^H$.

Next theorem gives the best multi-rank $\boldsymbol{r}$ approximation of $\mathcal{X}$:

Theorem B.1 (Eckart-Young theorem, Theorem 11 in Kilmer et al. 2021)**.** *If $\mathcal{X}=\mathcal{U}\diamond_\Phi\mathcal{S}\diamond_\Phi\mathcal{V}^H\in\mathbb{C}^{m\times n\times q}$ is the transformed t-SVD of $\mathcal{X}$. Define $H_{\boldsymbol{r}}(\mathcal{X})$ to be the approximation having multi-rank $\boldsymbol{r}$, that is,*

$$\hat{\mathcal{X}}_\Phi(:,:,i)=\hat{\mathcal{U}}_\Phi(:,1:r_i)\hat{\mathcal{S}}_\Phi(1:r_i,1:r_i,i)\hat{\mathcal{V}}_\Phi^H(:,1:r_i,i).$$

Then, $H_{\boldsymbol{r}}(\mathcal{X})$ is the best multi-rank $\boldsymbol{r}$ approximation to $\mathcal{X}$ in the Frobenius norm and

$$H_{\boldsymbol{r}}=\arg\min_{\mathit{rank}_m(\tilde{\mathcal{X}})\leq\boldsymbol{r}}\|\mathcal{X}-\tilde{\mathcal{X}}\|_F.$$

The following theorem is transformed t-QR decomposition of tensors.

Theorem B.2 (transformed t-QR)**.** *Let $\mathcal{X}\in\mathbb{C}^{m\times n\times q}$. Then it can be factorized as $\mathcal{X}=\mathcal{Q}\diamond_\Phi\mathcal{R}$, where $\mathcal{Q}\in\mathbb{C}^{m\times m\times q}$ is a orthogonal tensor, and $\mathcal{R}\in\mathbb{C}^{m\times n\times q}$ is an f-upper triangle tensor (each frontal slice is an upper triangular matrix).*

The tensor spectral norm $\|\mathcal{X}\|$ with respect to a unitary matrix Φ is defined as $\|\mathcal{X}\|=\|\bar{\mathcal{X}}\|$. With block-diagonal representation, the inner product of two tensors $\mathcal{X},\mathcal{Y}\in\mathbb{C}^{m\times n\times q}$ can also be calculated with the following property

$$\langle\mathcal{X},\mathcal{Y}\rangle=\langle\bar{\mathcal{X}},\bar{\mathcal{Y}}\rangle.$$

With above equality, the Frobenius of a tensor $\mathcal{X}\in\mathbb{C}^{m\times n\times q}$ can be represented as $\|\mathcal{X}\|_F=\|\bar{\mathcal{X}}\|_F$.

Some important singular values of tensor $\mathcal{X}$ are denoted as $\bar{\sigma}_1(\mathcal{X}) := \max\{\hat{\mathcal{S}}(i,i,c)|\hat{\mathcal{S}}(i,i,c) > 0, i \leq r_i\}$, $\bar{\sigma}_{s_r}(\mathcal{X}) := \min\{\hat{\mathcal{S}}_\Phi(i,i,c)|\hat{\mathcal{S}}_\Phi(i,i,c) > 0, i \leq r_i\}$, $\bar{\sigma}_{s_r+1}(\mathcal{X}) := \max\{\hat{\mathcal{S}}(i,i,c)|\hat{\mathcal{S}}(i,i,c) > 0, i > r_i\}$, $\bar{\sigma}_{rq}(\mathcal{X}) := \min\{\hat{\mathcal{S}}_\Phi(i,i,c)|\hat{\mathcal{S}}_\Phi(i,i,c) > 0, i \leq r\}$, and $\bar{\sigma}_{rq+1}(\mathcal{X}) := \max\{\hat{\mathcal{S}}(i,i,c)|\hat{\mathcal{S}}(i,i,c) > 0, i > r\}$ where $\mathcal{X} = \mathcal{U} \diamond_\Phi \mathcal{S} \diamond_\Phi \mathcal{V}^H$, r_i is the i-th element of multi-rank $\mathbf{r}$, and $r = \max\{r_i\}$ is the tubal rank.

A partitioned tensor is a tensor that is interpreted as having been broken into sections called subtensors or blocks. A partitioned tensor $\mathcal{A}$ with q row partitions and s column partitions with subtensors $\mathcal{A}_{ij}$ can be denoted by its i-th frontal slice as

$$\mathbf{A}^{(i)} = \begin{bmatrix} \mathbf{A}_{11}^{(i)} & \mathbf{A}_{12}^{(i)} & \cdots & \mathbf{A}_{1s}^{(i)} \\ \mathbf{A}_{21}^{(i)} & \mathbf{A}_{22}^{(i)} & \cdots & \mathbf{A}_{2s}^{(i)} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{A}_{q1}^{(i)} & \mathbf{A}_{q2}^{(i)} & \cdots & \mathbf{A}_{qs}^{i} \end{bmatrix}$$

where $\mathbf{A}_{ij}^{(i)}$ is the i-th frontal slice of $\mathcal{A}_{ij}$. For convenient, we also represent $\mathcal{A}$ as

$$\mathcal{A} = \begin{bmatrix} \mathcal{A}_{11} & \mathcal{A}_{12} & \cdots & \mathcal{A}_{1s} \\ \mathcal{A}_{21} & \mathcal{A}_{22} & \cdots & \mathcal{A}_{2s} \\ \vdots & \vdots & \ddots & \vdots \\ \mathcal{A}_{q1} & \mathcal{A}_{q2} & \cdots & \mathcal{A}_{qs} \end{bmatrix}.$$

Our analysis are based on symmetric tensors, but similar results can be obtained for asymmetric matrix by casting the general tensor problem to symmetric augmented tensors. Without loss of generality, assume $m \leq n$ and $dm \leq n < (d+1)m$ for some $d \geq 1$ and construct symmetric tensors $\hat{\mathcal{L}}$ and $\hat{\mathcal{S}}$ as

$$\hat{\mathcal{L}} := \left.\begin{bmatrix} \mathbf{0} & \cdots & \mathbf{0} & \mathcal{L} \\ \vdots & \ddots & \vdots & \vdots \\ \mathbf{0} & \cdots & \mathbf{0} & \mathcal{L} \\ \underbrace{\mathcal{L}^H \quad \cdots \quad \mathcal{L}^H}_{d \text{ times}} & & & \mathbf{0} \end{bmatrix}\right\} d \text{ times} \quad \text{and} \quad \hat{\mathcal{S}} := \left.\begin{bmatrix} \mathbf{0} & \cdots & \mathbf{0} & \mathcal{S} \\ \vdots & \ddots & \vdots & \vdots \\ \mathbf{0} & \cdots & \mathbf{0} & \mathcal{S} \\ \underbrace{\mathcal{S}^H \quad \cdots \quad \mathcal{S}^H}_{d \text{ times}} & & & \mathbf{0} \end{bmatrix}\right\} d \text{ times}.$$

It can be easily verified that $\hat{\mathcal{L}}$ is $\mathcal{O}(\mu)$-incoherence, and $\hat{\mathcal{S}}$ is $\mathcal{O}(\alpha)$-sparse where $\mathcal{O}(\mu)$ (resp. $\mathcal{O}(\alpha)$) means to hide the constant in front of μ (resp. α). Then our methods can be applied to asymmetric tensors with the same guarantee.

C. Extension of "Experiments"

C.1. Synthetic Data

C.1.1. Hyperparameter settings

We set $\beta = \frac{2\mu r}{n}, \beta_{\text{init}} = \frac{2\mu r \bar{\sigma}_1(\mathcal{L}^*)}{n\bar{\sigma}_1(\mathcal{D})}$ for APT and set $\beta = \frac{\mu r}{2n}, \beta_{\text{init}} = \frac{2\mu r \bar{\sigma}_1(\mathcal{L}^*)}{n\bar{\sigma}_1(\mathcal{D})}$ for EAPT. Transformed TRPCA and TRPCA use default hyperparameter $\lambda = 1/\sqrt{n}$ and $\lambda = 1/\sqrt{nq}$ respectively according to their theory. The weight of ETRPCA is set as $[\underbrace{1, \cdots, 1}_{167 \text{ times}}, \underbrace{1.1, \cdots, 1.1}_{167 \text{ times}}, \underbrace{1.5, \cdots, 1.5}_{166 \text{ times}}]$.

C.1.2. Robustness

We also conduct experiments to show the robustness of our methods under Gaussian noisy observation with different derivation σ in Table 4, i.e., the noisy observation $\mathcal{D}_n = \mathcal{D} + \mathcal{E}$ where $\mathcal{E}_{ijt} \sim N(0, \sigma^2)$. We set $n = 500, q = 5, r = 5, \alpha = 0.3$ in Table 4. As can be seen, our methods have smaller relative error and are more efficient than T-TRPCA.

C.2. HSI Denoising

Given a hyperspectral image $\mathcal{H} \in \mathbb{R}^{m \times n \times q}$, we permute the image into $\hat{\mathcal{H}} \in \mathbb{R}^{q \times m \times n}$ as shown in Figure 4 since the mode-1 unfolding matrix of $\hat{\mathcal{H}}$ has best low-rank property based on tensor-tensor product. So we apply this permutation to all methods based on tensor-tensor product.

Table 4. Relative error and running time of different methods under Gaussian noise with zero mean and different standard deviation.

	$\sigma = 0.5$		$\sigma = 1$		$\sigma = 1.5$	
	Rel Err	Time	Rel Err	Time	Rel Err	Time
T-TRPCA	0.1864	21.87	0.3413	21.54	0.4761	20.77
APT-DCT	**0.0476**	6.94	**0.0987**	7.11	0.1483	6.45
EAPT-DCT	**0.0476**	**1.76**	**0.0987**	**1.66**	**0.1480**	**1.80**

C.2.1. HYPERPARAMETER SETTINGS

We set $\lambda = 1/\sqrt{nq}$ of TRPCA and $\lambda = 1/\sqrt{n}$ for Transformed TRPCA. For ETRPCA, we set set $p = 0.9$, $w = [\underbrace{1, \cdots, 1}_{\text{5 times}}, \underbrace{1.1, \cdots, 1.1}_{\text{5 times}}, \underbrace{1.5, \cdots, 1.5}_{\text{21 times}}]$. We do not apply ket augmentation scheme to TTNN because the shape of our tensor is $512 \times 512 \times 31$ and set hyperparameters as $\lambda = 0.01, f = 0.01, \gamma = 0.001$ and $\delta = 0.001$. The parameter of SNN is set to $[40, 40, 40]$ because we find it performs well in most cases. λ and μ are set to be 0.5 and 0.1 in Atomic Norm. The rank of RTD is set to be 20 in this experiment. Here we empirically set the hyperparameters of EAPT as $\beta = \frac{2\mu r}{m+n}$, $\beta_{\text{init}} = 4\beta$, where we set $\mu = 10$. The multi-rank of our methods is set to be a descending sequence. For EAPT-FFT, the descending sequence of multi-rank is set to be $[\underbrace{4, \cdots, 4}_{\text{50 times}}, \underbrace{2, \cdots, 2}_{\text{50 times}}, 0, \cdots]$. For EAPT-DCT, the descending sequence of multi-rank is set to be $[\underbrace{4, \cdots, 4}_{\text{100 times}}, \underbrace{1, \cdots, 1}_{\text{200 times}}, 0, \cdots]$ and set to be $[\underbrace{5, \cdots, 5}_{\text{100 times}}, \underbrace{1, \cdots, 1}_{\text{200 times}}, 0, \cdots]$ for "sponges" data.

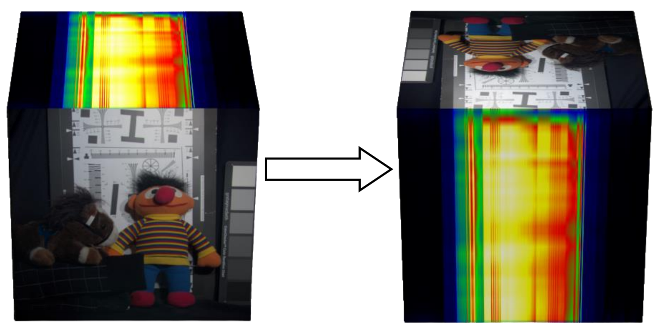

Figure 4. Illustration of the reshape scheme for hyperspectral image. The rank of the frontal slice for the right tensor is much lower than that of the frontal slice for the left tensor. According to Lemma D.2, t-SVD can better model the structure in the right tensor.

D. Proofs

D.1. Proof of Proposition 4.2

Here, we give the proof of Proposition 4.2.

Proof.

$$\begin{aligned} &P_{\mathbb{T}}(\mathcal{Z}) = \mathcal{U} \diamond_\Phi \mathcal{U}^H \diamond_\Phi \mathcal{Z} + \mathcal{Z} \diamond_\Phi \mathcal{V} \diamond_\Phi \mathcal{V}^H - \mathcal{U} \diamond_\Phi \mathcal{U}^H \diamond_\Phi \mathcal{Z} \diamond_\Phi \mathcal{V} \diamond_\Phi \mathcal{V}^H \\ =&\mathcal{U} \diamond_\Phi \mathcal{U}^H \diamond_\Phi \mathcal{Z} \diamond_\Phi (\mathcal{I} - \mathcal{V} \diamond_\Phi \mathcal{V}^H) + (\mathcal{I} - \mathcal{U} \diamond_\Phi \mathcal{U}^H) \diamond_\Phi \mathcal{Z} \diamond_\Phi \mathcal{V} \diamond_\Phi \mathcal{V}^H + \mathcal{U} \diamond_\Phi \mathcal{U}^H \diamond_\Phi \mathcal{Z} \diamond_\Phi \mathcal{V} \diamond_\Phi \mathcal{V}^H \\ =&\mathcal{U} \diamond_\Phi \mathcal{R}_2^H \diamond_\Phi \mathcal{Q}_2^H + \mathcal{Q}_1 \diamond_\Phi \mathcal{R}_1 \diamond_\Phi \mathcal{V}^H + \mathcal{U} \diamond_\Phi \mathcal{U}^H \diamond_\Phi \mathcal{Z} \diamond_\Phi \mathcal{V} \diamond_\Phi \mathcal{V}^H \\ =&\begin{bmatrix} \mathcal{U} & \mathcal{Q}_1 \end{bmatrix} \diamond_\Phi \begin{bmatrix} \mathcal{U}^H \diamond_\Phi \mathcal{Z} \diamond_\Phi \mathcal{B} & \mathcal{R}_2^H \\ \mathcal{R}_1 & \mathbf{0} \end{bmatrix} \diamond_\Phi \begin{bmatrix} \mathcal{V}^H \\ \mathcal{Q}_2^H \end{bmatrix} := \begin{bmatrix} \mathcal{U} & \mathcal{Q}_1 \end{bmatrix} \diamond_\Phi \mathcal{M} \diamond_\Phi \begin{bmatrix} \mathcal{V}^H \\ \mathcal{Q}_2^H \end{bmatrix} \end{aligned}$$

Then we have

$$P_{\mathbb{L}} \circ P_{\mathbb{T}}(\mathcal{Z}) = \begin{bmatrix} \mathcal{U} & \mathcal{Q}_1 \end{bmatrix} \diamond_\Phi P_{\mathbb{L}}(\mathcal{M}) \diamond_\Phi \begin{bmatrix} \mathcal{V}^H \\ \mathcal{Q}_2^H \end{bmatrix}.$$

□

Table 5. Sections and its corresponding lemmas / theorems / functions.

Section	Function	Lemma & Theorem
Basic lemmas	rank equivalence	Lemma D.2
	Weyl's inequality	Lemma D.3
	tensor sparse	Lemma D.4
initialization	—	Lemma D.5 and Theorem D.6
APT	singular value bound	Lemma D.21
	bound $\|\mathcal{L}-\mathcal{L}_k\|_\infty$	Lemma D.22
	bound $\|\mathcal{S}-\mathcal{S}_k\|_\infty$	Lemma D.23
	convergence	Theorem D.24
EAPT	projection & Trim	Lemma D.7 - D.13
	Some bounds	Lemma D.14 - D.16
	bound $\|\mathcal{L}-\mathcal{L}_k\|$	Lemma D.17
	bound $\|\mathcal{L}-\mathcal{L}_k\|_\infty$	Lemma D.18
	bound $\|\mathcal{S}-\mathcal{S}_k\|_\infty$	Lemma D.19
	convergence	Theorem D.20

D.2. Proof of Recovery Guarantees

The proof outline of our two methods follows as:

- In the t-th iteration, if $[\mathcal{S}]_{ijk}=0$, then $[\mathcal{S}_t]_{ijk}=0$.
  - If $[\mathcal{S}]_{ijk}=0$, then $T_{\zeta_t}(\mathcal{D}-\mathcal{L}_t)=T_{\zeta_t}(\mathcal{L}-\mathcal{L}_t)$. So we need $\zeta_t \geq \|\mathcal{L}-\mathcal{L}_t\|_\infty$.
  - $[\mathcal{S}-\mathcal{S}_t]_{ijk}$ is bounded by $\|\mathcal{L}-\mathcal{L}_t\|_\infty$ and ζ_t.
- Positive feedback in optimization process.
  - In $(t-1)$-th iteration, the hard thresholding parameter ζ_{t-1} can select $\mathcal{S}_{t-1}$ which includes all the zero elements in $\mathcal{S}$.
  - In the t-th iteration, the chosen $\mathcal{S}_{t-1}$ can make sure that $\|\mathcal{L}-\mathcal{L}_t\|_\infty$ is momently decreasing w.r.t. t.
  - Because $\|\mathcal{L}-\mathcal{L}_t\|_\infty$ is momently decreasing w.r.t. t, we can select smaller ζ_t.
  - Because $[\mathcal{S}-\mathcal{S}_t]_{ijk}$ is bounded by $\|\mathcal{L}-\mathcal{L}_t\|_\infty$ and ζ_t, we have smaller $[\mathcal{S}-\mathcal{S}_t]_{ijk}$.

We also present what the following theorems and lemmas do in the Table 5.

First, we give some notations which will be used in the proof. The optimal solution $\mathcal{L}^*$ and $\mathcal{S}^*$ are denoted here as $\mathcal{L}$ and $\mathcal{S}$ for simplicity. The (i,j)-th element of a matrix S is denote as $S_{i,j}$. $\mathcal{D}^{(c)}$ is shorthand for $\mathcal{D}(:,:,c)$. $\bar{\sigma}^{\mathcal{X}}_{i,c}$ denotes $\bar{\mathcal{S}}(i,i,c)$ where $\mathcal{X}=\mathcal{U}\diamond_\Phi \mathcal{S}\diamond_\Phi \mathcal{V}^H$ is transformed t-SVD of $\mathcal{X}$ and $\bar{\sigma}_{i,c}:=\bar{\sigma}^{\mathcal{L}}_{i,c}$.

We next defined the extension of eigenvalue in matrix to tensors as:

Definition D.1. Eigen fiber can been seen as an extension of eigenvalue in tensors. The definition of eigen fiber λ is:

$$\mathcal{D}\diamond_\Phi u = u\diamond_\Phi \lambda$$

where $\mathcal{D}\in\mathbb{R}^{n\times n\times q}$, $u\in\mathbb{R}^{n\times 1\times q}$, and $\lambda\in\mathbb{R}^{1\times 1\times q}$.

D.2.1. Basic Lemmas

Lemma D.2 (equivalence of block diagonal matrix and mode-1 unfolding matrix / rank equivalence)**.** *Let $\mathcal{X}\in\mathbb{R}^{m\times n\times q}$, the transformed t-product between two tensors is defined with unitary matrix Φ. Then*

$$\hat{\mathcal{X}}_\Phi(:,:,t)=\begin{bmatrix}\mathcal{X}(:,1,:) & \mathcal{X}(:,2,:) & \cdots & \mathcal{X}(:,n,:)\end{bmatrix}\cdot\begin{bmatrix}\Phi(t,:)^H & \mathbf{0} & \mathbf{0} & \mathbf{0}\\ \mathbf{0} & \Phi(t,:)^H & \mathbf{0} & \mathbf{0}\\ \vdots & \vdots & \ddots & \vdots\\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \Phi(t,:)^H\end{bmatrix}:=\mathcal{X}^{(1)}\tilde{\Phi}_t. \tag{8}$$

Proof. Assume $\mathcal{X} \in \mathbb{R}^{m\times n\times q}$ with

$$\hat{\mathcal{X}}_\Phi(j,k,t) = \sum_{c=1}^{q} \mathcal{X}(j,k,c)\Phi(t,c) = \mathcal{X}(j,k,:)\Phi(t,:)^H.$$

Then, we have

$$\hat{\mathcal{X}}_\Phi(:,:,t) = \mathcal{X}^{(1)}\tilde{\Phi}_t = \begin{bmatrix}\mathcal{X}(:,1,:) & \mathcal{X}(:,2,:) & \cdots & \mathcal{X}(:,n,:)\end{bmatrix} \cdot \begin{bmatrix} \Phi(t,:)^H & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \Phi(t,:)^H & \mathbf{0} & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \Phi(t,:)^H \end{bmatrix},$$

where $\mathcal{X}^{(1)} \in \mathbb{R}^{m\times nq}$ and $\tilde{\Phi}_t \in \mathbb{C}^{nq\times n}$.

□

Lemma D.3 (Tensor Weyl's inequality). *Let $\mathcal{A},\mathcal{B},\mathcal{C} \in \mathbb{R}^{n\times n\times q}$ be the symmetric tensors such that $\mathcal{A} = \mathcal{B} + \mathcal{C}$. Then, the inequality $|\bar{\sigma}_i^{\mathcal{A}} - \bar{\sigma}_i^{\mathcal{B}}| \leq \|\mathcal{C}\|$ holds for all i, where $\bar{\sigma}_i^{\mathcal{A}}$ and $\bar{\sigma}_i^{\mathcal{B}}$ represent the i^{th} singular values of $\bar{\mathcal{A}}$ and $\bar{\mathcal{B}}$ respectively.*

Proof. Because $\mathcal{A},\mathcal{B},\mathcal{C}$ are symmetric tensors, we know $\bar{\mathcal{A}}_\Phi, \bar{\mathcal{B}}_\Phi, \bar{\mathcal{C}}_\Phi$ are symmetric matrices such that $\bar{\mathcal{A}}_\Phi = \bar{\mathcal{B}}_\Phi + \bar{\mathcal{C}}_\Phi$. Following Weyl's inequality, we have $|\bar{\sigma}_i^{\mathcal{A}} - \bar{\sigma}_i^{\mathcal{B}}| \leq \|\bar{\mathcal{C}}_\Phi\| = \|\mathcal{C}\|$. □

Lemma D.4. *Let $\mathcal{S} \in \mathbb{R}^{n\times n\times q}$ satisfy assumption 1. Then, $\|\mathcal{S}\| \leq \alpha nq\|S\|_\infty$.*

Proof. Assume two unit vector $\mathbf{x} \in \mathbb{R}^n, \mathbf{y} \in \mathbb{R}^{nq}$,

$$\begin{aligned}\|\mathcal{S}\| = \left\|\bar{\mathcal{S}}\right\| &= \max_{c=1,\cdots,q} \left\|\hat{\mathcal{S}}_\Phi(:,:,c)\right\| = \max_{c=1,\cdots,q} \left\|\mathcal{S}^{(1)}\tilde{\Phi}_c\right\| = \|\mathcal{S}^{(1)}\| = \sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{c=1}^{q} \left(x_i \mathcal{S}_{ijc}^{(1)} y_{(j-1)q+c}\right) \\ &\leq \sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{c=1}^{q} \frac{1}{2}\left((x_i^2 + y_{(j-1)q+c}^2\right)\mathcal{S}_{ijc}^{(1)} \leq \frac{1}{2}\left(\alpha nq + \alpha nq\right)\|\mathcal{S}\|_\infty = \alpha nq\|\mathcal{S}\|_\infty,\end{aligned}$$

where the last inequality follows from Assumption 5.2. □

D.2.2. Initialization

Lemma D.5. *Let $\mathcal{S} \in \mathbb{R}^{n\times n\times q}$ be a sparse tensor satisfying Assumption 5.2. Let $\mathcal{U} \in \mathbb{R}^{n\times r\times q}$ be an orthogonal tensor with μ-incoherence, i.e., $\|\dot{e}_i^H * \mathcal{U}\|_F \leq \sqrt{\frac{\mu s_r}{nq}}$ for all i. Then $\|\dot{e}_i^H \mathcal{S}^a \mathcal{U}\|_F \leq \sqrt{\frac{\mu s_r}{nq}}(\alpha nq\sqrt{q}\|\mathcal{S}\|_\infty)^a$ for all i and $a \geq 0$.*

Proof. This proof is done by mathematical induction.
Base case: When $a = 0$, $\|\dot{e}_i^H \diamond_\Phi \mathcal{U}\|_F \leq \sqrt{\frac{\mu s_r}{nq}}$ is satisfied following from the assumption.
Induction Hypothesis: $\|\dot{e}_i^H \diamond_\Phi \mathcal{S}^a \diamond_\Phi \mathcal{U}\|_F \leq \max_l \sqrt{\frac{\mu s_r}{nq}}(\alpha nq\sqrt{q}\|\dot{e}_l^H \diamond_\Phi \mathcal{S}\|_F)^a$ for all i at the a^{th} power.

Induction Step: We have

$$\|\dot{e}_i^H \diamond_\Phi \mathcal{S}^{a+1} \diamond_\Phi \mathcal{U}\|_F^2 = \|\dot{e}_i^H \diamond_\Phi \mathcal{S} \diamond_\Phi \mathcal{S}^a \diamond_\Phi \mathcal{U}\|_F^2 = \|\bar{\dot{e}}_i^H \bar{\mathcal{S}}\bar{\mathcal{S}}^a\bar{\mathcal{U}}\|_F^2 = \sum_{c=0}^{q-1}\sum_{j=1}^{r}\left(\sum_{k=1}^{n}[\bar{\mathcal{S}}]_{cn+i,cn+k}[\bar{\mathcal{S}}^a\bar{\mathcal{U}}]_{cn+k,cr+j}\right)^2$$

$$=\sum_{c=0}^{q-1}\sum_{k_1,k_2}[\bar{\mathcal{S}}]_{cn+i,cn+k_1}[\bar{\mathcal{S}}]_{cn+i,cn+k_2}\sum_{j=1}^{r}[\bar{\mathcal{S}}^a\bar{\mathcal{U}}]_{cn+k_1,cr+j}[\bar{\mathcal{S}}^a\bar{\mathcal{U}}]_{cn+k_2,cr+j}$$

$$=\sum_{c=0}^{q-1}\sum_{k_1,k_2}[\bar{\mathcal{S}}]_{cn+i,cn+k_1}[\bar{\mathcal{S}}]_{cn+i,cn+k_2}\left\langle e_{cn+k_1}^H\bar{\mathcal{S}}^a\bar{\mathcal{U}}(:,cr+1:cr+r), e_{cn+k_2}^H\bar{\mathcal{S}}^a\bar{\mathcal{U}}(cr+1:cr+r)\right\rangle$$

$$\leq\sum_{c=0}^{q-1}\sum_{k_1,k_2}\left|[\bar{\mathcal{S}}]_{cn+i,cn+k_1}[\bar{\mathcal{S}}]_{cn+i,cn+k_2}\right|\|e_{cn+k_1}^H\bar{\mathcal{S}}^a\bar{\mathcal{U}}(:,cr+1:cr+r)\|_2\|e_{cn+k_2}^H\bar{\mathcal{S}}^a\bar{\mathcal{U}}(:,cr+1:cr+r)\|_2$$

$$\leq\sum_{k_1,k_2}\sum_{c=0}^{q-1}\left|[\bar{\mathcal{S}}]_{cn+i,cn+k_1}[\bar{\mathcal{S}}]_{cn+i,cn+k_2}\right|\|e_{cn+k_1}^H\bar{\mathcal{S}}^a\bar{\mathcal{U}}\|_2\|e_{cn+k_2}^H\bar{\mathcal{S}}^a\bar{\mathcal{U}}\|_2$$

$$\leq\|\dot{e}_l^H \diamond_\Phi \mathcal{S}^a \diamond_\Phi \mathcal{U}\|_F^2\sum_{c=0}^{q-1}\sum_{k_1,k_2}\left|[\bar{\mathcal{S}}]_{cn+i,cn+k_1}[\bar{\mathcal{S}}]_{cn+i,cn+k_2}\right|$$

$$=\|\dot{e}_l^H \diamond_\Phi \mathcal{S}^a \diamond_\Phi \mathcal{U}\|_F^2\sum_{c=1}^{q-1}\left(\|e_{cn+i}^H\bar{\mathcal{S}}(:,cn+1:cn+n)\|_1\right)^2 = \|\dot{e}_l^H \diamond_\Phi \mathcal{S}^a \diamond_\Phi \mathcal{U}\|_F^2\sum_{c=0}^{q-1}\left(\left\|\mathcal{S}^{(1)}(i,:)\tilde{\Phi}_{c+1}\right\|_1\right)^2$$

$$=\|\dot{e}_l^H \diamond_\Phi \mathcal{S}^a \diamond_\Phi \mathcal{U}\|_F^2\sum_{c=0}^{q-1}\left(\sum_{j=1}^{n}\left|\sum_{t=1}^{q}\mathcal{S}(i,j,t)\Phi(c+1,t)\right|\right)^2 \leq \|\dot{e}_l^H \diamond_\Phi \mathcal{S}^a \diamond_\Phi \mathcal{U}\|_F^2\sum_{c=0}^{q-1}\left(\sum_{j=1}^{n}\sum_{t=1}^{q}|\mathcal{S}(i,j,t)|\right)^2$$

$$\leq\|\dot{e}_l^H \diamond_\Phi \mathcal{S}^a \diamond_\Phi \mathcal{U}\|_F^2\sum_{c=0}^{q-1}(\alpha nq\|\mathcal{S}\|_\infty)^2 \leq \|\dot{e}_l^H \diamond_\Phi \mathcal{S}^a \diamond_\Phi \mathcal{U}\|_F^2(\alpha nq\sqrt{q}\|\mathcal{S}\|_\infty)^2 \leq \frac{\mu s_r}{nq}(\alpha nq\sqrt{q}\|\mathcal{S}\|_\infty)^{2(a+1)}.$$

Now, we have $\|\mathcal{U}^H \diamond_\Phi \mathcal{S}^{a+1} \diamond_\Phi \dot{e}_i\|_F \leq \max_l \sqrt{\frac{\mu s_r}{nq}}\left(\alpha nq\sqrt{q}\|\mathcal{S}\|_\infty\right)^{a+1}$. In the proof, we use the inequality

$$\|e_{cn+k_1}^T\bar{\mathcal{S}}^a\bar{\mathcal{U}}\|_2\|e_{cn+k_2}^T\bar{\mathcal{S}}^a\bar{\mathcal{U}}\|_2 \leq \|\bar{\dot{e}}_{k_1}^T\bar{\mathcal{S}}^a\bar{\mathcal{U}}\|_F\|\bar{\dot{e}}_{k_2}^T\bar{\mathcal{S}}^a\bar{\mathcal{U}}\|_F$$
$$= \|\dot{e}_{k_1}^T \diamond_\Phi \mathcal{S}^a \diamond_\Phi \mathcal{U}\|_F\|\dot{e}_{k_2}^T \diamond_\Phi \mathcal{S}^a \diamond_\Phi \mathcal{U}\|_F \leq \frac{\mu s_r}{nq}(\alpha nq\sqrt{q}\|\mathcal{S}\|_\infty)^{2a}.$$

□

Theorem D.6. *With the condition of Proposition 5.3, let $\mathcal{L} \in \mathbb{R}^{n\times n\times q}$ and $\mathcal{S} \in \mathbb{R}^{n\times n\times q}$ be two symmetric tensors satisfying Assumptions 5.1 and 5.2. If the thresholding parameters obey $\frac{\mu s_r\bar{\sigma}_1^{\mathcal{L}}}{nq\bar{\sigma}_1^{\mathcal{D}}} \leq \beta_{\text{init}} \leq \frac{3\mu s_r\bar{\sigma}_1^{\mathcal{L}}}{nq\bar{\sigma}_1^{\mathcal{D}}}$ and $\beta = \frac{\mu s_r}{2nq}$, then the outputs of Algorithm 1 satisfy*

$$\|\mathcal{L}-\mathcal{L}_0\| \leq 8\alpha\mu s_r\bar{\sigma}_1\left(\|\mathcal{L}-\mathcal{L}_0\|_\infty \leq \frac{\mu s_r}{4nq}\bar{\sigma}_1\right),\ \|\mathcal{S}-\mathcal{S}_0\|_\infty \leq \frac{\mu s_r}{nq}\bar{\sigma}_1, \quad \textit{and } supp(\mathcal{S}_0)\subset\Omega.$$

Proof. The proof can be partitioned into several parts.
(i) Note that $\mathcal{L}_{-1} = 0$ and

$$\|\mathcal{L}-\mathcal{L}_{-1}\|_\infty = \|\mathcal{L}\|_\infty = \max_{i,j,k}\left\langle \mathcal{U}\diamond_\Phi\Sigma\diamond_\Phi\mathcal{U}^H, \mathcal{E}_{ijk}\right\rangle = \max_{i,j,k}\left\langle \bar{\dot{e}}_i^H\bar{\mathcal{U}}\bar{\Sigma}\bar{\mathcal{U}}^H\bar{\dot{e}}_j, \bar{e}_k\right\rangle \leq \max_{i,j,k}\|\bar{\dot{e}}_i^H\bar{\mathcal{U}}\bar{\Sigma}\bar{\mathcal{U}}^H\bar{\dot{e}}_j\|_F\cdot\|\bar{e}_k\|_F$$
$$= \max_{i,j,k}\|\bar{\dot{e}}_i^H\bar{\mathcal{U}}\|_F\|\bar{\Sigma}\|\|\bar{\mathcal{U}}^H\bar{\dot{e}}_j\|_F \leq \max_{i,j,k}\|\dot{e}_i^H\diamond_\Phi\mathcal{U}\|_F\|\bar{\Sigma}\|\|\mathcal{U}^H\diamond_\Phi\dot{e}_j\|_F \leq \frac{\mu s_r}{nq}\bar{\sigma}_1,$$

where the last inequality follows from Assumption 5.1, i.e., $\mathcal{L}$ is μ-incoherence. Thus, with the choice of $\beta_{\text{init}} \geq \frac{\mu s_r\bar{\sigma}_1}{nq\bar{\sigma}_1^{\mathcal{D}}}$, we have $\|\mathcal{L}-\mathcal{L}_{-1}\|_\infty \leq \beta_{\text{init}}\bar{\sigma}_1^{\mathcal{D}} = \zeta_{-1}$. Since

$$[\mathcal{S}_{-1}]_{ijk} = [T_{\zeta_{-1}}(\mathcal{S}+\mathcal{L}-\mathcal{L}_{-1})]_{ijk} = \begin{cases} T_{\zeta_{-1}}([\mathcal{S}+\mathcal{L}-\mathcal{L}_{-1}]_{ijk}) & (i,j,k)\in\Omega \\ T_{\zeta_{-1}}([\mathcal{L}-\mathcal{L}_{-1}]_{ijk}) & (i,j,k)\in\Omega^c,\end{cases}$$

it follows that $[\mathcal{S}_{-1}]_{ijk} = 0$ for all $(i,j,k) \in \Omega^c$, i.e., $\Omega_{-1} := \text{supp}(\mathcal{S}_{-1}) \subset \Omega$. Moreover, for any entries of $\mathcal{S} - \mathcal{S}_{-1}$, we have

$$[\mathcal{S} - \mathcal{S}_{-1}]_{ijk} = \begin{cases} 0 \\ [\mathcal{L}_{-1} - \mathcal{L}]_{ijk} \\ [\mathcal{S}]_{ijk} \end{cases} \leq \begin{cases} 0 \\ \|\mathcal{L} - \mathcal{L}_{-1}\|_\infty \\ \|\mathcal{L} - \mathcal{L}_{-1}\|_\infty + \zeta_{-1} \end{cases} \leq \begin{cases} 0 & (i,j,k) \in \Omega^c \\ \frac{\mu s_r}{nq}\bar{\sigma}_1 & (i,j,k) \in \Omega_{-1} \\ \frac{4\mu s_r}{nq}\bar{\sigma}_1 & (i,j,k) \in \Omega \backslash \Omega_{-1} \end{cases},$$

where the last inequality follows from $\beta_{\text{init}} \leq \frac{3\mu s_r \bar{\sigma}_1}{nq\bar{\sigma}_1^{\mathcal{D}}}$, so that $\zeta_{-1} \leq \frac{3\mu s_r}{nq}\bar{\sigma}_1$. Therefore, we get

$$\text{supp}(\mathcal{S}_{-1}) \subset \Omega \quad \text{and} \quad \|\mathcal{S} - \mathcal{S}_{-1}\|_\infty \leq \frac{4\mu s_r}{nq}\bar{\sigma}_1.$$

By Lemma D.4, we also have $\|\mathcal{S} - \mathcal{S}_{-1}\|_2 \leq \alpha nq \|\mathcal{S} - \mathcal{S}_{-1}\|_\infty \leq 4\alpha\mu s_r \bar{\sigma}_1$. (ii) To bound the approximation error of $\mathcal{L}_0$ to $\mathcal{L}$ in terms of the spectral norm, note that

$$\begin{aligned} &\|\mathcal{L} - \mathcal{L}_0\| \leq \|\mathcal{L} - (\mathcal{D} - \mathcal{S}_{-1})\| + \|(\mathcal{D} - \mathcal{S}_{-1}) - \mathcal{L}_0\| \\ \leq &2\|\mathcal{L} - (\mathcal{D} - \mathcal{S}_{-1})\| = 2\|\mathcal{L} - (\mathcal{L} + \mathcal{S} - \mathcal{S}_{-1})\| = 2\|\mathcal{S} - \mathcal{S}_{-1}\|, \end{aligned}$$

where the second inequality follows from the fact $\mathcal{L}_0 = H_{\boldsymbol{r}}(\mathcal{D} - \mathcal{S}_{-1})$ is the best multi-rank $\boldsymbol{r}$ approximation of $\mathcal{D} - \mathcal{S}_{-1}$. It follows immediately that $\|\mathcal{L} - \mathcal{L}_0\| \leq 8\alpha\mu s_r \bar{\sigma}_1$. (iii) Since $\mathcal{D} = \mathcal{L} + \mathcal{S}$, we have $\mathcal{D} - \mathcal{S}_{-1} = \mathcal{L} + \mathcal{S} - \mathcal{S}_{-1}$. Let $\bar{\lambda}_i$ denotes the $(i,c)^{th}$ eigenvalue of $\bar{\mathcal{D}} - \bar{\mathcal{S}}_{-1}$ ordered by $|\bar{\lambda}_{1,c}| \geq |\bar{\lambda}_{2,c}| \geq \cdots \geq |\bar{\lambda}_{n,c}|$. The application of Tensor Weyl's inequality together with the bound of α implies that

$$|\bar{\sigma}_{i,c} - |\bar{\lambda}_{i,c}|| \leq \|\bar{\mathcal{S}} - \bar{\mathcal{S}}_{-1}\| \leq \frac{\bar{\sigma}_{s_r}}{8}$$

holds for all i. Consequently, we have

$$\frac{7}{8}\bar{\sigma}_{i,c} \leq |\bar{\lambda}_{i,c}| \leq \frac{9}{8}\bar{\sigma}_{i,c}, \quad \forall 1 \leq i \leq rq, \quad \frac{\|\mathcal{S} - \mathcal{S}_{-1}\|}{|\bar{\lambda}_{s_r}|} \leq \frac{1}{7}.$$

Let $\mathcal{D} - \mathcal{S}_{-1} = \begin{bmatrix}\mathcal{U}_0, \ddot{\mathcal{U}}_0\end{bmatrix} \diamond_\Phi \begin{bmatrix} \Lambda & 0 \\ 0 & \ddot{\Lambda} \end{bmatrix} \diamond_\Phi \begin{bmatrix}\mathcal{U}_0, \ddot{\mathcal{U}}_0\end{bmatrix}^H = \mathcal{U}_0 \diamond_\Phi \Lambda \diamond_\Phi \mathcal{U}_0^H + \ddot{\mathcal{U}}_0 \diamond_\Phi \ddot{\Lambda} \diamond_\Phi \ddot{\mathcal{U}}_0^H$ be its eigenvalue decomposition, where Λ has the multi-rank $\boldsymbol{r}$ largest eigenvalues in magnitude and $\ddot{\Lambda}$ contains the rest eigenvalues. Also, $\mathcal{U}_0$ contains the first $\boldsymbol{r}$ eigenvectors, and $\ddot{\mathcal{U}}_0$ has the rest. Notice $\mathcal{L}_0 = H_r(\mathcal{D} - \mathcal{S}_{-1}) = \mathcal{U}_0 \diamond_\Phi \Lambda \diamond_\Phi \mathcal{U}_0^H$ due to the symmetric setting. Denote $\mathcal{Z} = \mathcal{D} - \mathcal{S}_{-1} - \mathcal{L} = \mathcal{S} - \mathcal{S}_{-1}$. Let $u_i = \mathcal{U}(:, i, :)$ be the i^{th} eigenvector of $\mathcal{D} - \mathcal{S}_{-1} = \mathcal{L} + \mathcal{Z}$ and $\lambda_s = \Lambda(s, s, :)$ be the s^{th} eigen fiber of $\mathcal{D} - \mathcal{S}_{-1} = \mathcal{L} + \mathcal{Z}$. Then, we have $(\mathcal{D} - \mathcal{S}_{-1}) \diamond_\Phi u_i = u_i \diamond_\Phi \lambda_i$. We denote $\mathcal{D} - \mathcal{S}_{-1}$ as $\mathcal{M}$, then

$$\bar{\mathcal{M}}\bar{u}_i = \bar{u}_i\bar{\lambda}_i$$

which is

$$\begin{bmatrix} \bar{\mathcal{M}}^{(1)} & 0 & \cdots & 0 \\ 0 & \bar{\mathcal{M}}^{(2)} & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & \bar{\mathcal{M}}^{(q)} \end{bmatrix} \begin{bmatrix} \bar{u}_i^{(1)} & 0 & \cdots & 0 \\ 0 & \bar{u}_i^{(2)} & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & \bar{u}_i^{(q)} \end{bmatrix} = \begin{bmatrix} \bar{u}_i^{(1)} & 0 & \cdots & 0 \\ 0 & \bar{u}_i^{(2)} & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & \bar{u}_i^{(q)} \end{bmatrix} \begin{bmatrix} \bar{\lambda}_i^{(1)} & 0 & \cdots & 0 \\ 0 & \bar{\lambda}_i^{(2)} & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & \bar{\lambda}_i^{(q)} \end{bmatrix}.$$

This means $\bar{\mathcal{M}}^{(c)}\bar{u}_i^{(c)} = \bar{\lambda}_i^{(c)}\bar{u}_i^{(c)}$. Because $P_{T_k}(\mathcal{D} - \mathcal{S}_k)^{(c)} = \mathcal{Z}^{(c)} + \mathcal{L}^{(c)}$, we get $(\bar{\lambda}_i^{(c)} I - \bar{\mathcal{Z}}^{(c)})\bar{u}_i^{(c)} = \bar{\mathcal{L}}^{(c)}\bar{u}_i^{(c)}$. Then, we have

$$\bar{u}_i^{(c)} = \left(I - \frac{\bar{\mathcal{Z}}^{(c)}}{\bar{\lambda}_i^{(c)}}\right)^{-1} \frac{\bar{\mathcal{L}}^{(c)}}{\bar{\lambda}_i^{(c)}}\bar{u}_i^{(c)} = \left(I + \frac{\bar{\mathcal{Z}}^{(c)}}{\bar{\lambda}_i^{(c)}} + \left(\frac{\bar{\mathcal{Z}}^{(c)}}{\bar{\lambda}_i^{(c)}}\right)^2 + \cdots\right) \frac{\bar{\mathcal{L}}^{(c)}}{\bar{\lambda}_i^{(c)}}\bar{u}_i^{(c)}.$$

Combining all q slices, we have $\bar{u}_i = \left(I + \bar{\mathcal{E}}_i\bar{\mathcal{Z}} + (\bar{\mathcal{E}}_i\bar{\mathcal{Z}})^2 + \cdots\right)\bar{\mathcal{E}}_i\bar{\mathcal{L}}\bar{u}_i$ where

$$\bar{\mathcal{E}}_i = \begin{bmatrix} \frac{1}{\bar{\lambda}_i^{(1)}}I & 0 & \cdots & 0 \\ 0 & \frac{1}{\bar{\lambda}_i^{(2)}}I & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & \frac{1}{\bar{\lambda}_i^{(q)}}I \end{bmatrix}, \ \bar{\mathcal{Z}} = \begin{bmatrix} \bar{\mathcal{Z}}^{(1)} & 0 & \cdots & 0 \\ 0 & \bar{\mathcal{Z}}^{(2)} & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & \bar{\mathcal{Z}}^{(q)} \end{bmatrix}.$$

The above inequality is valid because of $\|\bar{\mathcal{E}}_i\bar{\mathcal{Z}}\| \leq \|\bar{\mathcal{Z}}\|\frac{1}{\bar{\lambda}_i^{(q)}} \leq \frac{\|\mathcal{Z}\|}{\bar{\lambda}_{sr}} \leq \frac{1}{7}$.

Then, we get $u_i = \left(\mathcal{I} + \mathcal{E}_i \diamond_\Phi \mathcal{Z} + (\mathcal{E}_i \diamond_\Phi \mathcal{Z})^2 + \cdots\right) \diamond_\Phi \mathcal{E}_i \diamond_\Phi \mathcal{L} \diamond_\Phi u_i$ for each u_i which implies

$$\begin{aligned}
&\mathcal{U}_0 \diamond_\Phi \Lambda \diamond_\Phi \mathcal{U}_0^H = \sum_{i=1}^r u_i \diamond_\Phi \lambda_i \diamond_\Phi u_i^H \\
=&\sum_{i=1}^r \left(\sum_{a\geq 0} (\mathcal{E}_i \diamond_\Phi \mathcal{Z})^a \diamond_\Phi \mathcal{E}_i \diamond_\Phi \mathcal{L}\right) \diamond_\Phi u_i \diamond_\Phi \lambda_i \diamond_\Phi u_i^H \left(\sum_{b\geq 0} (\mathcal{E}_i \diamond_\Phi \mathcal{Z})^b \diamond_\Phi \mathcal{E}_i \diamond_\Phi \mathcal{L}\right)^H \\
=&\sum_{a\geq 0} \mathcal{Z}^a \diamond_\Phi \mathcal{L} \diamond_\Phi \sum_{i=1}^r \left(\mathcal{E}_i^{a+1} \diamond_\Phi u_i \diamond_\Phi \lambda_i \diamond_\Phi u_i^H \diamond_\Phi \mathcal{E}_i^{b+1}\right) \diamond_\Phi \mathcal{L} \diamond_\Phi \sum_{b\geq 0} \mathcal{Z}^b.
\end{aligned}$$

To simplify the above formula, we have

$$\begin{aligned}
\bar{\mathcal{E}}_i\bar{u}_i &= \begin{bmatrix} \frac{1}{\bar{\lambda}_i^{(1)}}I & 0 & \cdots & 0 \\ 0 & \frac{1}{\bar{\lambda}_i^{(2)}}I & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & \frac{1}{\bar{\lambda}_i^{(q)}}I \end{bmatrix} \begin{bmatrix} \bar{u}_i^{(1)} & 0 & \cdots & 0 \\ 0 & \bar{u}_i^{(2)} & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & \bar{u}_i^{(q)} \end{bmatrix} \\
&= \begin{bmatrix} \bar{u}_i^{(1)} & 0 & \cdots & 0 \\ 0 & \bar{u}_i^{(2)} & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & \bar{u}_i^{(q)} \end{bmatrix} \begin{bmatrix} \frac{1}{\bar{\lambda}_i^{(1)}} & 0 & \cdots & 0 \\ 0 & \frac{1}{\bar{\lambda}_i^{(2)}} & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & \frac{1}{\bar{\lambda}_i^{(q)}} \end{bmatrix} = \bar{u}_i\bar{\mathcal{E}}_i.
\end{aligned}$$

Then, we have

$$\begin{aligned}
&\bar{\mathcal{E}}_i^{a+1}\bar{u}_i\bar{\lambda}_i\bar{u}_i^H\bar{\mathcal{E}}_i^{b+1} \\
&= \bar{u}_i \begin{bmatrix} \left(\frac{1}{\bar{\lambda}_i^{(1)}}\right)^{a+1} & 0 & \cdots & 0 \\ 0 & \left(\frac{1}{\bar{\lambda}_i^{(2)}}\right)^{a+1} & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & \left(\frac{1}{\bar{\lambda}_i^{(q)}}\right)^{a+1} \end{bmatrix} \begin{bmatrix} \bar{\lambda}_i^{(1)} & 0 & \cdots & 0 \\ 0 & \bar{\lambda}_i^{(2)} & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & \bar{\lambda}_i^{(q)} \end{bmatrix}. \\
&\qquad \begin{bmatrix} \left(\frac{1}{\bar{\lambda}_i^{(1)}}\right)^{b+1} & 0 & \cdots & 0 \\ 0 & \left(\frac{1}{\bar{\lambda}_i^{(2)}}\right)^{b+1} & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & \left(\frac{1}{\bar{\lambda}_i^{(q)}}\right)^{b+1} \end{bmatrix} \bar{u}_i^H \\
&= \bar{u}_i \begin{bmatrix} \left(\frac{1}{\bar{\lambda}_i^{(1)}}\right)^{a+b+1} & 0 & \cdots & 0 \\ 0 & \left(\frac{1}{\bar{\lambda}_i^{(2)}}\right)^{a+b+1} & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & \left(\frac{1}{\bar{\lambda}_i^{(q)}}\right)^{a+b+1} \end{bmatrix} \bar{u}_i^H = \bar{u}_i\bar{\Gamma}_i^{(a+b+1)}\bar{u}_i^H,
\end{aligned}$$

where we introduce new notation $\bar{\Gamma}_i^{(a+b+1)}$. Hence we have

$$\mathcal{E}_i^{a+1} \diamond_\Phi u_i \diamond_\Phi \lambda_i \diamond_\Phi u_i^H \diamond_\Phi \mathcal{E}_i^{b+1} = u_i \diamond_\Phi \Gamma_i^{(a+b+1)} \diamond_\Phi u_i^H.$$

Now, we get

$$\begin{aligned}\mathcal{U}_0 \diamond_\Phi \Lambda \diamond_\Phi \mathcal{U}_0^H &= \sum_{a\geq 0} \mathcal{Z}^a \diamond_\Phi \mathcal{L} \sum_{i=1}^{r} \left(\mathcal{E}_i^{a+1} \diamond_\Phi u_i \diamond_\Phi \lambda_i \diamond_\Phi u_i^H \diamond_\Phi \mathcal{E}_i^{b+1}\right) \diamond_\Phi \mathcal{L} \diamond_\Phi \sum_{b\geq 0} \mathcal{Z}^b \\ &= \sum_{a\geq 0} \mathcal{Z}^a \diamond_\Phi \mathcal{L} \diamond_\Phi \sum_{i=1}^{r} \left(u_i \diamond_\Phi \Gamma_i^{(a+b+1)} \diamond_\Phi u_i^H\right) \diamond_\Phi \mathcal{L} \diamond_\Phi \sum_{b\geq 0} \mathcal{Z}^b \\ &= \sum_{a,b\geq 0} \mathcal{Z}^a \diamond_\Phi \mathcal{L} \diamond_\Phi \mathcal{U}_0 \diamond_\Phi \Gamma^{(a+b+1)} \diamond_\Phi \mathcal{U}_0^H \diamond_\Phi \mathcal{L} \diamond_\Phi \mathcal{Z}^b.\end{aligned}$$

Thus, we have

$$\begin{aligned}&\|\mathcal{L}_0 - \mathcal{L}\|_\infty = \|\mathcal{U}_0 \diamond_\Phi \Lambda \diamond_\Phi \mathcal{U}_0^H - \mathcal{L}\|_\infty \\ =&\|\mathcal{L} \diamond_\Phi \mathcal{U}_0 \diamond_\Phi \Gamma^{(1)} \diamond_\Phi \mathcal{U}_0^H \diamond_\Phi \mathcal{L} - \mathcal{L} + \sum_{a+b>0} \mathcal{Z}^a \diamond_\Phi \mathcal{L} \diamond_\Phi \mathcal{U}_0 \diamond_\Phi \Gamma^{(a+b+1)} \diamond_\Phi \mathcal{U}_0^H \diamond_\Phi \mathcal{L} \diamond_\Phi \mathcal{Z}^b\|_\infty \\ \leq&\|\mathcal{L} \diamond_\Phi \mathcal{U}_0 \diamond_\Phi \Gamma^{(1)} \diamond_\Phi \mathcal{U}_0^H \diamond_\Phi \mathcal{L} - \mathcal{L}\|_\infty + \sum_{a+b>0} \|\mathcal{Z}^a \diamond_\Phi \mathcal{L} \diamond_\Phi \mathcal{U}_0 \diamond_\Phi \Gamma^{(a+b+1)} \diamond_\Phi \mathcal{U}_0^H \diamond_\Phi \mathcal{L} \diamond_\Phi \mathcal{Z}^b\|_\infty \\ :=&Y_0 + \sum_{a+b>0} Y_{ab}.\end{aligned}$$

We will handle Y_0 first. Recall that $\mathcal{L} = \mathcal{U} \diamond_\Phi \Sigma \diamond_\Phi \mathcal{V}^H$ is the t-SVD of the symmetric tensor $\mathcal{L}$ which obeys μ-incoherence, i.e., $\mathcal{U} \diamond_\Phi \mathcal{U}^H = \mathcal{V} \diamond_\Phi \mathcal{V}^H$ and $\|\dot{e}_i^H \diamond_\Phi \mathcal{U} \diamond_\Phi \mathcal{U}^H\|_F \leq \sqrt{\frac{\mu s_r}{nq}}$ for all i. So for each (i,j,k) entry of Y_0, one has

$$\begin{aligned}Y_0 =& \max_{i,j,k} \left|\left\langle \mathcal{L} \diamond_\Phi \mathcal{U}_0 \diamond_\Phi \Gamma^{(1)} \diamond_\Phi \mathcal{U}_0^H \diamond_\Phi \mathcal{L} - \mathcal{L}, \mathcal{E}_{ijk}\right\rangle\right| \\ =& \max_{i,j,k} \left|\left\langle \dot{e}_i^H \diamond_\Phi \mathcal{U} \diamond_\Phi \mathcal{U}^H (\mathcal{L} \diamond_\Phi \mathcal{U}_0 \diamond_\Phi \Gamma^{(1)} \diamond_\Phi \mathcal{U}_0^H \diamond_\Phi \mathcal{L} - \mathcal{L}) \diamond_\Phi \mathcal{U} \diamond_\Phi \mathcal{U}^H \diamond_\Phi \dot{e}_j, e_k\right\rangle\right| \\ \leq& \max_{i,j} \|\dot{e}_i^H \diamond_\Phi \mathcal{U} \diamond_\Phi \mathcal{U}^H (\mathcal{L} \diamond_\Phi \mathcal{U}_0 \diamond_\Phi \Gamma^{(1)} \diamond_\Phi \mathcal{U}_0^H \diamond_\Phi \mathcal{L} - \mathcal{L}) \diamond_\Phi \mathcal{U} \diamond_\Phi \mathcal{U}^H \diamond_\Phi \dot{e}_j\|_F \cdot \|e_k\|_F \\ \leq& \max_{i,j} \|\dot{e}_i^H \diamond_\Phi \mathcal{U} \diamond_\Phi \mathcal{U}^H\|_F \cdot \|\mathcal{L} \diamond_\Phi \mathcal{U}_0 \diamond_\Phi \Gamma^{(1)} \diamond_\Phi \mathcal{U}_0^H \diamond_\Phi \mathcal{L} - \mathcal{L}\| \cdot \|\mathcal{U} \diamond_\Phi \mathcal{U}^H \diamond_\Phi \dot{e}_j\|_F \\ \leq& \frac{\mu s_r}{nq} \|\mathcal{L} \diamond_\Phi \mathcal{U}_0 \diamond_\Phi \Gamma^{(1)} \diamond_\Phi \mathcal{U}_0^H \diamond_\Phi \mathcal{L} - \mathcal{L}\|,\end{aligned}$$

where the first inequality follows from the fact $\mathcal{U} \diamond_\Phi \mathcal{U}^H \diamond_\Phi \mathcal{L} = \mathcal{L} \diamond_\Phi \mathcal{U} \diamond_\Phi \mathcal{U}^H = \mathcal{L}$. Since $\mathcal{L} = \mathcal{U}_0 \diamond_\Phi \Lambda \diamond_\Phi \mathcal{U}_0^H + \ddot{\mathcal{U}}_0 \diamond_\Phi \ddot{\Lambda} \diamond_\Phi \ddot{\mathcal{U}}_0^H - \mathcal{Z}$, there holds that

$$\begin{aligned}&\|\mathcal{L} \diamond_\Phi \mathcal{U}_0 \diamond_\Phi \Gamma^{(1)} \diamond_\Phi \mathcal{U}_0^H \diamond_\Phi \mathcal{L} - \mathcal{L}\| \\ =&\|(\mathcal{U}_0 \diamond_\Phi \Lambda \diamond_\Phi \mathcal{U}_0^H + \ddot{\mathcal{U}}_0 \diamond_\Phi \ddot{\Lambda} \diamond_\Phi \ddot{\mathcal{U}}_0^H - \mathcal{Z}) \diamond_\Phi \mathcal{U}_0 \diamond_\Phi \Gamma^{(1)} \diamond_\Phi \mathcal{U}_0^H \diamond_\Phi (\mathcal{U}_0 \diamond_\Phi \Lambda \diamond_\Phi \mathcal{U}_0^H + \ddot{\mathcal{U}}_0 \diamond_\Phi \ddot{\Lambda} \diamond_\Phi \ddot{\mathcal{U}}_0^H - \mathcal{Z}) - \mathcal{L}\| \\ =&\|\mathcal{U}_0 \diamond_\Phi \Lambda \diamond_\Phi \Gamma^{(1)} \diamond_\Phi \Lambda \diamond_\Phi \mathcal{U}_0^H - \mathcal{L} - \mathcal{U}_0 \diamond_\Phi \Lambda \diamond_\Phi \Gamma^{(1)} \diamond_\Phi \mathcal{U}_0^H \diamond_\Phi \mathcal{Z} - \mathcal{Z} \diamond_\Phi \mathcal{U}_0 \diamond_\Phi \Gamma^{(1)} \diamond_\Phi \Lambda \diamond_\Phi \mathcal{U}_0^H \\ &+ \mathcal{Z} \diamond_\Phi \mathcal{U}_0 \diamond_\Phi \Gamma^{(1)} \diamond_\Phi \mathcal{U}_0^H \diamond_\Phi \mathcal{Z}\| \\ \leq&\|\mathcal{Z} - \ddot{\mathcal{U}}_0 \diamond_\Phi \ddot{\Lambda} \diamond_\Phi \ddot{\mathcal{U}}_0^H\| + \|\mathcal{U}_0 \diamond_\Phi \Lambda \diamond_\Phi \Gamma^{(1)} \diamond_\Phi \mathcal{U}_0^H \diamond_\Phi \mathcal{Z}\| + \|\mathcal{Z} \diamond_\Phi \mathcal{U}_0 \diamond_\Phi \Gamma^{(1)} \diamond_\Phi \Lambda \diamond_\Phi \mathcal{U}_0^H\| \\ &+ \|\mathcal{Z} \diamond_\Phi \mathcal{U}_0 \diamond_\Phi \Gamma^{(1)} \diamond_\Phi \mathcal{U}_0^H \diamond_\Phi \mathcal{Z}\| \\ =&\|\bar{\mathcal{Z}} - \bar{\ddot{\mathcal{U}}}_{k+1}\bar{\ddot{\Lambda}}\bar{\ddot{\mathcal{U}}}_{k+1}^H\| + \|\bar{\mathcal{U}}_0\bar{\Lambda}\bar{\Gamma}^{(1)}\bar{\mathcal{U}}_0^H\bar{\mathcal{Z}}\| + \|\bar{\mathcal{Z}}\bar{\mathcal{U}}_0\bar{\Gamma}^{(1)}\bar{\Lambda}\bar{\mathcal{U}}_0^H\| + \|\bar{\mathcal{Z}}\bar{\mathcal{U}}_0\bar{\Gamma}^{(1)}\bar{\mathcal{U}}_0^H\bar{\mathcal{Z}}\| \\ \leq&\|\bar{\mathcal{Z}} - \bar{\ddot{\mathcal{U}}}_{k+1}\bar{\ddot{\Lambda}}\bar{\ddot{\mathcal{U}}}_{k+1}^H\| + 2\|\bar{\mathcal{Z}}\| + \frac{\|Z\|^2}{|\bar{\lambda}_{s_r}|} \leq \|\bar{\ddot{\mathcal{U}}}_{k+1}\bar{\ddot{\Lambda}}\bar{\ddot{\mathcal{U}}}_{k+1}^H\| + 4\|\mathcal{Z}\| \leq |\bar{\lambda}_{rq+1}| + 4\|\mathcal{Z}\| \leq 5\|\mathcal{Z}\|,\end{aligned}$$

where the last inequality use the fact $\frac{\|\mathcal{Z}\|}{|\bar{\lambda}_{s_r}|} \leq \frac{1}{7}$ and $|\bar{\lambda}_{s_r+1}| \leq \|\mathcal{Z}\|$ because of $\bar{\sigma}_{rq+1} = 0$ and Weyl's inequality. Thus we have

$$Y_0 \leq \frac{5\mu s_r}{nq}\|\mathcal{Z}\| \leq 5\alpha\mu s_r\|\mathcal{Z}\|_\infty.$$

Next, we derive an upper bound for the rest part. Note that

$$
\begin{aligned}
&Y_{ab} = \max_{i,j,k} \left\langle \mathcal{Z}^a \diamond_\Phi \mathcal{L} \diamond_\Phi \mathcal{U}_0 \diamond_\Phi \Gamma^{(a+b+1)} \diamond_\Phi \mathcal{U}_0^H \diamond_\Phi \mathcal{L} \diamond_\Phi \mathcal{Z}^b, \mathcal{E}_{ijk} \right\rangle \\
=&\max_{i,j,k} \left\langle (\dot{e}_i^H \diamond_\Phi \mathcal{Z}^a \diamond_\Phi \mathcal{U} \diamond_\Phi \mathcal{U}^H) \diamond_\Phi \mathcal{L} \diamond_\Phi \mathcal{U}_0 \diamond_\Phi \Gamma^{(a+b+1)} \diamond_\Phi \mathcal{U}_0^H \diamond_\Phi \mathcal{L} \diamond_\Phi (\mathcal{U} \diamond_\Phi \mathcal{U}^H \diamond_\Phi \mathcal{Z}^b \diamond_\Phi \dot{e}_j), e_k \right\rangle \\
\le&\max_{i,j} \|(\dot{e}_i^H \diamond_\Phi \mathcal{Z}^a \diamond_\Phi \mathcal{U} \diamond_\Phi \mathcal{U}^H) \diamond_\Phi \mathcal{L} \diamond_\Phi \mathcal{U}_0 \diamond_\Phi \Gamma^{(a+b+1)} \diamond_\Phi \mathcal{U}_0^H \diamond_\Phi \mathcal{L} \diamond_\Phi (\mathcal{U} \diamond_\Phi \mathcal{U}^H \diamond_\Phi \mathcal{Z}^b \diamond_\Phi \dot{e}_j)\|_F \cdot \|e_k\|_F \\
\le&\max_{i,j} \|\dot{e}_i^H \diamond_\Phi \mathcal{Z}^a \diamond_\Phi \mathcal{U}\|_F \cdot \|\mathcal{L} \diamond_\Phi \mathcal{U}_0 \diamond_\Phi \Gamma^{(a+b+1)} \diamond_\Phi \mathcal{U}_0^H \diamond_\Phi \mathcal{L}\| \cdot \|\mathcal{U}^H \diamond_\Phi \mathcal{Z}^b \diamond_\Phi \dot{e}_j\|_F \\
\le&\max_{l} \frac{\mu s_r}{nq} (\alpha nq\sqrt{q}\|\mathcal{Z}\|_\infty)^{a+b} \|\mathcal{L} \diamond_\Phi \mathcal{U}_0 \diamond_\Phi \Gamma^{(a+b+1)} \diamond_\Phi \mathcal{U}_0^H \diamond_\Phi \mathcal{L}\| \\
\le&\alpha\mu s_r\sqrt{q}\|\mathcal{Z}\|_\infty \left(\frac{\bar{\sigma}_{s_r}}{8}\right)^{a+b-1} \|\mathcal{L} \diamond_\Phi \mathcal{U}_0 \diamond_\Phi \Gamma^{(a+b+1)} \diamond_\Phi \mathcal{U}_0^H \diamond_\Phi \mathcal{L}\|,
\end{aligned}
$$

where last inequality uses the bound of α ($\alpha \le \frac{1}{32\mu s_r \kappa\sqrt{q}}$ in Proposition 5.3). Furthermore, by using $\mathcal{L} = \mathcal{U}_0 \diamond_\Phi \Lambda \diamond_\Phi \mathcal{U}_0^H + \ddot{\mathcal{U}}_0 \diamond_\Phi \ddot{\Lambda} \diamond_\Phi \ddot{\mathcal{U}}_0^H - \mathcal{Z}$ again, we get

$$
\begin{aligned}
&\|\mathcal{L} \diamond_\Phi \mathcal{U}_0 \diamond_\Phi \Gamma^{(a+b+1)} \diamond_\Phi \mathcal{U}_0^H \diamond_\Phi \mathcal{L}\| \\
=&\|(\mathcal{U}_0 \diamond_\Phi \Lambda \diamond_\Phi \mathcal{U}_0^H + \ddot{\mathcal{U}}_0 \diamond_\Phi \ddot{\Lambda} \diamond_\Phi \ddot{\mathcal{U}}_0^H - \mathcal{Z}) \diamond_\Phi \mathcal{U}_0 \diamond_\Phi \Gamma^{(a+b+1)} \diamond_\Phi \mathcal{U}_0^H \diamond_\Phi (\mathcal{U}_0 \diamond_\Phi \Lambda \diamond_\Phi \mathcal{U}_0^H + \ddot{\mathcal{U}}_0 \diamond_\Phi \ddot{\Lambda} \diamond_\Phi \ddot{\mathcal{U}}_0^H - \mathcal{Z})\| \\
=&\|\mathcal{U}_0 \diamond_\Phi \Lambda \diamond_\Phi \Gamma^{(a+b+1)} \diamond_\Phi \Lambda \diamond_\Phi \mathcal{U}_0^H - \mathcal{U}_0 \diamond_\Phi \Lambda \diamond_\Phi \Gamma^{(a+b+1)} \diamond_\Phi \mathcal{U}_0^H \diamond_\Phi \mathcal{Z} - \mathcal{Z} \diamond_\Phi \mathcal{U}_0 \diamond_\Phi \Gamma^{(a+b+1)} \diamond_\Phi \Lambda \diamond_\Phi \mathcal{U}_0^H \\
&+ \mathcal{Z} \diamond_\Phi \mathcal{U}_0 \diamond_\Phi \Gamma^{(a+b+1)} \diamond_\Phi \mathcal{U}_0^H \diamond_\Phi \mathcal{Z}\| \\
\le&\|\bar{\mathcal{U}}_0\bar{\Lambda}\bar{\Gamma}^{(a+b+1)}\bar{\Lambda}\bar{\mathcal{U}}_0^H - \bar{\mathcal{U}}_0\bar{\Lambda}\bar{\Gamma}^{(a+b+1)}\bar{\mathcal{U}}_0^H\bar{\mathcal{Z}} - \bar{\mathcal{Z}}\bar{\mathcal{U}}_0\bar{\Gamma}^{(a+b+1)}\bar{\Lambda}\bar{\mathcal{U}}_0^H + \bar{\mathcal{Z}}\bar{\mathcal{U}}_0\bar{\Gamma}^{(a+b+1)}\bar{\mathcal{U}}_0^H\bar{\mathcal{Z}}\| \\
\le&|\bar{\lambda}_{s_r}|^{-(a+b-1)} + |\bar{\lambda}_{s_r}|^{-(a+b)}\|\mathcal{Z}\| + |\bar{\lambda}_{s_r}|^{-(a+b)}\|\mathcal{Z}\| + |\bar{\lambda}_{s_r}|^{-(a+b+1)}\|\mathcal{Z}\|^2 \\
=&|\bar{\lambda}_{s_r}|^{-(a+b-1)} \left(1 + \frac{2\|\mathcal{Z}\|}{\bar{\lambda}_{s_r}} + \left(\frac{\|\mathcal{Z}\|}{|\bar{\lambda}_{s_r}|}\right)^2\right) \\
=&|\bar{\lambda}_{s_r}|^{-(a+b-1)} \left(1 + \frac{\|\mathcal{Z}\|}{|\bar{\lambda}_{s_r}|}\right)^2 \le 2|\bar{\lambda}_{s_r}|^{-(a+b-1)} \le 2\left(\frac{7}{8}\bar{\sigma}_{s_r}\right)^{-(a+b-1)}.
\end{aligned}
$$

Then, we have

$$
\begin{aligned}
\sum_{a+b>0} Y_{ab} \le \sum_{a+b>0} 2\alpha\mu s_r\sqrt{q}\|\mathcal{Z}\|_\infty \left(\frac{\frac{1}{8}\bar{\sigma}_{s_r}}{\frac{7}{8}\bar{\sigma}_{s_r}}\right)^{a+b-1} \le 2\alpha\mu s_r\sqrt{q}\|\mathcal{Z}\|_\infty \sum_{a+b>0} \left(\frac{1}{7}\right)^{a+b-1} \\
\le 2\alpha\mu s_r\sqrt{q}\|\mathcal{Z}\|_\infty \left(\frac{1}{1-\frac{1}{7}}\right)^2 \le 3\alpha\mu s_r\sqrt{q}\|\mathcal{Z}\|_\infty.
\end{aligned}
$$

Finally, combining them together gives

$$
\|\mathcal{L}_0 - \mathcal{L}\|_\infty = Y_0 + \sum_{a+b>0} Y_{ab} \le 5\alpha\mu s_r\|\mathcal{Z}\|_\infty + 3\alpha\mu s_r\sqrt{q}\|\mathcal{Z}\|_\infty \le \frac{\mu s_r}{4nq}\bar{\sigma}_1,
$$

where the last inequality uses the bound of α ($\alpha \le \frac{1}{128\mu s_r\sqrt{q}}$) in Proposition 5.3.
(iv) From the thresholding rule, we know that

$$
[\mathcal{S}_0]_{ijk} = [T_{\zeta_0}(\mathcal{S} + \mathcal{L} - \mathcal{L}_0)]_{ijk} = \begin{cases} T_{\zeta_0}([\mathcal{S} + \mathcal{L} - \mathcal{L}_0]_{ijk}) & (i,j,k) \in \Omega \\ T_{\zeta_0}([\mathcal{L} - \mathcal{L}_0]_{ijk}) & (i,j,k) \in \Omega^c \end{cases}.
$$

So the above inequality and $\zeta_0 = \frac{\mu s_r}{2nq}\bar{\lambda}_1$ imply $[\mathcal{S}]_{ijk} = 0$ for all $(i,j,k) \in \Omega^c$, i.e., supp$(\mathcal{S}_0) := \Omega_0 \subset \Omega$. Also, for any entries of $\mathcal{S} - \mathcal{S}_0$, there holds that

$$
[\mathcal{S} - \mathcal{S}_0]_{ijk} = \begin{cases} 0 \\ [\mathcal{L}_0 - \mathcal{L}]_{ijk} \\ [\mathcal{S}]_{ijk} \end{cases} \le \begin{cases} 0 \\ \|\mathcal{L} - \mathcal{L}_0\|_\infty \\ \|\mathcal{L} - \mathcal{L}_0\|_\infty + \zeta_0 \end{cases} \le \begin{cases} 0 & (i,j,k) \in \Omega^c \\ \frac{\mu s_r}{4nq}\bar{\sigma}_1 & (i,j,k) \in \Omega_0 \\ \frac{\mu s_r}{nq}\bar{\sigma}_1 & (i,j,k) \in \Omega \backslash \Omega_0 \end{cases}.
$$

Here the last inequality implies $\zeta_0 = \frac{\mu s_r}{2nq}\bar{\lambda}_1 \leq \frac{3\mu s_r}{4nq}\bar{\sigma}_1$. Therefore, we have

$$\text{supp}(\mathcal{S}_0) \subset \Omega \quad \text{and} \quad \|\mathcal{S} - \mathcal{S}_0\|_\infty \leq \frac{\mu s_r}{nq}\bar{\sigma}_1.$$

□

D.2.3. Local Convergence of Alternating Projection in EAPT

Lemma D.7 (Bounds for Projections). *Let $\mathcal{L}_k = \mathcal{U}_k \diamond_\Phi \Sigma_k \diamond_\Phi \mathcal{V}_k^H$ be a multi-rank $\boldsymbol{k}$ tensor, and T_k be the tangent space of the multi-rank $\boldsymbol{k}$ tensor manifold at $\mathcal{L}_k$. The implicit rank of $\mathcal{L}_k$ is s_k and the tubal rank is k. Let $\mathcal{L} = \mathcal{U} \diamond_\Phi \Sigma \diamond_\Phi \mathcal{V}^H$ be another multi-rank $\boldsymbol{k}$ tensor, and T be the corresponding tangent space. Then,*

$$\|\mathcal{U}_k \diamond_\Phi \mathcal{U}_k^H - \mathcal{U} \diamond_\Phi \mathcal{U}^H\| \leq \frac{\|\mathcal{L}_k - \mathcal{L}\|}{\sigma_{\min}(\mathcal{L})}, \quad \|\mathcal{V}_k \diamond_\Phi \mathcal{V}_k^H - \mathcal{V} \diamond_\Phi \mathcal{V}^H\| \leq \frac{\|\mathcal{L}_k - \mathcal{L}\|}{\sigma_{\min}(\mathcal{L})} \tag{9}$$

$$\|\mathcal{U}_k \diamond_\Phi \mathcal{U}_k^H - \mathcal{U} \diamond_\Phi \mathcal{U}^H\|_F \leq \frac{\sqrt{2}\|\mathcal{L}_k - \mathcal{L}\|_F}{\sigma_{\min}(\mathcal{L})}, \quad \|\mathcal{V}_k \diamond_\Phi \mathcal{V}_k^H - \mathcal{V} \diamond_\Phi \mathcal{V}^H\|_F \leq \frac{\sqrt{2}\|\mathcal{L}_k - \mathcal{L}\|_F}{\sigma_{\min}(\mathcal{L})} \tag{10}$$

$$\|(I - P_{T_k})\mathcal{L}\|_F \leq \frac{\|\mathcal{L}_k - \mathcal{L}\|_F^2}{\sigma_{\min}(\mathcal{L})}, \quad \|P_{T_k} - P_T\| \leq \frac{2\|\mathcal{L}_k - \mathcal{L}\|_F}{\sigma_{\min}(\mathcal{L})}. \tag{11}$$

Proof. We only prove the left inequalities of Eqs (9) and Eq.(10) because the right inequalities can be easily established. The left inequality of Eq.(9) follows directly from calculations

$$\begin{aligned}
&\|\mathcal{U}_k \diamond_\Phi \mathcal{U}_k^H - \mathcal{U} \diamond_\Phi \mathcal{U}^H\| = \|\bar{\mathcal{U}}\bar{\mathcal{U}}^H(\bar{\mathcal{U}}\bar{\mathcal{U}}^H - \bar{\mathcal{U}}_k\bar{\mathcal{U}}_k^H)\| = \|\bar{\mathcal{U}}\bar{\mathcal{U}}^H(\bar{\mathcal{I}} - \bar{\mathcal{U}}_k\bar{\mathcal{U}}_k^H)\| = \|(\bar{\mathcal{I}} - \bar{\mathcal{U}}_k\bar{\mathcal{U}}_k^H)\bar{\mathcal{U}}\bar{\mathcal{U}}^H\| \\
=&\|(\bar{\mathcal{I}} - \bar{\mathcal{U}}_k\bar{\mathcal{U}}_k^H)\bar{\mathcal{L}}\bar{\mathcal{V}}\bar{\Sigma}^{-1}\bar{\mathcal{U}}^H\| = \|(\bar{\mathcal{I}} - \bar{\mathcal{U}}_k\bar{\mathcal{U}}_k^H)(\bar{\mathcal{L}}_k - \bar{\mathcal{L}})\bar{\mathcal{V}}\bar{\Sigma}^{-1}\bar{\mathcal{U}}^H\| \\
\leq&\|\bar{\mathcal{I}} - \bar{\mathcal{U}}_k\bar{\mathcal{U}}_k^H\| \cdot \|\bar{\mathcal{L}}_k - \bar{\mathcal{L}}\| \cdot \|\bar{\mathcal{V}}\| \cdot \|\bar{\Sigma}^{-1}\| \cdot \|\bar{\mathcal{U}}^H\| \leq \frac{\|\bar{\mathcal{L}}_k - \bar{\mathcal{L}}\|}{\sigma_{\min}(\bar{\mathcal{L}})} = \frac{\|\mathcal{L}_k - \mathcal{L}\|}{\sigma_{\min}(\mathcal{L})}.
\end{aligned}$$

To prove the left inequality of Eq.(10), we first show that

$$\|(\mathcal{I} - \mathcal{U}_k \diamond_\Phi \mathcal{U}_k^H) \diamond_\Phi \mathcal{U} \diamond_\Phi \mathcal{U}^H\|_F = \|\mathcal{U}_k \diamond_\Phi \mathcal{U}_k^H \diamond_\Phi (\mathcal{I} - \mathcal{U} \diamond_\Phi \mathcal{U}^H)\|_F.$$

Eq.(10) can be obtained by noting that

$$\begin{aligned}
\|(\mathcal{I} - \mathcal{U}_k \diamond_\Phi \mathcal{U}_k^H) \diamond_\Phi \mathcal{U} \diamond_\Phi \mathcal{U}^H\|_F^2 &= \|(\bar{\mathcal{I}} - \bar{\mathcal{U}}_k\bar{\mathcal{U}}_k^H)\bar{\mathcal{U}}\bar{\mathcal{U}}^H\|_F^2 = \left\langle (\bar{\mathcal{I}} - \bar{\mathcal{U}}_k\bar{\mathcal{U}}_k^H)\bar{\mathcal{U}}\bar{\mathcal{U}}^H, (\bar{\mathcal{I}} - \bar{\mathcal{U}}_k\bar{\mathcal{U}}_k^H)\bar{\mathcal{U}}\bar{\mathcal{U}}^H \right\rangle, \\
&= \left\langle \bar{\mathcal{U}}_k\bar{\mathcal{U}}_k^H, \bar{\mathcal{I}} - \bar{\mathcal{U}}\bar{\mathcal{U}}^H \right\rangle = s_k - \left\langle \bar{\mathcal{U}}_k\bar{\mathcal{U}}_k^H, \bar{\mathcal{U}}\bar{\mathcal{U}}^H \right\rangle,
\end{aligned}$$

and

$$\begin{aligned}
&\|\mathcal{U}_k \diamond_\Phi \mathcal{U}_k^H \diamond_\Phi (\mathcal{I} - \mathcal{U} \diamond_\Phi \mathcal{U}^H)\|_F^2 = \|\bar{\mathcal{U}}_k\bar{\mathcal{U}}_k^H(\bar{\mathcal{I}} - \bar{\mathcal{U}}\bar{\mathcal{U}}^H)\|_F^2 \\
=&\left\langle \bar{\mathcal{U}}_k\bar{\mathcal{U}}_k^H(\bar{\mathcal{I}} - \bar{\mathcal{U}}\bar{\mathcal{U}}^H), \bar{\mathcal{U}}_k\bar{\mathcal{U}}_k^H(\bar{\mathcal{I}} - \bar{\mathcal{U}}\bar{\mathcal{U}}^H) \right\rangle \\
=&\left\langle \bar{\mathcal{I}} - \bar{\mathcal{U}}_k\bar{\mathcal{U}}_k^H, \bar{\mathcal{U}}\bar{\mathcal{U}}^H \right\rangle = s_k - \left\langle \bar{\mathcal{U}}_k\bar{\mathcal{U}}_k^H, \bar{\mathcal{U}}\bar{\mathcal{U}}^H \right\rangle.
\end{aligned}$$

Then, it follows that

$$\begin{aligned}
&\|\mathcal{U}_k \diamond_\Phi \mathcal{U}_k^H - \mathcal{U} \diamond_\Phi \mathcal{U}^H\|_F \leq \sqrt{2}\|(\mathcal{I} - \mathcal{U}_k \diamond_\Phi \mathcal{U}_k^H) \diamond_\Phi \mathcal{U} \diamond_\Phi \mathcal{U}^H\|_F \\
=&\sqrt{2}\|(\bar{\mathcal{I}} - \bar{\mathcal{U}}_k\bar{\mathcal{U}}_k^H)\bar{\mathcal{L}}\bar{\mathcal{V}}\bar{\Sigma}^{-1}\bar{\mathcal{U}}^H\|_F = \sqrt{2}\|(\bar{\mathcal{I}} - \bar{\mathcal{U}}_k\bar{\mathcal{U}}_k^H)(\bar{\mathcal{L}}_k - \bar{\mathcal{L}})\bar{\mathcal{V}}\bar{\Sigma}^{-1}\bar{\mathcal{U}}^H\|_F \\
\leq&\sqrt{2}\|\bar{\mathcal{I}} - \bar{\mathcal{U}}_k\bar{\mathcal{U}}_k^H\| \cdot \|\bar{\mathcal{L}}_k - \bar{\mathcal{L}}\|_F \cdot \|\bar{\mathcal{V}}\| \cdot \|\bar{\Sigma}^{-1}\| \cdot \|\bar{\mathcal{U}}^H\| \\
\leq&\frac{\sqrt{2}\|\bar{\mathcal{L}}_k - \bar{\mathcal{L}}\|_F}{\sigma_{\min}(\bar{\mathcal{L}})} = \frac{\sqrt{2}\|\mathcal{L}_k - \mathcal{L}\|_F}{\sigma_{\min}(\mathcal{L})}.
\end{aligned}$$

Next, we prove the left inequality of Eq.(11). First we note that $P_T(\mathcal{L}) = \mathcal{L}$. So

$$\begin{aligned}
&(I - P_{T_k})\mathcal{L} = (P_T - P_{T_k})\mathcal{L} \\
=&(\mathcal{U} \diamond_\Phi \mathcal{U}^H - \mathcal{U}_k \diamond_\Phi \mathcal{U}_k^H) \diamond_\Phi \mathcal{L} + \mathcal{L} \diamond_\Phi (\mathcal{V} \diamond_\Phi \mathcal{V}^H - \mathcal{V}_k \diamond_\Phi \mathcal{V}_k^H) \\
&- (\mathcal{U} \diamond_\Phi \mathcal{U}^H - \mathcal{U}_k \diamond_\Phi \mathcal{U}_k^H) \diamond_\Phi \mathcal{L} \diamond_\Phi \mathcal{V}_k \diamond_\Phi \mathcal{V}_k^H - \mathcal{U} \diamond_\Phi \mathcal{U}^H \diamond_\Phi \mathcal{L} \diamond_\Phi (\mathcal{V} \diamond_\Phi \mathcal{V}^H - \mathcal{V}_k \diamond_\Phi \mathcal{V}_k^H) \\
=&(\mathcal{U} \diamond_\Phi \mathcal{U}^H - \mathcal{U}_k \diamond_\Phi \mathcal{U}_k^H) \diamond_\Phi \mathcal{L} \diamond_\Phi (\mathcal{I} - \mathcal{V}_k \diamond_\Phi \mathcal{V}_k^H) + (\mathcal{I} - \mathcal{U} \diamond_\Phi \mathcal{U}^H) \diamond_\Phi \mathcal{L} \diamond_\Phi (\mathcal{V} \diamond_\Phi \mathcal{V}^H - \mathcal{V}_k \diamond_\Phi \mathcal{V}_k^H) \\
=&(\mathcal{U} \diamond_\Phi \mathcal{U}^H - \mathcal{U}_k \diamond_\Phi \mathcal{U}_k^H) \diamond_\Phi \mathcal{L} \diamond_\Phi (\mathcal{I} - \mathcal{V}_k \diamond_\Phi \mathcal{V}_k^H) = (\mathcal{U} \diamond_\Phi \mathcal{U}^H - \mathcal{U}_k \diamond_\Phi \mathcal{U}_k^H) \diamond_\Phi (\mathcal{L} - \mathcal{L}_k) \diamond_\Phi (\mathcal{I} - \mathcal{V}_k \diamond_\Phi \mathcal{V}_k^H),
\end{aligned}$$

where the last two inequalities follow from the fact $(\mathcal{I} - \mathcal{U} \diamond_\Phi \mathcal{U}^H) \diamond_\Phi \mathcal{L} = 0$ and $\mathcal{L}_k \diamond_\Phi (\mathcal{I} - \mathcal{V}_k \diamond_\Phi \mathcal{V}_k^H) = 0$. Taking the Frobenius norm on both sides gives

$$\|(I - P_{T_k})\mathcal{L}\|_F \le \|\mathcal{U} \diamond_\Phi \mathcal{U}^H - \mathcal{U}_k \diamond_\Phi \mathcal{U}_k^H\| \cdot \|\mathcal{L}_k - \mathcal{L}\|_F \cdot \|\mathcal{I} - \mathcal{V}_k \diamond_\Phi \mathcal{V}_k^H\| \le \frac{\|\mathcal{L}_k - \mathcal{L}\|_F^2}{\sigma_{\min}(\mathcal{L})}.$$

Now, we give the proof of the right side of Eq.(11). For any tensor $\mathcal{Z}$, we have

$$\begin{aligned}
&(P_{T_k} - P_T)\mathcal{Z} = \mathcal{U}_k \diamond_\Phi \mathcal{U}_k^H \diamond_\Phi \mathcal{Z} + \mathcal{Z} \diamond_\Phi \mathcal{V}_k \diamond_\Phi \mathcal{V}_k^H - \mathcal{U}_k \diamond_\Phi \mathcal{U}_k^H \diamond_\Phi \mathcal{Z} \diamond_\Phi \mathcal{V}_k \diamond_\Phi \mathcal{V}_k^H \\
&- \mathcal{U} \diamond_\Phi \mathcal{U}^H \diamond_\Phi \mathcal{Z} - \mathcal{Z} \diamond_\Phi \mathcal{V} \diamond_\Phi \mathcal{V}^H + \mathcal{U} \diamond_\Phi \mathcal{U}^H \diamond_\Phi \mathcal{Z} \diamond_\Phi \mathcal{V} \diamond_\Phi \mathcal{V}^H \\
=&(\mathcal{U}_k \diamond_\Phi \mathcal{U}_k^H - \mathcal{U} \diamond_\Phi \mathcal{U}^H) \diamond_\Phi \mathcal{Z} + \mathcal{Z} \diamond_\Phi (\mathcal{V}_k \diamond_\Phi \mathcal{V}_k^H - \mathcal{V} \diamond_\Phi \mathcal{V}^H) \\
&- (\mathcal{U}_k \diamond_\Phi \mathcal{U}_k^H - \mathcal{U} \diamond_\Phi \mathcal{U}^H) \diamond_\Phi \mathcal{Z} \diamond_\Phi \mathcal{V} \diamond_\Phi \mathcal{V}^H - \mathcal{U}_k \diamond_\Phi \mathcal{U}_k^H \diamond_\Phi \mathcal{Z} \diamond_\Phi (\mathcal{V}_k \diamond_\Phi \mathcal{V}_k^H - \mathcal{V} \diamond_\Phi \mathcal{V}^H) \\
=&(\mathcal{U}_k \diamond_\Phi \mathcal{U}_k^H - \mathcal{U} \diamond_\Phi \mathcal{U}^H) \diamond_\Phi \mathcal{Z} \diamond_\Phi (\mathcal{I} - \mathcal{V} \diamond_\Phi \mathcal{V}^H) + (\mathcal{I} - \mathcal{U}_k \diamond_\Phi \mathcal{U}_k^H) \diamond_\Phi \mathcal{Z} \diamond_\Phi (\mathcal{V}_k \diamond_\Phi \mathcal{V}_k^H - \mathcal{V} \diamond_\Phi \mathcal{V}^H).
\end{aligned}$$

Taking Frobenius norm on both sides gives

$$\begin{aligned}
&\|(P_{T_k} - P_T)\mathcal{Z}\|_F \\
\le&\|\mathcal{U}_k \diamond_\Phi \mathcal{U}_k^H - \mathcal{U} \diamond_\Phi \mathcal{U}^H\| \cdot \|\mathcal{Z}\|_F \cdot \|\mathcal{I} - \mathcal{V} \diamond_\Phi \mathcal{V}^H\| + \|\mathcal{I} - \mathcal{U}_k \diamond_\Phi \mathcal{U}_k^H\| \cdot \|\mathcal{Z}\|_F \cdot \|\mathcal{V}_k \diamond_\Phi \mathcal{V}_k^H - \mathcal{V} \diamond_\Phi \mathcal{V}^H\| \\
\le&\frac{2\|\mathcal{L}_k - \mathcal{L}\|_F}{\sigma_{\min}(\mathcal{L})}\|\mathcal{Z}\|_F.
\end{aligned}$$

□

Lemma D.8 (Chordal and Projection Distances: the matrix version in (Wei et al., 2016))**.** *Let $\mathcal{U}_k, \mathcal{U} \in \mathbb{R}^{n\times k\times q}$ be two orthogonal tensors of multi-rank $\boldsymbol{k}$. Then there exists a $k \times k \times q$ unitary tensor $\mathcal{Q}$ such that*

$$\|\mathcal{U}_k - \mathcal{U} \diamond_\Phi \mathcal{Q}\|_F \le \|\mathcal{U}_k \diamond_\Phi \mathcal{U}_k^H - \mathcal{U} \diamond_\Phi \mathcal{U}^H\|_F.$$

Proof. Since

$$\|\mathcal{U}_k - \mathcal{U} \diamond_\Phi \mathcal{Q}\|_F^2 = \left\langle \bar{\mathcal{U}}_k - \bar{\mathcal{U}}\bar{\mathcal{Q}}, \bar{\mathcal{U}}_k - \bar{\mathcal{U}}\bar{\mathcal{Q}} \right\rangle = 2s_k - 2\left\langle \mathcal{U}_k, \mathcal{U} \diamond_\Phi \mathcal{Q} \right\rangle = 2s_k - 2\left\langle \bar{\mathcal{U}}_k, \bar{\mathcal{U}}\bar{\mathcal{Q}} \right\rangle$$

and

$$\begin{aligned}
\|\mathcal{U}_k \diamond_\Phi \mathcal{U}_k^H - \mathcal{U} \diamond_\Phi \mathcal{U}^H\|_F =& \left\langle \bar{\mathcal{U}}_k\bar{\mathcal{U}}_k^H - \bar{\mathcal{U}}\bar{\mathcal{U}}^H, \bar{\mathcal{U}}_k\bar{\mathcal{U}}_k^H - \bar{\mathcal{U}}\bar{\mathcal{U}}^H \right\rangle \\
=&2s_k - 2\left\langle \mathcal{U}_k \diamond_\Phi \mathcal{U}_k^H, \mathcal{U} \diamond_\Phi \mathcal{U}^H \right\rangle = 2s_k - 2\left\langle \bar{\mathcal{U}}_k\bar{\mathcal{U}}_k^H, \bar{\mathcal{U}}\bar{\mathcal{U}}^H \right\rangle,
\end{aligned}$$

it suffices to show that there exists a $\mathcal{Q}$ such that

$$\left\langle \mathcal{U}_k, \mathcal{U} \diamond_\Phi \mathcal{Q} \right\rangle \ge \left\langle \mathcal{U}_k \diamond_\Phi \mathcal{U}_k^H, \mathcal{U} \diamond_\Phi \mathcal{U}^H \right\rangle \Leftrightarrow \left\langle \bar{\mathcal{U}}_k, \bar{\mathcal{U}}\bar{\mathcal{Q}} \right\rangle \ge \left\langle \bar{\mathcal{U}}_k\bar{\mathcal{U}}_k^H, \bar{\mathcal{U}}\bar{\mathcal{U}}^H \right\rangle.$$

It is equivalent to show that

$$\left\langle \mathcal{U}^H \diamond_\Phi \mathcal{U}_k, \mathcal{Q} \right\rangle \ge \left\langle \mathcal{U}^H \diamond_\Phi \mathcal{U}_k, \mathcal{U}^H \diamond_\Phi \mathcal{U}_k \right\rangle \Leftrightarrow \left\langle \bar{\mathcal{U}}^H\bar{\mathcal{U}}_k, \bar{\mathcal{Q}} \right\rangle \ge \left\langle \bar{\mathcal{U}}^H\bar{\mathcal{U}}_k, \bar{\mathcal{U}}^H\bar{\mathcal{U}}_k \right\rangle$$

for some unitary tensor $\mathcal{Q} \in \mathbb{R}^{k\times k\times q}$. Let $\mathcal{U}^H \diamond_\Phi \mathcal{U}_k = \mathcal{Q}_1 \diamond_\Phi \Gamma \diamond_\Phi \mathcal{Q}_2^H$ be the transformed t-SVD of $\mathcal{U}^H \diamond_\Phi \mathcal{U}_k$. Then we have $\bar{\Gamma}(i,i,s) \le 1 (1 \le i \le k_s)$, and we can choose $\mathcal{Q} = \mathcal{Q}_1 \diamond_\Phi \mathcal{Q}_2^H$. □

Lemma D.9. *Let $\mathcal{Z}_k = \mathcal{U}_k \diamond_\Phi \Sigma_k \diamond_\Phi \mathcal{V}_k^H$ be a multi-rank $\mathbf{r}$ tensor such that*

$$\|\mathcal{Z}_k - \mathcal{L}\|_F \le \frac{\sigma_{\min}(\mathcal{L})}{10\sqrt{2}}.$$

*Then the tensor $\hat{\mathcal{Z}}_k$ returned by **Trim** algorithm satisfies*

$$\|\hat{\mathcal{U}}_k^H \diamond_\Phi \dot{e}_i\|_F \le \frac{10}{9}\sqrt{\frac{\mu s_r}{nq}} \text{ and } \|\hat{\mathcal{V}}_k \diamond_\Phi \dot{e}_j\|_F \le \frac{10}{9}\sqrt{\frac{\mu s_r}{nq}}$$

for $1 \le i, j \le n$, where $\hat{\mathcal{Z}} = \hat{\mathcal{U}} \diamond_\Phi \hat{\Sigma} \diamond_\Phi \hat{\mathcal{V}}^H$. Furthermore, $\|\hat{\mathcal{Z}}_k - \mathcal{L}\|_F \le 8\kappa\|\mathcal{Z}_k - \mathcal{L}\|_F$.

Proof. For simplicity, let $d = \|\mathcal{Z}_k - \mathcal{L}\|_F$. By Lemma D.7, we have

$$\|\mathcal{U}_k \diamond_\Phi \mathcal{U}_k^H - \mathcal{U} \diamond_\Phi \mathcal{U}^H\|_F \le \frac{\sqrt{2}d}{\sigma_{\min}(\mathcal{L})} \quad \text{and} \quad \|\mathcal{V}_k \diamond_\Phi \mathcal{V}_k^H - \mathcal{V} \diamond_\Phi \mathcal{V}^H\|_F \le \frac{\sqrt{2}d}{\sigma_{\min}(\mathcal{L})}.$$

This together with Lemma D.8 implies that there exist two unitary tensors $\mathcal{Q}_u \in \mathbb{R}^{r\times r\times q}$ and $\mathcal{Q}_v \in \mathbb{R}^{r\times r\times q}$ such that

$$\|\mathcal{U}_k - \mathcal{U} \diamond_\Phi \mathcal{Q}_u\|_F \le \frac{\sqrt{2}d}{\sigma_{\min}(\mathcal{L})} \quad \text{and} \quad \|\mathcal{V}_k - \mathcal{V} \diamond_\Phi \mathcal{Q}_v\|_F \le \frac{\sqrt{2}d}{\sigma_{\min}(\mathcal{L})}.$$

It follows that

$$\begin{aligned}
&\|\Sigma_k - \mathcal{Q}_u^H \diamond_\Phi \Sigma \diamond_\Phi \mathcal{Q}_v\|_F = \|\bar{\Sigma}_k - \bar{\mathcal{Q}}_u^H \bar{\Sigma}\bar{\mathcal{Q}}_v\|_F = \|\bar{\mathcal{U}}_k^H \bar{\mathcal{Z}}_k \bar{\mathcal{V}}_k - (\bar{\mathcal{U}}\bar{\mathcal{Q}}_u)^H \bar{\mathcal{L}}(\bar{\mathcal{V}}\bar{\mathcal{Q}}_v)\|_F \\
\le&\|\bar{\mathcal{U}}_k^H \bar{\mathcal{Z}}_k\bar{\mathcal{V}}_k - (\bar{\mathcal{U}}\bar{\mathcal{Q}}_u)^H \bar{\mathcal{Z}}_k\bar{\mathcal{V}}_k\|_F + \|(\bar{\mathcal{U}}\bar{\mathcal{Q}}_u)^H \bar{\mathcal{Z}}_k\bar{\mathcal{V}}_k - (\bar{\mathcal{U}}\bar{\mathcal{Q}}_u)^H \bar{\mathcal{L}}\bar{\mathcal{V}}_k\|_F + \|(\bar{\mathcal{U}}\bar{\mathcal{Q}}_u)^H \bar{\mathcal{L}}\bar{\mathcal{V}}_k - (\bar{\mathcal{U}}\bar{\mathcal{Q}}_u)^H \bar{\mathcal{L}}(\bar{\mathcal{V}}\bar{\mathcal{Q}}_v)\|_F \\
\le&\|\bar{\mathcal{U}}_k - \bar{\mathcal{U}}\bar{\mathcal{Q}}_u\|_F \cdot \|\bar{\mathcal{Z}}_k\| + \|\bar{\mathcal{Z}}_k - \bar{\mathcal{L}}\|_F + \|\bar{\mathcal{L}}\| \cdot \|\bar{\mathcal{V}}_k - \bar{\mathcal{V}}\bar{\mathcal{Q}}_v\|_F \\
\le&\frac{\sqrt{2}d(d + \sigma_{\max}(\mathcal{L}))}{\sigma_{\min}(\mathcal{L})} + d + \frac{\sqrt{2}d\sigma_{\max}(\mathcal{L})}{\sigma_{\min}(\mathcal{L})} \le 4\kappa d,
\end{aligned}$$

where κ is the condition number of $\bar{\mathcal{L}}$, and we have used the assumption $d \le \sigma_{\min}(\mathcal{L})/10\sqrt{2} \le \sigma_{\max}(\mathcal{L})/10\sqrt{2}$ and the fact

$$\|\mathcal{Z}_k\| \le \|\mathcal{L}\| + \|\mathcal{Z}_k - \mathcal{L}\| \le \sigma_{\max}(\mathcal{L}) + d$$

in the last two inequalities.
Recall that $\tilde{\mathcal{A}}_k$ and $\tilde{\mathcal{B}}_k$ in **Trim** algorithm are defined as

$$\begin{aligned}
\tilde{\mathcal{A}}_k^H(:, i, :) &= \min\left(\|\mathcal{U}_k^H(:, i, :)\|_F, \sqrt{\frac{\mu s_r}{nq}}\right) \frac{\mathcal{U}_k^H(:, i, :)}{\|\mathcal{U}_k^H(:, i, :)\|_F}, \\
\tilde{\mathcal{B}}_k^H(:, i, :) &= \min\left(\|\mathcal{V}_k^H(:, i, :)\|_F, \sqrt{\frac{\mu s_r}{nq}}\right) \frac{\mathcal{V}_k^H(:, i, :)}{\|\mathcal{V}_k^H(:, i, :)\|_F}.
\end{aligned}$$

Because

$$\|\dot{e}_i^H \diamond_\Phi \mathcal{U} \diamond_\Phi \mathcal{Q}_u\|_F \le \sqrt{\frac{\mu s_r}{nq}} \quad \text{and} \quad \|\dot{e}_i^H \diamond_\Phi \mathcal{V} \diamond_\Phi \mathcal{Q}_v\|_F \le \sqrt{\frac{\mu s_r}{nq}},$$

we have

$$\begin{aligned}
\|\dot{e}_i \diamond_\Phi \left(\tilde{\mathcal{A}}_k - \mathcal{U} \diamond_\Phi \mathcal{Q}_u\right)\|_F &\le \|\dot{e}_i \diamond_\Phi \left(\mathcal{U}_k - \mathcal{U} \diamond_\Phi \mathcal{Q}_u\right)\|_F, \\
\|\dot{e}_i \diamond_\Phi \left(\tilde{\mathcal{B}}_k - \mathcal{V} \diamond_\Phi \mathcal{Q}_v\right)\|_F &\le \|\dot{e}_i \diamond_\Phi \left(\mathcal{V}_k - \mathcal{V} \diamond_\Phi \mathcal{Q}_v\right)\|_F.
\end{aligned}$$

Therefore,

$$\begin{aligned}
\|\tilde{\mathcal{A}}_k - \mathcal{U} \diamond_\Phi \mathcal{Q}_u\|_F &\le \|\mathcal{U}_k - \mathcal{U} \diamond_\Phi \mathcal{Q}_u\|_F \le \frac{\sqrt{2}d}{\sigma_{\min}(\mathcal{L})}, \\
\|\tilde{\mathcal{B}}_k - \mathcal{V} \diamond_\Phi \mathcal{Q}_v\|_F &\le \|\mathcal{V}_k - \mathcal{V} \diamond_\Phi \mathcal{Q}_v\|_F \le \frac{\sqrt{2}d}{\sigma_{\min}(\mathcal{L})}.
\end{aligned}$$

Since $\hat{\mathcal{Z}}_k = \tilde{\mathcal{A}}_k \diamond_\Phi \Sigma_k \diamond_\Phi \tilde{\mathcal{B}}_k^H$ by **Trim** algorithm, we have

$$\begin{aligned}
&\|\hat{\mathcal{Z}}_k - \mathcal{L}\|_F = \|\tilde{\mathcal{A}}_k \diamond_\Phi \Sigma_k \diamond_\Phi \tilde{\mathcal{B}}_k^H - (\mathcal{U} \diamond_\Phi \mathcal{Q}_u) \diamond_\Phi (\mathcal{Q}_u^H \diamond_\Phi \Sigma \diamond_\Phi \mathcal{Q}_v) \diamond_\Phi (\mathcal{V} \diamond_\Phi \mathcal{Q}_v)^H\|_F \\
=&\|\bar{\tilde{\mathcal{A}}}_k \bar{\Sigma}_k \bar{\tilde{\mathcal{B}}}_k^H - (\bar{\mathcal{U}}\bar{\mathcal{Q}}_u)(\bar{\mathcal{Q}}_u^H \bar{\Sigma} \bar{\mathcal{Q}}_v)(\bar{\mathcal{V}}\bar{\mathcal{Q}}_v)^H\|_F \\
\leq&\|\bar{\tilde{\mathcal{A}}}_k \bar{\Sigma}_k \bar{\tilde{\mathcal{B}}}_k^H - (\bar{\mathcal{U}}\bar{\mathcal{Q}}_u)\bar{\Sigma}_k \bar{\tilde{\mathcal{B}}}_k^H\|_F + \|(\bar{\mathcal{U}}\bar{\mathcal{Q}}_u)\bar{\Sigma}_k \bar{\tilde{\mathcal{B}}}_k^H - (\bar{\mathcal{U}}\bar{\mathcal{Q}}_u)(\bar{\mathcal{Q}}_u^H \bar{\Sigma} \bar{\mathcal{Q}}_v)\bar{\tilde{\mathcal{B}}}_k^H\|_F \\
&+ \|(\bar{\mathcal{U}}\bar{\mathcal{Q}}_u)(\bar{\mathcal{Q}}_u^H \bar{\Sigma} \bar{\mathcal{Q}}_v)\bar{\tilde{\mathcal{B}}}_k^H - (\bar{\mathcal{U}}\bar{\mathcal{Q}}_u)(\bar{\mathcal{Q}}_u^H \bar{\Sigma} \bar{\mathcal{Q}}_v)(\bar{\mathcal{V}}\bar{\mathcal{Q}}_v)^H\|_F \\
\leq&\|\bar{\tilde{\mathcal{A}}}_k - \bar{\mathcal{U}}\bar{\mathcal{Q}}_u\|_F \|\bar{\Sigma}_k\| \|\bar{\tilde{\mathcal{B}}}_k\| + \|\bar{\Sigma}_k - \bar{\mathcal{Q}}_u^H \bar{\Sigma} \bar{\mathcal{Q}}_v\|_F \|\bar{\tilde{\mathcal{B}}}_k\| + \|\bar{\Sigma}\| \|\bar{\tilde{\mathcal{B}}}_k - \bar{\mathcal{V}}\bar{\mathcal{Q}}_v\|_F \\
\leq&\frac{\sqrt{2}d}{\sigma_{\min}(\mathcal{L})} (\sigma_{\max}(\mathcal{L}) + d) \left(\frac{\sqrt{2}d}{\sigma_{\min}(\mathcal{L})} + 1 \right) + 4\kappa d \left(\frac{\sqrt{2}d}{\sigma_{\min}(\mathcal{L})} + 1 \right) + \sigma_{\max}(\mathcal{L}) \frac{\sqrt{2}d}{\sigma_{\min}(\mathcal{L})} \leq 8\kappa d,
\end{aligned}$$

where we use the fact

$$\|\Sigma_k\| = \|\mathcal{Z}_k\| = \|\mathcal{L}\| + \|\mathcal{Z}_k - \mathcal{L}\| \leq \sigma_{\max}(\mathcal{L}) + d.$$

It remains to estimate the incoherence of $\hat{\mathcal{Z}}_k$. Because $\tilde{\mathcal{A}}_k$ and $\tilde{\mathcal{B}}_k$ are not necessarily orthogonal, we consider their QR factorizations:

$$\tilde{\mathcal{A}}_k = \tilde{\mathcal{U}}_k \diamond_\Phi \mathcal{R}_u \quad \text{and} \quad \tilde{\mathcal{B}}_k = \tilde{\mathcal{V}}_k \diamond_\Phi \mathcal{R}_v.$$

First note that

$$\begin{aligned}
\sigma_{\min}(\tilde{\mathcal{A}}_k) \geq& 1 - \|\tilde{\mathcal{A}}_k - \mathcal{U} \diamond_\Phi \mathcal{Q}_u\| \geq 1 - \frac{\sqrt{2}d}{\sigma_{\min}(\mathcal{L})} \geq \frac{9}{10}, \\
\sigma_{\min}(\tilde{\mathcal{B}}_k) \geq& 1 - \|\tilde{\mathcal{B}}_k - \mathcal{V} \diamond_\Phi \mathcal{Q}_v\| \geq 1 - \frac{\sqrt{2}d}{\sigma_{\min}(\mathcal{L})} \geq \frac{9}{10}.
\end{aligned}$$

We give the proof of the first inequality in the above inequalities. We assume $\mathcal{X}_0 \in \mathbb{R}^{r \times r \times q}$, $\|\mathcal{X}_0\|_F = 1$ and $\tilde{\mathcal{A}}_k^H \diamond_\Phi \tilde{\mathcal{A}}_k \diamond_\Phi \mathcal{X}_0 = \sigma_{\min}(\tilde{\mathcal{A}}_k)\mathcal{X}_0$

$$\begin{aligned}
&1 = \|\mathcal{U} \diamond_\Phi \mathcal{Q}_u \diamond_\Phi \mathcal{X}_0\|_F \leq \|(\mathcal{U} \diamond_\Phi \mathcal{Q}_u - \tilde{\mathcal{A}}_k + \tilde{\mathcal{A}}_k) \diamond_\Phi \mathcal{X}\|_F \\
&\leq \|(\mathcal{U} \diamond_\Phi \mathcal{Q}_u - \tilde{\mathcal{A}}_k) \diamond_\Phi \mathcal{X}\|_F + \|\tilde{\mathcal{A}}_k \diamond_\Phi \mathcal{X}\|_F \leq \|\mathcal{U} \diamond_\Phi \mathcal{Q}_u - \tilde{\mathcal{A}}_k\| + \sigma_{\min}(\tilde{\mathcal{A}}_k).
\end{aligned}$$

Therefore,

$$\|\mathcal{R}_u^{-1}\| \leq \frac{10}{9} \quad \text{and} \quad \|\mathcal{R}_v^{-1}\| \leq \frac{10}{9}.$$

Consequently,

$$\begin{aligned}
\|\dot{e}_i^H \diamond_\Phi \hat{\mathcal{U}}_k\|_F = \|\dot{e}_i^H \diamond_\Phi \tilde{\mathcal{U}}_k\|_F = \|\dot{e}_i^H \diamond_\Phi \tilde{\mathcal{A}}_k \diamond_\Phi \mathcal{R}_u^{-1}\|_F \leq \|\dot{e}_i^H \diamond_\Phi \tilde{\mathcal{A}}_k\|_F \|\mathcal{R}_u^{-1}\| \leq \frac{10}{9}\sqrt{\frac{\mu s_r}{nq}}, \\
\|\dot{e}_i^H \diamond_\Phi \hat{\mathcal{V}}_k\|_F = \|\dot{e}_i^H \diamond_\Phi \tilde{\mathcal{V}}_k\|_F = \|\dot{e}_i^H \diamond_\Phi \tilde{\mathcal{B}}_k \diamond_\Phi \mathcal{R}_v^{-1}\|_F \leq \|\dot{e}_i^H \diamond_\Phi \tilde{\mathcal{B}}_k\|_F \|\mathcal{R}_v^{-1}\| \leq \frac{10}{9}\sqrt{\frac{\mu s_r}{nq}}.
\end{aligned}$$

□

Lemma D.10 (Trim Property)**.** *Let **Trim** be the algorithm defined in Algorithm 3. If $\mathcal{L}_k \in \mathbb{R}^{n \times n \times q}$ is a multi-rank $\boldsymbol{r}$ tensor with*

$$\|\mathcal{L}_k - \mathcal{L}\| \leq \frac{\bar{\sigma}_{s_r}}{20\sqrt{s_r}}, \tag{12}$$

then the trim output with level $\sqrt{\frac{\mu s_r}{nq}}$ satisfies

$$\|\tilde{\mathcal{L}}_k - \mathcal{L}\|_F \leq 8\kappa \|\mathcal{L}_k - \mathcal{L}\|_F, \tag{13}$$

$$\max_{i \in [n]} \|\mathcal{A}_k^H \diamond_\Phi \dot{e}_i\|_F \leq \frac{10}{9}\sqrt{\frac{\mu s_r}{nq}}, \quad \textit{and} \quad \max_{j \in [n]} \|\mathcal{B}_k^H \diamond_\Phi \dot{e}_i\|_F \leq \frac{10}{9}\sqrt{\frac{\mu s_r}{nq}}, \tag{14}$$

where $\tilde{\mathcal{L}}_k = \mathcal{A}_k \diamond_\Phi \Sigma_k \diamond_\Phi \mathcal{B}_k^H$ is the transformed t-SVD of $\tilde{\mathcal{L}}_k$. Furthermore, it follows that

$$\|\tilde{\mathcal{L}}_k - \mathcal{L}\| \leq 8\kappa\sqrt{2s_r} \|\mathcal{L}_k - \mathcal{L}\|. \tag{15}$$

Proof. Since both $\mathcal{L}$ and $\mathcal{L}_k$ are multi-rank $\boldsymbol{r}$ tensor, $\mathcal{L}_k - \mathcal{L}$ is multi-rank at most $2\boldsymbol{r}$. So

$$\|\mathcal{L}_k - \mathcal{L}\|_F = \|\bar{\mathcal{L}}_k - \bar{\mathcal{L}}\|_F \leq \sqrt{2s_r}\|\bar{\mathcal{L}}_k - \bar{\mathcal{L}}\| = \sqrt{2s_r}\|\mathcal{L}_k - \mathcal{L}\| \leq \sqrt{2s_r}\frac{\bar{\sigma}_{s_r}}{20\sqrt{s_r}} = \frac{\bar{\sigma}_{s_r}}{10\sqrt{2}}.$$

The **Trim** incoherence follows directly from Lemma D.9. Furthermore, we have

$$\|\tilde{\mathcal{L}}_k - \mathcal{L}\|_F \leq 8\kappa\|\mathcal{L}_k - \mathcal{L}\|_F$$

and

$$\|\tilde{\mathcal{L}}_k - \mathcal{L}\| = \|\bar{\tilde{\mathcal{L}}}_k - \bar{\mathcal{L}}\| \leq 8\sqrt{2s_r}\kappa\|\bar{\mathcal{L}}_k - \bar{\mathcal{L}}\| = 8\kappa\sqrt{2s_r}\|\mathcal{L}_k - \mathcal{L}\|.$$

□

Lemma D.11. *Let $\mathcal{L} = \mathcal{U} \diamond_\Phi \Sigma \diamond_\Phi \mathcal{V}^H$ and $\tilde{\mathcal{L}}_k = \mathcal{A}_k \diamond_\Phi \Sigma_k \diamond_\Phi \mathcal{B}_k^H$ be the transformed t-SVD of two multi-rank $\boldsymbol{r}$ tensors, then*

$$\|\mathcal{U} \diamond_\Phi \mathcal{U}^H - \mathcal{A}_k \diamond_\Phi \mathcal{A}_k^H\| \leq \frac{\|\tilde{\mathcal{L}}_k - \mathcal{L}\|}{\bar{\sigma}_{s_r}}, \|\mathcal{V} \diamond_\Phi \mathcal{V}^H - \mathcal{B}_k \diamond_\Phi \mathcal{B}_k^H\| \leq \frac{\|\tilde{\mathcal{L}}_k - \mathcal{L}\|}{\bar{\sigma}_{s_r}}, \tag{16}$$

and

$$\|(I - P_{T_k})\mathcal{L}\| \leq \frac{\|\tilde{\mathcal{L}}_k - \mathcal{L}\|^2}{\bar{\sigma}_{s_r}}. \tag{17}$$

Proof. From Lemma D.7 we get Eq.(16). Since $\mathcal{L} = \mathcal{U} \diamond_\Phi \mathcal{U}^H \diamond_\Phi \mathcal{L}$ and $\tilde{\mathcal{L}}_k \diamond_\Phi (\mathcal{I} - \mathcal{B}_k \diamond_\Phi \mathcal{B}^H) = 0$, we have

$$\begin{aligned}
&\|(I - P_{T_k})\mathcal{L}\| = \|(\mathcal{I} - \mathcal{A}_k \diamond_\Phi \mathcal{A}_k^H) \diamond_\Phi \mathcal{L} \diamond_\Phi (\mathcal{I} - \mathcal{B}_k \diamond_\Phi \mathcal{B}_k^H)\| \\
=&\|(\mathcal{I} - \mathcal{A}_k \diamond_\Phi \mathcal{A}_k^H) \diamond_\Phi \mathcal{U} \diamond_\Phi \mathcal{U}^H \diamond_\Phi \mathcal{L} \diamond_\Phi (\mathcal{I} - \mathcal{B}_k \diamond_\Phi \mathcal{B}_k^H)\| \\
=&\|(\mathcal{U} \diamond_\Phi \mathcal{U}^H - \mathcal{A}_k \diamond_\Phi \mathcal{A}_k^H) \diamond_\Phi \mathcal{U} \diamond_\Phi \mathcal{U}^H \diamond_\Phi \mathcal{L} \diamond_\Phi (\mathcal{I} - \mathcal{B}_k \diamond_\Phi \mathcal{B}_k^H)\| \\
=&\|(\mathcal{U} \diamond_\Phi \mathcal{U}^H - \mathcal{A}_k \diamond_\Phi \mathcal{A}_k^H) \diamond_\Phi \mathcal{L} \diamond_\Phi (\mathcal{I} - \mathcal{B}_k \diamond_\Phi \mathcal{B}_k^H)\| \\
=&\|(\mathcal{U} \diamond_\Phi \mathcal{U}^H - \mathcal{A}_k \diamond_\Phi \mathcal{A}_k^H) \diamond_\Phi (\mathcal{L} - \tilde{\mathcal{L}}_k) \diamond_\Phi (\mathcal{I} - \mathcal{B}_k \diamond_\Phi \mathcal{B}_k^H)\| \\
=&\|(\bar{\mathcal{U}}\bar{\mathcal{U}}^H - \bar{\mathcal{A}}_k\bar{\mathcal{A}}_k^H)(\bar{\mathcal{L}} - \bar{\tilde{\mathcal{L}}}_k)(\bar{\mathcal{I}} - \bar{\mathcal{B}}_k\bar{\mathcal{B}}_k^H)\| \leq \|\bar{\mathcal{U}}\bar{\mathcal{U}}^H - \bar{\mathcal{A}}_k\bar{\mathcal{A}}_k^H\| \cdot \|\bar{\mathcal{L}} - \bar{\tilde{\mathcal{L}}}_k\| \cdot \|\bar{\mathcal{I}} - \bar{\mathcal{B}}_k\bar{\mathcal{B}}_k^H\| \\
\leq&\frac{\|\bar{\mathcal{L}} - \bar{\tilde{\mathcal{L}}}_k\|^2}{\bar{\sigma}_{s_r}} = \frac{\|\tilde{\mathcal{L}}_k - \mathcal{L}\|^2}{\bar{\sigma}_{s_r}}.
\end{aligned}$$

□

Lemma D.12. *Let $\mathcal{S} \in \mathbb{R}^{n\times n\times q}$ be a symmetric tensor satisfying Assumption 5.2. Let $\tilde{\mathcal{L}}_k \in \mathbb{R}^{n\times n\times q}$ be a multi-rank $\boldsymbol{r}$ tensor with $\frac{100}{81}$-incoherence. That is,*

$$\max_{i\in[n]}\|\mathcal{A}_k^H \diamond_\Phi \dot{e}_i\|_F \leq \frac{10}{9}\sqrt{\frac{\mu s_r}{nq}} \quad \textit{and} \quad \max_{j\in[n]}\|\mathcal{B}_k^H \diamond_\Phi \dot{e}_j\|_F \leq \frac{10}{9}\sqrt{\frac{\mu s_r}{nq}},$$

where $\tilde{\mathcal{L}}_k = \mathcal{A}_k \diamond_\Phi \Sigma_k \diamond_\Phi \mathcal{B}_k^H$ is the transformed t-SVD of $\tilde{\mathcal{L}}_k$. If supp$(\mathcal{S}_k) \subset \Omega$, then

$$\|P_{T_k}(\mathcal{S} - \mathcal{S}_k)\|_\infty \leq 4\alpha\mu s_r\|\mathcal{S} - \mathcal{S}_k\|_\infty.$$

Proof. By the incoherence assumption of $\tilde{\mathcal{L}}_k$ and the sparsity assumption of $\mathcal{S}-\mathcal{S}_k$, we have

$$
\begin{aligned}
&[P_{T_k}(\mathcal{S}-\mathcal{S}_k)]_{ijt} = \langle P_{T_k}(\mathcal{S}-\mathcal{S}_k), \mathcal{E}_{ijt}\rangle \\
=&\left\langle \mathcal{A}_k \diamond_\Phi \mathcal{A}_k^H \diamond_\Phi (\mathcal{S}-\mathcal{S}_k) + (\mathcal{S}-\mathcal{S}_k)\diamond_\Phi \mathcal{B}_k \diamond_\Phi \mathcal{B}_k^H - \mathcal{A}_k \diamond_\Phi \mathcal{A}_k^H \diamond_\Phi (\mathcal{S}-\mathcal{S}_k)\diamond_\Phi \mathcal{B}_k \diamond_\Phi \mathcal{B}_k^H, \mathcal{E}_{ijt}\right\rangle \\
=&\left\langle \bar{\mathcal{A}}_k\bar{\mathcal{A}}_k^H(\bar{\mathcal{S}}-\bar{\mathcal{S}}_k) + (\bar{\mathcal{S}}-\bar{\mathcal{S}}_k)\bar{\mathcal{B}}_k\bar{\mathcal{B}}_k^H - \bar{\mathcal{A}}_k\bar{\mathcal{A}}_k^H(\bar{\mathcal{S}}-\bar{\mathcal{S}}_k)\bar{\mathcal{B}}_k\bar{\mathcal{B}}_k^H, \bar{\mathcal{E}}_{ijt}\right\rangle \\
=&\left\langle \bar{\mathcal{S}}-\bar{\mathcal{S}}_k, \bar{\mathcal{A}}_k\bar{\mathcal{A}}_k^H\bar{\mathcal{E}}_{ijt} + \bar{\mathcal{E}}_{ijt}\bar{\mathcal{B}}_k\bar{\mathcal{B}}_k^H - \bar{\mathcal{A}}_k\bar{\mathcal{A}}_k^H\bar{\mathcal{E}}_{ijt}\bar{\mathcal{B}}_k\bar{\mathcal{B}}_k^H\right\rangle \\
=&\left\langle \mathcal{S}-\mathcal{S}_k, \mathcal{A}_k \diamond_\Phi \mathcal{A}_k^H \diamond_\Phi \mathcal{E}_{ijt} + \mathcal{E}_{ijt}\diamond_\Phi \mathcal{B}_k \diamond_\Phi \mathcal{B}_k^H - \mathcal{A}_k \diamond_\Phi \mathcal{A}_k^H \diamond_\Phi \mathcal{E}_{ijt}\diamond_\Phi \mathcal{B}_k \diamond_\Phi \mathcal{B}_k^H\right\rangle \\
=&\left\langle (\mathcal{S}-\mathcal{S}_k)\diamond_\Phi \dot{e}_j, \mathcal{A}_k \diamond_\Phi \mathcal{A}_k^H \diamond_\Phi \dot{e}_i \diamond_\Phi e_t\right\rangle + \left\langle \dot{e}_i^H \diamond_\Phi (\mathcal{S}-\mathcal{S}_k), e_t \diamond_\Phi \dot{e}_j^H \diamond_\Phi \mathcal{B}_k \diamond_\Phi \mathcal{B}_k^H\right\rangle \\
&-\left\langle (\mathcal{S}-\mathcal{S}_k), \mathcal{A}_k \diamond_\Phi \mathcal{A}_k^H \diamond_\Phi \mathcal{E}_{ijt}\diamond_\Phi \mathcal{B}_k \diamond_\Phi \mathcal{B}_k^H\right\rangle \\
\leq&\sum_{c\in[q]}\left\langle (\mathcal{S}-\mathcal{S}_k)(:,j,c), (\mathcal{A}_k \diamond_\Phi \mathcal{A}_k^H \diamond_\Phi \dot{e}_i \diamond_\Phi e_t)(:,:,c)\right\rangle + \sum_{c\in[q]}\left\langle (\mathcal{S}-\mathcal{S}_k)(i,:,c), (e_t \diamond_\Phi \dot{e}_j^H \diamond_\Phi \mathcal{B}_k \diamond_\Phi \mathcal{B}_k^H)(:,:,c)\right\rangle \\
&+\|\bar{\mathcal{S}}-\bar{\mathcal{S}}_k\|\|\bar{\mathcal{A}}_k\bar{\mathcal{A}}_k^H\bar{\mathcal{E}}_{ijt}\bar{\mathcal{B}}_k\bar{\mathcal{B}}_k^H\|_* \\
\leq&\sum_{c\in[q]}\|(\mathcal{S}-\mathcal{S}_k)(:,j,c)\|_\infty\cdot\|(\mathcal{A}_k \diamond_\Phi \mathcal{A}_k^H \diamond_\Phi \dot{e}_i \diamond_\Phi e_t)(:,:,c)\|_1 \\
&+\sum_{c\in[q]}\|(\mathcal{S}-\mathcal{S}_k)(i,:,c)\|_\infty\cdot\|(e_t \diamond_\Phi \dot{e}_j^H \diamond_\Phi \mathcal{B}_k \diamond_\Phi \mathcal{B}_k^H)(:,:,c)\|_1 \\
&+\alpha nq\|\mathcal{S}-\mathcal{S}_k\|_\infty\|\mathcal{A}_k \diamond_\Phi \mathcal{A}_k^H \diamond_\Phi \mathcal{E}_{ijt}\diamond_\Phi \mathcal{B}_k \diamond_\Phi \mathcal{B}_k^H\|_F \\
\leq&\|\mathcal{S}-\mathcal{S}_k\|_\infty\left(\|\mathcal{A}_k \diamond_\Phi \mathcal{A}_k^H \diamond_\Phi \dot{e}_i \diamond_\Phi e_t\|_1 + \|e_t \diamond_\Phi \dot{e}_j^H \diamond_\Phi \mathcal{B}_k \diamond_\Phi \mathcal{B}_k^H\|_1\right) + \alpha nq\frac{100\mu s_r}{81nq}\|\mathcal{S}-\mathcal{S}_k\|_\infty \\
=&\|\mathcal{S}-\mathcal{S}_k\|_\infty\left(\sum_{a|(a,i,t)\in\Omega}\|\dot{e}_a^H \diamond_\Phi \mathcal{A}_k \diamond_\Phi \mathcal{A}_k^H \diamond_\Phi \dot{e}_i \diamond_\Phi e_t\|_1 + \sum_{b|(j,b,t)\in\Omega}\|e_t \diamond_\Phi \dot{e}_j^H \diamond_\Phi \mathcal{B}_k \diamond_\Phi \mathcal{B}_k^H \diamond_\Phi \dot{e}_b\|_1\right) \\
&+\frac{100\alpha\mu s_r}{81}\|\mathcal{S}-\mathcal{S}_k\|_\infty \\
=&\|\mathcal{S}-\mathcal{S}_k\|_\infty\left(\sum_{a|(a,i,t)\in\Omega}\|\dot{e}_a^H \diamond_\Phi \mathcal{A}_k \diamond_\Phi \mathcal{A}_k^H \diamond_\Phi \dot{e}_i \diamond_\Phi e_t\|_F + \sum_{b|(j,b,t)\in\Omega}\|\dot{e}_j^H \diamond_\Phi \mathcal{B}_k \diamond_\Phi \mathcal{B}_k^H \diamond_\Phi \dot{e}_b \diamond_\Phi e_t\|_F\right) \\
&+\frac{100\alpha\mu s_r}{81}\|\mathcal{S}-\mathcal{S}_k\|_\infty \qquad \text{one element in the } l_1 \text{ norm} \\
\leq&2\alpha n\frac{100\mu s_r}{81nq}\|\mathcal{S}-\mathcal{S}_k\|_\infty + \frac{100\alpha\mu s_r}{81}\|\mathcal{S}-\mathcal{S}_k\|_\infty = \left(\frac{200\alpha\mu s_r}{81q} + \frac{100\alpha\mu s_r}{81}\right)\|\mathcal{S}-\mathcal{S}_k\|_\infty \leq 4\alpha\mu s_r\|\mathcal{S}-\mathcal{S}_k\|_\infty.
\end{aligned}
$$

This proof uses the fact that $\|X\|_1 \leq \sqrt{r}\|X\|_F$, $\|X\|_* \leq \sqrt{r}\|X\|_F$, $q \geq 1$ where X is a rank-r matrix. We also using the following inequality in the proof:

$$
\begin{aligned}
&\|\mathcal{A}_k \diamond_\Phi \mathcal{A}_k^H \diamond_\Phi \mathcal{E}_{ijt}\diamond_\Phi \mathcal{B}_k \diamond_\Phi \mathcal{B}_k^H\|_F = \|\bar{\mathcal{A}}_k\bar{\mathcal{A}}_k^H\bar{\mathcal{E}}_{ijt}\bar{\mathcal{B}}_k\bar{\mathcal{B}}_k^H\|_F = \sqrt{\mathrm{Tr}(\bar{\mathcal{A}}_k\bar{\mathcal{A}}_k^H\bar{\mathcal{E}}_{ijt}\bar{\mathcal{B}}_k\bar{\mathcal{B}}_k^H\bar{\mathcal{B}}_k\bar{\mathcal{B}}_k^H\bar{\mathcal{E}}_{ijt}^H\bar{\mathcal{A}}_k\bar{\mathcal{A}}_k^H)} \\
=&\|\bar{\mathcal{A}}_k\bar{\mathcal{A}}_k^H\bar{\mathcal{E}}_{ijt}\bar{\mathcal{B}}_k\|_F = \|\bar{\mathcal{B}}_k^H\bar{\mathcal{E}}_{ijt}^H\bar{\mathcal{A}}_k\bar{\mathcal{A}}_k^H\|_F = \sqrt{\mathrm{Tr}(\bar{\mathcal{B}}_k^H\bar{\mathcal{E}}_{ijt}^H\bar{\mathcal{A}}_k\bar{\mathcal{A}}_k^H\bar{\mathcal{A}}_k\bar{\mathcal{A}}_k^H\bar{\mathcal{E}}_{ijt}\bar{\mathcal{B}}_k)} = \|\bar{\mathcal{B}}_k^H\bar{\mathcal{E}}_{ijt}^H\bar{\mathcal{A}}_k\|_F \\
\leq&\|\bar{\mathcal{B}}_k^H\bar{e}_j\|_F\|\bar{e}_k^H\|_F\|\bar{\bar{e}}_i^H\bar{\mathcal{A}}_k\|_F = \|\mathcal{B}_k^H \diamond_\Phi \dot{e}_j\|_F\|e_t^H\|_F\|\dot{e}_i^H \diamond_\Phi \mathcal{A}_k\|_F \leq \frac{100\mu s_r}{81nq}.
\end{aligned}
$$

□

Lemma D.13. *Under the symmetric setting, i.e.,* $\mathcal{U}\diamond_\Phi\mathcal{U}^H = \mathcal{V}\diamond_\Phi\mathcal{V}^H$ *where* $\mathcal{U}\in\mathbb{R}^{n\times r\times q}$ *are two orthogonal tensors, we have*

$$\|P_T(\mathcal{Z})\| \leq \sqrt{\frac{4}{3}}\|\mathcal{Z}\|$$

for any symmetric tensor $\mathcal{Z}\in\mathbb{R}^{n\times n\times q}$*. Moreover, the bound is tight.*

Proof. First note that

$$P_T(\mathcal{Z}) = \mathcal{U} \diamond_\Phi \mathcal{U}^H \diamond_\Phi \mathcal{Z} + \mathcal{Z} \diamond_\Phi \mathcal{U} \diamond_\Phi \mathcal{U}^H - \mathcal{U} \diamond_\Phi \mathcal{U}^H \diamond_\Phi \mathcal{Z} \diamond_\Phi \mathcal{U} \diamond_\Phi \mathcal{U}^H.$$

Then, we have

$$\begin{aligned}&\|P_T(\mathcal{Z})\| = \|\mathcal{U} \diamond_\Phi \mathcal{U}^H \diamond_\Phi \mathcal{Z} + \mathcal{Z} \diamond_\Phi \mathcal{U} \diamond_\Phi \mathcal{U}^H - \mathcal{U} \diamond_\Phi \mathcal{U}^H \diamond_\Phi \mathcal{Z} \diamond_\Phi \mathcal{U} \diamond_\Phi \mathcal{U}^H\| \\ =&\|\bar{\mathcal{U}}\bar{\mathcal{U}}^H\bar{\mathcal{Z}} + \bar{\mathcal{Z}}\bar{\mathcal{U}}\bar{\mathcal{U}}^H - \bar{\mathcal{U}}\bar{\mathcal{U}}^H\bar{\mathcal{Z}}\bar{\mathcal{U}}\bar{\mathcal{U}}^H\| \le \sqrt{\frac{4}{3}}\|\bar{\mathcal{Z}}\| = \sqrt{\frac{4}{3}}\|\mathcal{Z}\|,\end{aligned}$$

where the last inequality follows from Lemma 8 in Cai et al. 2019. □

Lemma D.14. *Let $\mathcal{U} \in \mathbb{R}^{n\times r\times q}$ be an orthogonal tensor with μ-incoherence, i.e., $\|\mathcal{U}^H \diamond_\Phi \dot{e}_i\|_F \le \sqrt{\frac{\mu s_r}{nq}}$ for all $i \in [n]$. Then, for any $\mathcal{Z} \in \mathbb{R}^{n\times n\times q}$, the inequality*

$$\|\mathcal{U}^H \diamond_\Phi \mathcal{Z}^a \diamond_\Phi \dot{e}_i\|_F \le \max_l \sqrt{\frac{\mu s_r}{nq}}(\sqrt{n}\|\mathcal{Z} \diamond_\Phi \dot{e}_l\|_F)^a$$

holds for all i and $a \ge 0$.

Proof. This proof is done by mathematical induction.
Base case: When $a = 0$, $\|\dot{e}_i^H \diamond_\Phi \mathcal{U}\|_F \le \sqrt{\frac{\mu s_r}{nq}}$ is satisfied following from the assumption.
Induction Hypothesis: $\|\dot{e}_i^H \diamond_\Phi \mathcal{Z}^a \diamond_\Phi \mathcal{U}\|_F \le \max_l \sqrt{\frac{\mu s_r}{nq}}(\sqrt{n}\|\dot{e}_l^H \diamond_\Phi \mathcal{Z}\|_F)^a$ for all i at the a^{th} power.
Induction Step: We have

$$\begin{aligned}
&\|\dot{e}_i^H \diamond_\Phi \mathcal{Z}^{a+1} \diamond_\Phi \mathcal{U}\|_F^2 = \|\dot{e}_i^H \diamond_\Phi \mathcal{Z} \diamond_\Phi \mathcal{Z}^a \diamond_\Phi \mathcal{U}\|_F^2 = \|\bar{\dot{e}}_i^H \bar{\mathcal{Z}}\bar{\mathcal{Z}}^a\bar{\mathcal{U}}\|_F^2 \\
=&\sum_{c=0}^{q-1}\sum_{j=1}^{r}\left(\sum_{k=1}^{n}[\bar{\mathcal{Z}}]_{cn+i,cn+k}[\bar{\mathcal{Z}}^a\bar{\mathcal{U}}]_{cn+k,cr+j}\right)^2 \\
=&\sum_{c=0}^{q-1}\sum_{k_1,k_2}[\bar{\mathcal{Z}}]_{cn+i,cn+k_1}[\bar{\mathcal{Z}}]_{cn+i,cn+k_2}\sum_{j=1}^{r}[\bar{\mathcal{Z}}^a\bar{\mathcal{U}}]_{cn+k_1,cr+j}[\bar{\mathcal{Z}}^a\bar{\mathcal{U}}]_{cn+k_2,cr+j} \\
=&\sum_{c=0}^{q-1}\sum_{k_1,k_2}[\bar{\mathcal{Z}}]_{cn+i,cn+k_1}[\bar{\mathcal{Z}}]_{cn+i,cn+k_2}\left\langle e_{cn+k_1}^H\bar{\mathcal{Z}}^a\bar{\mathcal{U}}(:,cr+1:cr+r), e_{cn+k_2}^H\bar{\mathcal{Z}}^a\bar{\mathcal{U}}(:,cr+1:cr+r)\right\rangle \\
\le&\sum_{c=0}^{q-1}\sum_{k_1,k_2}\left|[\bar{\mathcal{Z}}]_{cn+i,cn+k_1}[\bar{\mathcal{Z}}]_{cn+i,cn+k_2}\right|\|e_{cn+k_1}^H\bar{\mathcal{Z}}^a\bar{\mathcal{U}}(:,cr+1:cr+r)\|_2\|e_{cn+k_2}^H\bar{\mathcal{Z}}^a\bar{\mathcal{U}}(:,cr+1:cr+r)\|_2 \\
&\qquad \|\cdot\|_2 \text{ is the } l_2 \text{ norm of vectors.} \\
\le&\sum_{k_1,k_2}\sum_{c=0}^{q-1}\left|[\bar{\mathcal{Z}}]_{cn+i,cn+k_1}[\bar{\mathcal{Z}}]_{cn+i,cn+k_2}\right|\|e_{cn+k_1}^H\bar{\mathcal{Z}}^a\bar{\mathcal{U}}\|_2\|e_{cn+k_2}^H\bar{\mathcal{Z}}^a\bar{\mathcal{U}}\|_2 \\
\le&\max_l\frac{\mu s_r}{nq}(\sqrt{n}\|\dot{e}_l^H \diamond_\Phi \mathcal{Z}\|_F)^{2a}\sum_{c=0}^{q-1}\sum_{k_1,k_2}\left|[\bar{\mathcal{Z}}]_{cn+i,cn+k_1}[\bar{\mathcal{Z}}]_{cn+i,cn+k_2}\right| \\
=&\max_l\frac{\mu s_r}{nq}(\sqrt{n}\|\dot{e}_l^H \diamond_\Phi \mathcal{Z}\|_F)^{2a}\sum_{c=0}^{q-1}\left(\|e_{cn+i}^H\bar{\mathcal{Z}}\left(:,cn+1:cn+n\right)\|_1\right)^2 \\
\le&\max_l\frac{\mu s_r}{nq}(\sqrt{n}\|\dot{e}_l^H \diamond_\Phi \mathcal{Z}\|_F)^{2a}\sum_{c=0}^{q-1}\left(\sqrt{n}\|e_{cn+i}^H\bar{\mathcal{Z}}\left(:,cn+1:cn+n\right)\|_2\right)^2 \\
=&\max_l\frac{\mu s_r}{nq}(\sqrt{n}\|\dot{e}_l^H \diamond_\Phi \mathcal{Z}\|_F)^{2a}\left(\sqrt{n}\|\bar{\dot{e}}_i^H\bar{\mathcal{Z}}\|_F\right)^2 \\
=&\max_l\frac{\mu s_r}{nq}(\sqrt{n}\|\dot{e}_l^H \diamond_\Phi \mathcal{Z}\|_F)^{2a}\left(\sqrt{n}\|\dot{e}_i^H \diamond_\Phi \mathcal{Z}\|_F\right)^2 \le \max_l\frac{\mu s_r}{nq}(\sqrt{n}\|\dot{e}_l^H \diamond_\Phi \mathcal{Z}\|_F)^{2a+2},
\end{aligned}$$

where we have used the inequality $\|x\|_1 \le \sqrt{n}\|x\|_2, x \in \mathbb{R}^n$. Now, we have

$$\|\mathcal{U}^H \diamond_\Phi \mathcal{Z}^{a+1} \diamond_\Phi \dot{e}_i\|_F \le \max_l \sqrt{\frac{\mu s_r}{nq}}(\sqrt{n}\|\mathcal{Z} \diamond_\Phi \dot{e}_l\|_F)^{a+1}.$$

In the proof, we have used the inequality

$$\|e_{cn+k_1}^H \bar{\mathcal{Z}}^a \bar{\mathcal{U}}\|_2 \|e_{cn+k_2}^H \bar{\mathcal{Z}}^a \bar{\mathcal{U}}\|_2 \le \left(\max_l \sqrt{\frac{\mu s_r}{nq}}(\sqrt{n}\|\dot{e}_l^H \diamond_\Phi \mathcal{Z}\|_F)^a\right)^2 = \max_l \frac{\mu s_r}{nq}(\sqrt{n}\|\dot{e}_l^H \diamond_\Phi \mathcal{Z}\|_F)^{2a}.$$

□

Lemma D.15. *With the condition of Theorem 5.5, let $\mathcal{L} \in \mathbb{R}^{n\times n\times q}$ and $\mathcal{S} \in \mathbb{R}^{n\times n\times q}$ be two symmetric tensors satisfying Assumptions 5.1 and 5.2. Let $\tilde{\mathcal{L}}_k \in \mathbb{R}^{n\times n\times q}$ be the trim output of $\mathcal{L}_k$. If*

$$\|\mathcal{L} - \mathcal{L}_k\| \le 8\alpha\mu s_r \gamma^k \bar{\sigma}_1, \quad \|\mathcal{S} - \mathcal{S}_k\|_\infty \le \frac{\mu s_r}{nq}\gamma^k \bar{\sigma}_1, \quad \mathit{supp}(\mathcal{S}_k) \subset \Omega.$$

then,

$$\|(P_{T_k} - I)\mathcal{L} + P_{T_k}(\mathcal{S} - \mathcal{S}_k)\| \le \tau\gamma^{k+1}\bar{\sigma}_{s_r}$$

and

$$\max_l \sqrt{n}\|\dot{e}_l^H \diamond_\Phi [(P_{T_k} - I)\mathcal{L} + P_{T_k}(\mathcal{S} - \mathcal{S}_k)]\|_F \le v\gamma^k \bar{\sigma}_{s_r}$$

hold for all $k \ge 0$, provided $1 > \gamma \ge 512\tau s_r \kappa^2 + \frac{1}{\sqrt{12}}$. Here $\tau = 4\alpha\mu s_r\kappa$ and $v = \tau(48\sqrt{\frac{\mu}{q}}s_r\kappa + \frac{\mu s_r}{\sqrt{q}})$.

Proof. Since $\mathcal{D} = \mathcal{L} + \mathcal{S}$, we have

$$\begin{aligned}
&\|(P_{T_k} - I)\mathcal{L} + P_{T_k}(\mathcal{S} - \mathcal{S}_k)\| \le \|(P_{T_k} - I)\mathcal{L}\| + \|P_{T_k}(\mathcal{S} - \mathcal{S}_k)\| \le \frac{\|\tilde{\mathcal{L}}_k - \mathcal{L}\|^2}{\bar{\sigma}_{s_r}} + \sqrt{\frac{4}{3}}\|\mathcal{S} - \mathcal{S}_k\| \\
\le& \frac{(8\sqrt{2s_r}\kappa)^2\|\mathcal{L}_k - \mathcal{L}\|^2}{\bar{\sigma}_{s_r}} + \sqrt{\frac{4}{3}}\alpha nq\|\mathcal{S} - \mathcal{S}_k\|_\infty \le 128 \cdot 8\alpha\mu s_r^2\kappa^3\|\mathcal{L} - \mathcal{L}_k\| + \sqrt{\frac{4}{3}}\alpha nq\|\mathcal{S} - \mathcal{S}_k\|_\infty \\
\le& 128 \cdot 64\alpha^2\mu^2 s_r^3\kappa^3\gamma^k\bar{\sigma}_1 + \sqrt{\frac{4}{3}}\alpha\mu s_r\gamma^k\bar{\sigma}_1 = \left(512 \cdot 4\alpha\mu s_r^2\kappa^3 + \frac{1}{4}\sqrt{\frac{4}{3}}\right)4\alpha\mu s_r\gamma^k\bar{\sigma}_1 \\
=& \left(512\tau s_r\kappa^2 + \sqrt{\frac{1}{12}}\right)4\alpha\mu s_r\gamma^k\bar{\sigma}_1 \le 4\alpha\mu s_r\gamma^{k+1}\bar{\sigma}_1 = \tau\gamma^{k+1}\bar{\sigma}_{s_r},
\end{aligned}$$

where we have used lemmas above and the fact $\frac{\|\mathcal{L} - \mathcal{L}_k\|}{\bar{\sigma}_{s_r}} \le 8\alpha\mu s_r\kappa$.

To compute the bound of $\max_l \sqrt{n}\|\dot{e}_l^H \diamond_\Phi [(P_{T_k} - I)\mathcal{L} + P_{T_k}(\mathcal{S} - \mathcal{S}_k)]\|_F$, we first note that

$$\begin{aligned}
&\max_l \|\dot{e}_l^H \diamond_\Phi (I - P_{T_k})\mathcal{L}\|_F = \max_l \|\dot{e}_l^H \diamond_\Phi (\mathcal{U} \diamond_\Phi \mathcal{U}^H - \mathcal{A} \diamond_\Phi \mathcal{A}^H) \diamond_\Phi (\mathcal{L} - \tilde{\mathcal{L}}_k) \diamond_\Phi (\mathcal{I} - \mathcal{A} \diamond_\Phi \mathcal{A}^H)\|_F \\
=& \max_l \|\bar{\dot{e}}_l^H(\bar{\mathcal{U}}\bar{\mathcal{U}}^H - \bar{\mathcal{A}}\bar{\mathcal{A}}^H)(\bar{\mathcal{L}} - \bar{\tilde{\mathcal{L}}}_k)(\bar{\mathcal{I}} - \bar{\mathcal{A}}\bar{\mathcal{A}}^H)\|_F \le \max_l \|\bar{\dot{e}}_l^H(\bar{\mathcal{U}}\bar{\mathcal{U}}^H - \bar{\mathcal{A}}\bar{\mathcal{A}}^H)\|_F\|(\bar{\mathcal{L}} - \bar{\tilde{\mathcal{L}}}_k)(\bar{\mathcal{I}} - \bar{\mathcal{A}}\bar{\mathcal{A}}^H)\| \\
\le& \max_l \|\bar{\dot{e}}_l^H(\bar{\mathcal{U}}\bar{\mathcal{U}}^H - \bar{\mathcal{A}}\bar{\mathcal{A}}^H)\|_F\|\bar{\mathcal{L}} - \bar{\tilde{\mathcal{L}}}_k\|\|\bar{\mathcal{I}} - \bar{\mathcal{A}}\bar{\mathcal{A}}^H\| = \max_l \|\dot{e}_l^H \diamond_\Phi (\mathcal{U} \diamond_\Phi \mathcal{U}^H - \mathcal{A} \diamond_\Phi \mathcal{A}^H)\|_F\|\bar{\mathcal{L}} - \bar{\tilde{\mathcal{L}}}_k\|\|\bar{\mathcal{I}} - \bar{\mathcal{A}}\bar{\mathcal{A}}^H\| \\
\le& \left(\frac{19}{9}\sqrt{\frac{\mu s_r}{nq}}\right)\|\bar{\mathcal{L}} - \bar{\tilde{\mathcal{L}}}_k\| = \frac{19}{9}\sqrt{\frac{\mu s_r}{nq}}\|\mathcal{L} - \tilde{\mathcal{L}}_k\|,
\end{aligned}$$

where we have used the fact $\mathcal{L}$ is μ-incoherence, $\tilde{\mathcal{L}}_k$ is $\frac{100}{81}\mu$-incoherence and $\|X\| \leq \|X\|_F$. Hence, for all $k \geq 0$, we have

$$\begin{aligned}
&\max_l \sqrt{n}\|\dot{e}_l^H \diamond_\Phi [(P_{T_k} - I)\mathcal{L} + P_{T_k}(\mathcal{S} - \mathcal{S}_k)]\|_F \leq \max_l \sqrt{n}\|\dot{e}_l^H \diamond_\Phi (P_{T_k} - I)\mathcal{L}\|_F + \sqrt{n}\|\dot{e}_l^H \diamond_\Phi P_{T_k}(\mathcal{S} - \mathcal{S}_k)\|_F \\
\leq& \max_l \sqrt{n}\|\dot{e}_l^H \diamond_\Phi (P_{T_k} - I)\mathcal{L}\|_F + \sqrt{n}\|\bar{e}_l^H P_{T_k}(\bar{\mathcal{S}} - \bar{\mathcal{S}}_k)\|_F \\
=& \max_l \sqrt{n}\|\dot{e}_l^H \diamond_\Phi (P_{T_k} - I)\mathcal{L}\|_F + \sqrt{n}\|\dot{e}_l^H \diamond_\Phi P_{T_k}(\mathcal{S} - \mathcal{S}_k)\|_F \\
\leq& \frac{19\sqrt{n}}{9}\sqrt{\frac{\mu s_r}{nq}}\|\mathcal{L} - \tilde{\mathcal{L}}_k\| + n\sqrt{q}\|P_{T_k}(\mathcal{S} - \mathcal{S}_k)\|_\infty \leq \frac{19}{9}8\kappa s_r\sqrt{\frac{2\mu}{q}}\|\mathcal{L} - \mathcal{L}_k\| + 4\alpha\mu s_r n\sqrt{q}\|\mathcal{S} - \mathcal{S}_k\|_\infty \\
\leq& \frac{19}{9}8\kappa s_r\sqrt{\frac{2\mu}{q}} \cdot 8\sqrt{\frac{\alpha}{q}}\mu s_r\gamma^k\bar{\sigma}_1 + 4\alpha\mu s_r n\sqrt{q} \cdot \frac{\mu s_r}{nq}\gamma^k\bar{\sigma}_1 \leq v\gamma^k\bar{\sigma}_{s_r},
\end{aligned}$$

where we have used Lemma D.10 and Lemma D.12. □

Lemma D.16. *With the condition of Theorem 5.5, let $\mathcal{L} \in \mathbb{R}^{n\times n\times q}$ and $\mathcal{S} \in \mathbb{R}^{n\times n\times q}$ be two symmetric tensors satisfying Assumptions 5.1 and 5.2. Let $\tilde{\mathcal{L}}_k \in \mathbb{R}^{n\times n\times q}$ be the trim output of $\mathcal{L}_k$. If*

$$\|\mathcal{L} - \mathcal{L}_k\| \leq 8\alpha\mu s_r\gamma^k\bar{\sigma}_1, \quad \|\mathcal{S} - \mathcal{S}_k\|_\infty \leq \frac{\mu s_r}{nq}\gamma^k\bar{\sigma}_1, \text{ and } supp(\mathcal{S}_k) \subset \Omega,$$

then $|\bar{\sigma}_{i,c} - |\bar{\lambda}_{i,c}^{(k)}|| \leq \tau\bar{\sigma}_{rq}$ and $(1-2\tau)\gamma^j\bar{\sigma}_{t,c'} \leq |\bar{\lambda}_{i,c}^{(k)}| + \gamma^j|\bar{\lambda}_{t,c'}^{(k)}| \leq (1+2\tau)\gamma^j\bar{\sigma}_{t,c'}$ hold for all $k \geq 0$, $i > r_c$, and $j \leq k+1$, provided $1 > \gamma \geq 512\tau s_r\kappa^2 + \frac{1}{\sqrt{12}}$. Here $|\bar{\lambda}_{i,c}^{(k)}|$ is the $(i,c)^{th}$ singular value of $P_{T_k}(\bar{\mathcal{D}} - \bar{\mathcal{S}}_k)$.

Proof. Since $\mathcal{D} = \mathcal{L} + \mathcal{S}$, we have $P_{T_k}(\mathcal{D} - \mathcal{S}_k) = P_{T_k}(\mathcal{L} + \mathcal{S} - \mathcal{S}_k) = \mathcal{L} + (P_{T_k} - I)\mathcal{L} + P_{T_k}(\mathcal{S} - \mathcal{S}_k)$. Hence, by Tensor Weyl's inequality and Lemma D.15, we can see that

$$|\bar{\sigma}_{i,c} - |\bar{\lambda}_{i,c}^{(k)}|| \leq \|(P_{T_k} - I)\mathcal{L} + P_{T_k}(\mathcal{S} - \mathcal{S}_k)\| \leq \tau\gamma^{k+1}\bar{\sigma}_{s_r} \leq \tau\bar{\sigma}_{s_r}$$

holds for all i and $k \geq 0$.
Notice that $\mathcal{L}$ is a multi-rank $\boldsymbol{r}$ tensors, which implies $\bar{\sigma}_{i,c} = 0, i > r_c$, so we have

$$\begin{aligned}
&||\bar{\lambda}_{i,c}^{(k)}| + \gamma^j|\bar{\lambda}_{t,c'}^{(k)}| - \gamma^j\bar{\sigma}_{t,c'}| = \left||\bar{\lambda}_{i,c}^{(k)}| - \bar{\sigma}_{i,c} + \gamma^j|\bar{\lambda}_{t,c'}^{(k)}| - \gamma^j\bar{\sigma}_{t,c'}\right| \\
\leq& \tau\gamma^{k+1}\bar{\sigma}_{s_r} + \tau\gamma^{j+k+1}\bar{\sigma}_{s_r} \leq (1+\gamma^{k+1})\tau\gamma^j\bar{\sigma}_{s_r} \leq 2\tau\gamma^j\bar{\sigma}_1.
\end{aligned}$$

□

Lemma D.17. *With the condition of Theorem 5.5, let $\mathcal{L} \in \mathbb{R}^{n\times n\times q}$ and $\mathcal{S} \in \mathbb{R}^{n\times n\times q}$ be two symmetric tensors satisfying Assumptions 5.1 and 5.2, respectively. Let $\tilde{\mathcal{L}}_k \in \mathbb{R}^{n\times n\times q}$ be the trim output of $\mathcal{L}_k$. If*

$$\|\mathcal{L} - \mathcal{L}_k\| \leq 8\alpha\mu s_r\gamma^k\bar{\sigma}_1, \ \|\mathcal{S} - \mathcal{S}_k\|_\infty \leq \frac{\mu s_r}{nq}\gamma^k\bar{\sigma}_1, \text{ and } \quad supp(\mathcal{S}_k) \subset \Omega,$$

then we have $\|\mathcal{L} - \mathcal{L}_{k+1}\| \leq 8\alpha\mu s_r\gamma^{k+1}\bar{\sigma}_1$, provided $1 > \gamma \geq 512\tau s_r\kappa^2 + \frac{1}{\sqrt{12}}$.

Proof. A direct calculation yields

$$\begin{aligned}
&\|\mathcal{L} - \mathcal{L}_{k+1}\| \leq \|\mathcal{L} - P_{T_k}(\mathcal{D} - \mathcal{S}_k)\| + \|P_{T_k}(\mathcal{D} - \mathcal{S}_k) - \mathcal{L}_{k+1}\| \leq 2\|\mathcal{L} - P_{T_k}(\mathcal{D} - \mathcal{S}_k)\| \\
=& 2\|\mathcal{L} - P_{T_k}(\mathcal{L} + \mathcal{S} - \mathcal{S}_k)\| = 2\|(P_{T_k} - I)\mathcal{L} + P_{T_k}(\mathcal{S} - \mathcal{S}_k)\| \leq 2 \cdot \tau\gamma^{k+1}\bar{\sigma}_{s_r} = 2 \cdot 4\alpha\mu s_r\kappa\gamma^{k+1}\bar{\sigma}_{rq} = 8\alpha\mu s_r\gamma^{k+1}\bar{\sigma}_1,
\end{aligned}$$

where the second inequality follows from the fact $\mathcal{L}_{k+1} = H_r(P_{T_k}(\mathcal{D} - \mathcal{S}_k))$ is the best multi-rank $\boldsymbol{r}$ approximation of $P_{T_k}(\mathcal{D} - \mathcal{S}_k)$, and the last inequality follows from Lemma D.15. □

Lemma D.18. *With the condition of Theorem 5.5, let $\mathcal{L} \in \mathbb{R}^{n\times n\times q}$ and $\mathcal{S} \in \mathbb{R}^{n\times n\times q}$ be two symmetric tensors satisfying Assumptions 5.1, 5.2. Let $\tilde{\mathcal{L}}_k \in \mathbb{R}^{n\times n\times q}$ be the trim output of $\mathcal{L}_k$. If*

$$\|\mathcal{L} - \mathcal{L}_k\| \leq 8\alpha\mu s_r\gamma^k\bar{\sigma}_1, \ \|\mathcal{S} - \mathcal{S}_k\|_\infty \leq \frac{\mu s_r}{nq}\gamma^k\bar{\sigma}_1, \text{ and } supp(\mathcal{S}_k) \subset \Omega,$$

then we have $\|\mathcal{L} - \mathcal{L}_{k+1}\|_\infty \leq \left(\frac{1}{2} - \tau\right)\frac{\mu s_r}{nq}\gamma^{k+1}\bar{\sigma}_1$, provided $1 > \gamma \geq \max\left\{512\tau s_r\kappa^2 + \frac{1}{\sqrt{12}}, \frac{2v}{(1-12\tau)(1-\tau-v)^2}\right\}$ and $\tau < \frac{1}{12}$.

Proof. Let $P_{T_k}(\mathcal{D}-\mathcal{S}_k) = \begin{bmatrix}\mathcal{U}_{k+1} & \ddot{\mathcal{U}}_{k+1}\end{bmatrix} \diamond_\Phi \begin{bmatrix}\Lambda & 0 \\ 0 & \ddot{\Lambda}\end{bmatrix} \diamond_\Phi \begin{bmatrix}\mathcal{U}_{k+1}^H \\ \ddot{\mathcal{U}}_{k+1}^H\end{bmatrix} = \mathcal{U}_{k+1} \diamond_\Phi \Lambda \diamond_\Phi \mathcal{U}_{k+1}^H + \ddot{\mathcal{U}}_{k+1} \diamond_\Phi \ddot{\Lambda} \diamond_\Phi \ddot{\mathcal{U}}_{k+1}^H$ be its eigenvalue decomposition. We use the lighter notion $\lambda_i = \Lambda(i,i,:) \in \mathbb{R}^{1\times 1\times q}$ $(1 \le i \le n)$ for the eigen fiber of $P_{T_k}(\mathcal{D}-\mathcal{S}_k)$ at the k-th iteration and assume they are ordered by $\|\lambda_1\|_F \ge \|\lambda_2\|_F \ge \cdots \ge \|\lambda_n\|_F$. Moreover, Λ has its r largest eigenvalues in Frobenius norm, $\mathcal{U}_{k+1}$ contains the first r eigenvectors, and $\ddot{\mathcal{U}}_{k+1}$ has the rest. Also, the multi-rank of $\mathcal{U}_{k+1} \diamond_\Phi \Lambda \diamond_\Phi \mathcal{U}_{k+1}^H$ is $\boldsymbol{r}$. It follows that $\mathcal{L}_{k+1} = H_r(P_{T_k}(\mathcal{D}-\mathcal{S}_k)) = \mathcal{U}_{k+1} \diamond_\Phi \Lambda \diamond_\Phi \mathcal{U}_{k+1}^H$.
Denote $\mathcal{Z} = P_{T_k}(\mathcal{D}-\mathcal{S}_k) - \mathcal{L} = (P_{T_k} - I)\mathcal{L} + P_{T_k}(\mathcal{S}-\mathcal{S}_k)$. Let $u_i = \mathcal{U}_{k+1}(:,i,:)$ be the i^{th} eigenvector of $P_{T_k}(\mathcal{D}-\mathcal{S}_k)$. That is, $P_{T_k}(\mathcal{D}-\mathcal{S}_k) \diamond_\Phi u_i = u_i \diamond_\Phi \lambda_i$. Then we denote $P_{T_k}(\mathcal{D}-\mathcal{S}_k)$ as $\mathcal{M}$

$$\bar{\mathcal{M}}\bar{u}_i = \bar{u}_i\bar{\lambda}_i$$

which is

$$\begin{bmatrix}\bar{\mathcal{M}}^{(1)} & 0 & \cdots & 0 \\ 0 & \bar{\mathcal{M}}^{(2)} & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & \bar{\mathcal{M}}^{(q)}\end{bmatrix} \begin{bmatrix}\bar{u}_i^{(1)} & 0 & \cdots & 0 \\ 0 & \bar{u}_i^{(2)} & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & \bar{u}_i^{(q)}\end{bmatrix} = \begin{bmatrix}\bar{u}_i^{(1)} & 0 & \cdots & 0 \\ 0 & \bar{u}_i^{(2)} & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & \bar{u}_i^{(q)}\end{bmatrix} \begin{bmatrix}\bar{\lambda}_i^{(1)} & 0 & \cdots & 0 \\ 0 & \bar{\lambda}_i^{(2)} & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & \bar{\lambda}_i^{(q)}\end{bmatrix}$$

This means $\bar{\mathcal{M}}^{(c)}\bar{u}_i^{(c)} = \bar{\lambda}_i^{(c)}\bar{u}_i^{(c)}$. Because $\mathcal{M}^{(c)} = P_{T_k}(\mathcal{D}-\mathcal{S}_k)^{(c)} = \mathcal{Z}^{(c)} + \mathcal{L}^{(c)}$, we get $(\bar{\lambda}_i^{(c)} I - \bar{\mathcal{Z}}^{(c)})\bar{u}_i^{(c)} = \bar{\mathcal{L}}^{(c)}\bar{u}_i^{(c)}$. Then we have

$$\bar{u}_i^{(c)} = \left(I - \frac{\bar{\mathcal{Z}}^{(c)}}{\bar{\lambda}_i^{(c)}}\right)^{-1} \frac{\bar{\mathcal{L}}^{(c)}}{\bar{\lambda}_i^{(c)}}\bar{u}_i^{(c)} = \left(I + \frac{\bar{\mathcal{Z}}^{(c)}}{\bar{\lambda}_i^{(c)}} + \left(\frac{\bar{\mathcal{Z}}^{(c)}}{\bar{\lambda}_i^{(c)}}\right)^2 + \cdots\right) \frac{\bar{\mathcal{L}}^{(c)}}{\bar{\lambda}_i^{(c)}}\bar{u}_i^{(c)}$$

Combining all q slices, we have

$$\bar{u}_i = \left(I + \bar{\mathcal{E}}_i\bar{\mathcal{Z}} + (\bar{\mathcal{E}}_i\bar{\mathcal{Z}})^2 + \cdots\right)\bar{\mathcal{E}}_i\bar{\mathcal{L}}\bar{u}_i$$

where

$$\bar{\mathcal{E}}_i = \begin{bmatrix}\frac{1}{\bar{\lambda}_i^{(1)}}I & 0 & \cdots & 0 \\ 0 & \frac{1}{\bar{\lambda}_i^{(2)}}I & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & \frac{1}{\bar{\lambda}_i^{(q)}}I\end{bmatrix}, \quad \bar{\mathcal{Z}} = \begin{bmatrix}\bar{\mathcal{Z}}^{(1)} & 0 & \cdots & 0 \\ 0 & \bar{\mathcal{Z}}^{(2)} & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & \bar{\mathcal{Z}}^{(q)}\end{bmatrix}$$

The above inequality is valid because of Lemmas D.15 and D.16 (because $\mathcal{Z}$ is symmetric, its eigenvalues and singular values coincide):

$$\|\bar{\mathcal{E}}_i\bar{\mathcal{Z}}\| \le \frac{\|Z\|}{|\bar{\lambda}_{s_r}|} \le \frac{\tau\gamma^{k+1}\bar{\sigma}_{s_r}}{(1-\tau)\bar{\sigma}_{s_r}} = \frac{\tau}{1-\tau} < 1.$$

Then we get $u_i = \left(\mathcal{I} + \mathcal{E}_i \diamond_\Phi \mathcal{Z} + (\mathcal{E}_i \diamond_\Phi \mathcal{Z})^2 + \cdots\right) \diamond_\Phi \mathcal{E}_i \diamond_\Phi \mathcal{L} \diamond_\Phi u_i$ which implies

$$\begin{aligned}
&\mathcal{U}_{k+1} \diamond_\Phi \Lambda \diamond_\Phi \mathcal{U}_{k+1}^H = \sum_{i=1}^r u_i \diamond_\Phi \lambda_i \diamond_\Phi u_i^H \\
&= \sum_{i=1}^r \left(\sum_{a\ge 0} (\mathcal{E}_i \diamond_\Phi \mathcal{Z})^a \diamond_\Phi \mathcal{E}_i \diamond_\Phi \mathcal{L}\right) \diamond_\Phi u_i \diamond_\Phi \lambda_i \diamond_\Phi u_i^H \left(\sum_{b\ge 0} (\mathcal{E}_i \diamond_\Phi \mathcal{Z})^b \diamond_\Phi \mathcal{E}_i \diamond_\Phi \mathcal{L}\right)^H \\
&= \sum_{a\ge 0} \mathcal{Z}^a \diamond_\Phi \mathcal{L} \diamond_\Phi \sum_{i=1}^r \left(\mathcal{E}_i^{a+1} \diamond_\Phi u_i \diamond_\Phi \lambda_i \diamond_\Phi u_i^H \diamond_\Phi \mathcal{E}_i^{b+1}\right) \diamond_\Phi \mathcal{L} \diamond_\Phi \sum_{b\ge 0} \mathcal{Z}^b
\end{aligned}$$

To simplify the above formula, we have

$$\bar{\mathcal{E}}_i \bar{u}_i = \begin{bmatrix} \frac{1}{\bar{\lambda}_i^{(1)}} I & 0 & \cdots & 0 \\ 0 & \frac{1}{\bar{\lambda}_i^{(2)}} I & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & \frac{1}{\bar{\lambda}_i^{(q)}} I \end{bmatrix} \begin{bmatrix} \bar{u}_i^{(1)} & 0 & \cdots & 0 \\ 0 & \bar{u}_i^{(2)} & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & \bar{u}_i^{(q)} \end{bmatrix}$$
$$= \begin{bmatrix} \bar{u}_i^{(1)} & 0 & \cdots & 0 \\ 0 & \bar{u}_i^{(2)} & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & \bar{u}_i^{(q)} \end{bmatrix} \begin{bmatrix} \frac{1}{\bar{\lambda}_i^{(1)}} & 0 & \cdots & 0 \\ 0 & \frac{1}{\bar{\lambda}_i^{(2)}} & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & \frac{1}{\bar{\lambda}_i^{(q)}} \end{bmatrix} = \bar{u}_i \bar{\mathcal{E}}_i.$$

Then we have

$$\begin{aligned} &\bar{\mathcal{E}}_i^{a+1} \bar{u}_i \bar{\lambda}_i \bar{u}_i^H \bar{\mathcal{E}}_i^{b+1} \\ =&\bar{u}_i \begin{bmatrix} \left(\frac{1}{\bar{\lambda}_i^{(1)}}\right)^{a+1} & 0 & \cdots & 0 \\ 0 & \left(\frac{1}{\bar{\lambda}_i^{(2)}}\right)^{a+1} & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & \left(\frac{1}{\bar{\lambda}_i^{(q)}}\right)^{a+1} \end{bmatrix} \begin{bmatrix} \bar{\lambda}_i^{(1)} & 0 & \cdots & 0 \\ 0 & \bar{\lambda}_i^{(2)} & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & \bar{\lambda}_i^{(q)} \end{bmatrix} \cdot \\ &\begin{bmatrix} \left(\frac{1}{\bar{\lambda}_i^{(1)}}\right)^{b+1} & 0 & \cdots & 0 \\ 0 & \left(\frac{1}{\bar{\lambda}_i^{(2)}}\right)^{b+1} & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & \left(\frac{1}{\bar{\lambda}_i^{(q)}}\right)^{b+1} \end{bmatrix} \bar{u}_i^H \\ =&\bar{u}_i \begin{bmatrix} \left(\frac{1}{\bar{\lambda}_i^{(1)}}\right)^{a+b+1} & 0 & \cdots & 0 \\ 0 & \left(\frac{1}{\bar{\lambda}_i^{(2)}}\right)^{a+b+1} & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & \left(\frac{1}{\bar{\lambda}_i^{(q)}}\right)^{a+b+1} \end{bmatrix} \bar{u}_i^H = \bar{u}_i \bar{\Gamma}_i^{(a+b+1)} \bar{u}_i^H, \end{aligned}$$

where we introduce new notation $\bar{\Gamma}_i^{(a+b+1)}$. Hence we have

$$\mathcal{E}_i^{a+1} \diamond_\Phi u_i \diamond_\Phi \lambda_i \diamond_\Phi u_i^H \diamond_\Phi \mathcal{E}_i^{b+1} = u_i \diamond_\Phi \Gamma_i^{(a+b+1)} \diamond_\Phi u_i^H.$$

Now, we get

$$\begin{aligned} \mathcal{U}_{k+1} \diamond_\Phi \Lambda \diamond_\Phi \mathcal{U}_{k+1}^H &= \sum_{a\geq 0} \mathcal{Z}^a \diamond_\Phi \mathcal{L} \diamond_\Phi \sum_{i=1}^r \left(\mathcal{E}_i^{a+1} \diamond_\Phi u_i \diamond_\Phi \lambda_i \diamond_\Phi u_i^H \diamond_\Phi \mathcal{E}_i^{b+1}\right) \diamond_\Phi \mathcal{L} \diamond_\Phi \sum_{b\geq 0} \mathcal{Z}^b \\ &= \sum_{a\geq 0} \mathcal{Z}^a \diamond_\Phi \mathcal{L} \diamond_\Phi \sum_{i=1}^r \left(u_i \diamond_\Phi \Gamma_i^{(a+b+1)} \diamond_\Phi u_i^H\right) \diamond_\Phi \mathcal{L} \diamond_\Phi \sum_{b\geq 0} \mathcal{Z}^b \\ &= \sum_{a,b\geq 0} \mathcal{Z}^a \diamond_\Phi \mathcal{L} \diamond_\Phi \mathcal{U}_{k+1} \diamond_\Phi \Gamma^{(a+b+1)} \diamond_\Phi \mathcal{U}_{k+1}^H \diamond_\Phi \mathcal{L} \diamond_\Phi \mathcal{Z}^b \end{aligned}$$

Thus, we have

$$
\begin{aligned}
&\|\mathcal{L}_{k+1}-\mathcal{L}\|_\infty = \|\mathcal{U}_{k+1}\diamond_\Phi \Lambda \diamond_\Phi \mathcal{U}_{k+1}^H - \mathcal{L}\|_\infty \\
=&\|\mathcal{L}\diamond_\Phi \mathcal{U}_{k+1}\diamond_\Phi \Gamma^{(1)}\diamond_\Phi \mathcal{U}_{k+1}^H \diamond_\Phi \mathcal{L} - \mathcal{L} + \sum_{a+b>0} \mathcal{Z}^a \diamond_\Phi \mathcal{L}\diamond_\Phi \mathcal{U}_{k+1}\diamond_\Phi \Gamma^{(a+b+1)}\diamond_\Phi \mathcal{U}_{k+1}^H\diamond_\Phi \mathcal{L}\diamond_\Phi \mathcal{Z}^b\|_\infty \\
\leq&\|\mathcal{L}\diamond_\Phi \mathcal{U}_{k+1}\diamond_\Phi \Gamma^{(1)}\diamond_\Phi \mathcal{U}_{k+1}^H \diamond_\Phi \mathcal{L} - \mathcal{L}\|_\infty + \sum_{a+b>0} \|\mathcal{Z}^a \diamond_\Phi \mathcal{L}\diamond_\Phi \mathcal{U}_{k+1}\diamond_\Phi \Gamma^{(a+b+1)}\diamond_\Phi \mathcal{U}_{k+1}^H\diamond_\Phi \mathcal{L}\diamond_\Phi \mathcal{Z}^b\|_\infty \\
:=&Y_0 + \sum_{a+b>0} Y_{ab}
\end{aligned}
$$

We will handle Y_0 first. Recall that $\mathcal{L}=\mathcal{U}\diamond_\Phi \Sigma \diamond_\Phi \mathcal{V}^H$ is the transformed t-SVD of the symmetric tensor $\mathcal{L}$ which obeys μ-incoherence, i.e., $\mathcal{U}\diamond_\Phi \mathcal{U}^H = \mathcal{V}\diamond_\Phi \mathcal{V}^H$ and $\|\dot{e}_i^H \diamond_\Phi \mathcal{U}\diamond_\Phi \mathcal{U}^H\|_F \leq \sqrt{\frac{\mu s_r}{nq}}$ for all i. So for each (i,j,k) entry of Y_0, one has

$$
\begin{aligned}
&Y_0 = \max_{i,j,k} \left|\left\langle \mathcal{L}\diamond_\Phi \mathcal{U}_{k+1}\diamond_\Phi \Gamma^{(1)}\diamond_\Phi \mathcal{U}_{k+1}^H\diamond_\Phi \mathcal{L} - \mathcal{L}, \mathcal{E}_{ijk}\right\rangle\right| \\
=&\max_{i,j,k} \left|\left\langle \dot{e}_i^H \diamond_\Phi \mathcal{U}\diamond_\Phi \mathcal{U}^H \diamond_\Phi (\mathcal{L}\diamond_\Phi \mathcal{U}_{k+1}\diamond_\Phi \Gamma^{(1)}\diamond_\Phi \mathcal{U}_{k+1}^H\diamond_\Phi \mathcal{L} - \mathcal{L})\diamond_\Phi \mathcal{U}\diamond_\Phi \mathcal{U}^H\diamond_\Phi \dot{e}_j, e_k\right\rangle\right| \\
\leq&\max_{i,j} \|\dot{e}_i^H \diamond_\Phi \mathcal{U}\diamond_\Phi \mathcal{U}^H \diamond_\Phi (\mathcal{L}\diamond_\Phi \mathcal{U}_{k+1}\diamond_\Phi \Gamma^{(1)}\diamond_\Phi \mathcal{U}_{k+1}^H\diamond_\Phi \mathcal{L} - \mathcal{L})\diamond_\Phi \mathcal{U}\diamond_\Phi \mathcal{U}^H\diamond_\Phi \dot{e}_j\|_F \cdot \|e_k\|_F \\
\leq&\max_{i,j} \|\dot{e}_i^H \diamond_\Phi \mathcal{U}\diamond_\Phi \mathcal{U}^H\|_F \cdot \|\mathcal{L}\diamond_\Phi \mathcal{U}_{k+1}\diamond_\Phi \Gamma^{(1)}\diamond_\Phi \mathcal{U}_{k+1}^H\diamond_\Phi \mathcal{L} - \mathcal{L}\| \cdot \|\mathcal{U}\diamond_\Phi \mathcal{U}^H \diamond_\Phi \dot{e}_j\|_F \\
\leq&\frac{\mu s_r}{nq}\|\mathcal{L}\diamond_\Phi \mathcal{U}_{k+1}\diamond_\Phi \Gamma^{(1)}\diamond_\Phi \mathcal{U}_{k+1}^H\diamond_\Phi \mathcal{L} - \mathcal{L}\|,
\end{aligned}
$$

where the first inequality follows from the fact $\mathcal{U}\diamond_\Phi \mathcal{U}^H \diamond_\Phi \mathcal{L} = \mathcal{L}\diamond_\Phi \mathcal{U}\diamond_\Phi \mathcal{U}^H = \mathcal{L}$. Since $\mathcal{L} = \mathcal{U}_{k+1}\diamond_\Phi \Lambda\diamond_\Phi \mathcal{U}_{k+1}^H + \ddot{\mathcal{U}}_{k+1}\diamond_\Phi \ddot{\Lambda}\diamond_\Phi \ddot{\mathcal{U}}_{k+1}^H - \mathcal{Z}$, there holds that

$$
\begin{aligned}
&\|\mathcal{L}\diamond_\Phi \mathcal{U}_{k+1}\diamond_\Phi \Gamma^{(1)}\diamond_\Phi \mathcal{U}_{k+1}^H\diamond_\Phi \mathcal{L} - \mathcal{L}\| \\
=&\|(\mathcal{U}_{k+1}\diamond_\Phi \Lambda\diamond_\Phi \mathcal{U}_{k+1}^H + \ddot{\mathcal{U}}_{k+1}\diamond_\Phi \ddot{\Lambda}\diamond_\Phi \ddot{\mathcal{U}}_{k+1}^H - \mathcal{Z})\diamond_\Phi \mathcal{U}_{k+1}\diamond_\Phi \Gamma^{(1)}\diamond_\Phi \mathcal{U}_{k+1}^H \\
&\diamond_\Phi (\mathcal{U}_{k+1}\diamond_\Phi \Lambda\diamond_\Phi \mathcal{U}_{k+1}^H + \ddot{\mathcal{U}}_{k+1}\diamond_\Phi \ddot{\Lambda}\diamond_\Phi \ddot{\mathcal{U}}_{k+1}^H - \mathcal{Z}) - \mathcal{L}\| \\
=&\|\mathcal{U}_{k+1}\diamond_\Phi \Lambda\diamond_\Phi \Gamma^{(1)}\diamond_\Phi \Lambda\diamond_\Phi \mathcal{U}_{k+1}^H - \mathcal{L} - \mathcal{U}_{k+1}\diamond_\Phi \Lambda\diamond_\Phi \Gamma^{(1)}\diamond_\Phi \mathcal{U}_{k+1}^H\diamond_\Phi \mathcal{Z} \\
&- \mathcal{Z}\diamond_\Phi \mathcal{U}_{k+1}\diamond_\Phi \Gamma^{(1)}\diamond_\Phi \Lambda\diamond_\Phi \mathcal{U}_{k+1}^H + \mathcal{Z}\diamond_\Phi \mathcal{U}_{k+1}\diamond_\Phi \Gamma^{(1)}\diamond_\Phi \mathcal{U}_{k+1}^H\diamond_\Phi \mathcal{Z}\| \\
\leq&\|\mathcal{Z} - \ddot{\mathcal{U}}_{k+1}\diamond_\Phi \ddot{\Lambda}\diamond_\Phi \ddot{\mathcal{U}}_{k+1}^H\| + \|\mathcal{U}_{k+1}\diamond_\Phi \Lambda\diamond_\Phi \Gamma^{(1)}\diamond_\Phi \mathcal{U}_{k+1}^H\diamond_\Phi \mathcal{Z}\| \\
&+ \|\mathcal{Z}\diamond_\Phi \mathcal{U}_{k+1}\diamond_\Phi \Gamma^{(1)}\diamond_\Phi \Lambda\diamond_\Phi \mathcal{U}_{k+1}^H\| + \|\mathcal{Z}\diamond_\Phi \mathcal{U}_{k+1}\diamond_\Phi \Gamma^{(1)}\diamond_\Phi \mathcal{U}_{k+1}^H\diamond_\Phi \mathcal{Z}\| \\
=&\|\bar{\mathcal{Z}} - \bar{\ddot{\mathcal{U}}}_{k+1}\bar{\ddot{\Lambda}}\bar{\ddot{\mathcal{U}}}_{k+1}^H\| + \|\bar{\mathcal{U}}_{k+1}\bar{\Lambda}\bar{\Gamma}^{(1)}\bar{\mathcal{U}}_{k+1}^H\bar{\mathcal{Z}}\| + \|\bar{\mathcal{Z}}\bar{\mathcal{U}}_{k+1}\bar{\Gamma}^{(1)}\bar{\Lambda}\bar{\mathcal{U}}_{k+1}^H\| + \|\bar{\mathcal{Z}}\bar{\mathcal{U}}_{k+1}\bar{\Gamma}^{(1)}\bar{\mathcal{U}}_{k+1}^H\bar{\mathcal{Z}}\| \\
\leq&\|\bar{\mathcal{Z}} - \bar{\ddot{\mathcal{U}}}_{k+1}\bar{\ddot{\Lambda}}\bar{\ddot{\mathcal{U}}}_{k+1}^H\| + 2\|\bar{\mathcal{Z}}\| + \frac{\|Z\|^2}{|\bar{\lambda}_{s_r}|} \leq \|\bar{\ddot{\mathcal{U}}}_{k+1}\bar{\ddot{\Lambda}}\bar{\ddot{\mathcal{U}}}_{k+1}^H\| + 4\|\mathcal{Z}\| \\
\leq&|\bar{\lambda}_{s_r+1}| + 4\|\mathcal{Z}\| \leq 5\|\mathcal{Z}\| \leq 5\tau\gamma^{k+1}\bar{\sigma}_1,
\end{aligned}
$$

where the last inequality follows from Lemma D.15, and notice $\frac{\|\mathcal{Z}\|}{|\bar{\lambda}_{s_r}|} \leq \frac{\tau}{1-\tau} < 1$ since $\tau < \frac{1}{2}$, and $|\bar{\lambda}_{s_r+1}| \leq \|\mathcal{Z}\|$ since $\mathcal{L}$ is a multi-rank $\boldsymbol{r}$ tensor (using Tensor Weyl's inequality). Thus we have $Y_0 \leq \frac{\mu s_r}{nq}5\tau\gamma^{k+1}\bar{\sigma}_1$. Next, we derive an upper bound for the rest part. Note that

$$
\begin{aligned}
&Y_{ab} = \max_{i,j,k} \left\langle \mathcal{Z}^a \diamond_\Phi \mathcal{L} \diamond_\Phi \mathcal{U}_{k+1} \diamond_\Phi \Gamma^{(a+b+1)} \diamond_\Phi \mathcal{U}_{k+1}^H \diamond_\Phi \mathcal{L} \diamond_\Phi \mathcal{Z}^b, \mathcal{E}_{ijk} \right\rangle \\
&= \max_{i,j,k} \left\langle (\dot{e}_i^H \diamond_\Phi \mathcal{Z}^a \diamond_\Phi \mathcal{U} \diamond_\Phi \mathcal{U}^H) \diamond_\Phi \mathcal{L} \diamond_\Phi \mathcal{U}_{k+1} \diamond_\Phi \Gamma^{(a+b+1)} \diamond_\Phi \mathcal{U}_{k+1}^H \diamond_\Phi \mathcal{L} \diamond_\Phi (\mathcal{U} \diamond_\Phi \mathcal{U}^H \diamond_\Phi \mathcal{Z}^b \diamond_\Phi \dot{e}_j), e_k \right\rangle \\
&\le \max_{i,j} \|(\dot{e}_i^H \diamond_\Phi \mathcal{Z}^a \diamond_\Phi \mathcal{U} \diamond_\Phi \mathcal{U}^H) \diamond_\Phi \mathcal{L} \diamond_\Phi \mathcal{U}_{k+1} \diamond_\Phi \Gamma^{(a+b+1)} \diamond_\Phi \mathcal{U}_{k+1}^H \diamond_\Phi \mathcal{L} \diamond_\Phi (\mathcal{U} \diamond_\Phi \mathcal{U}^H \diamond_\Phi \mathcal{Z}^b \diamond_\Phi \dot{e}_j)\|_F \cdot \|e_k\|_F \\
&\le \max_{i,j} \|\dot{e}_i^H \diamond_\Phi \mathcal{Z}^a \diamond_\Phi \mathcal{U}\|_F \cdot \|\mathcal{L} \diamond_\Phi \mathcal{U}_{k+1} \diamond_\Phi \Gamma^{(a+b+1)} \diamond_\Phi \mathcal{U}_{k+1}^H \diamond_\Phi \mathcal{L}\| \cdot \|\mathcal{U}^H \diamond_\Phi \mathcal{Z}^b \diamond_\Phi \dot{e}_j\|_F \\
&\le \max_{l} \frac{\mu s_r}{nq} (\sqrt{n}\|\dot{e}_l^H \diamond_\Phi \mathcal{Z}\|_F)^{a+b} \|\mathcal{L} \diamond_\Phi \mathcal{U}_{k+1} \diamond_\Phi \Gamma^{(a+b+1)} \diamond_\Phi \mathcal{U}_{k+1}^H \diamond_\Phi \mathcal{L}\|,
\end{aligned}
$$

where the last inequality follows from Lemma D.14. Furthermore, by using $\mathcal{L} = \mathcal{U}_{k+1} \diamond_\Phi \Lambda \diamond_\Phi \mathcal{U}_{k+1}^H + \ddot{\mathcal{U}}_{k+1} \diamond_\Phi \ddot{\Lambda} \diamond_\Phi \ddot{\mathcal{U}}_{k+1}^H - \mathcal{Z}$ again, we get

$$
\begin{aligned}
&\|\mathcal{L} \diamond_\Phi \mathcal{U}_{k+1} \diamond_\Phi \Gamma^{(a+b+1)} \diamond_\Phi \mathcal{U}_{k+1}^H \diamond_\Phi \mathcal{L}\| \\
=&\|(\mathcal{U}_{k+1} \diamond_\Phi \Lambda \diamond_\Phi \mathcal{U}_{k+1}^H + \ddot{\mathcal{U}}_{k+1} \diamond_\Phi \ddot{\Lambda} \diamond_\Phi \ddot{\mathcal{U}}_{k+1}^H - \mathcal{Z}) \diamond_\Phi \mathcal{U}_{k+1} \diamond_\Phi \Gamma^{(a+b+1)} \diamond_\Phi \mathcal{U}_{k+1}^H \\
&\diamond_\Phi (\mathcal{U}_{k+1} \diamond_\Phi \Lambda \diamond_\Phi \mathcal{U}_{k+1}^H + \ddot{\mathcal{U}}_{k+1} \diamond_\Phi \ddot{\Lambda} \diamond_\Phi \ddot{\mathcal{U}}_{k+1}^H - \mathcal{Z})\| \\
=&\|\mathcal{U}_{k+1} \diamond_\Phi \Lambda \diamond_\Phi \Gamma^{(a+b+1)} \diamond_\Phi \Lambda \diamond_\Phi \mathcal{U}_{k+1}^H - \mathcal{U}_{k+1} \diamond_\Phi \Lambda \diamond_\Phi \Gamma^{(a+b+1)} \diamond_\Phi \mathcal{U}_{k+1}^H \diamond_\Phi \mathcal{Z} \\
&- \mathcal{Z} \diamond_\Phi \mathcal{U}_{k+1} \diamond_\Phi \Gamma^{(a+b+1)} \diamond_\Phi \Lambda \diamond_\Phi \mathcal{U}_{k+1}^H + \mathcal{Z} \diamond_\Phi \mathcal{U}_{k+1} \diamond_\Phi \Gamma^{(a+b+1)} \diamond_\Phi \mathcal{U}_{k+1}^H \diamond_\Phi \mathcal{Z}\| \\
\le&\|\bar{\mathcal{U}}_{k+1}\bar{\Lambda}\bar{\Gamma}^{(a+b+1)}\bar{\Lambda}\bar{\mathcal{U}}_{k+1}^H - \bar{\mathcal{U}}_{k+1}\bar{\Lambda}\bar{\Gamma}^{(a+b+1)}\bar{\mathcal{U}}_{k+1}^H\bar{\mathcal{Z}} - \bar{\mathcal{Z}}\bar{\mathcal{U}}_{k+1}\bar{\Gamma}^{(a+b+1)}\bar{\Lambda}\bar{\mathcal{U}}_{k+1}^H + \bar{\mathcal{Z}}\bar{\mathcal{U}}_{k+1}\bar{\Gamma}^{(a+b+1)}\bar{\mathcal{U}}_{k+1}^H\bar{\mathcal{Z}}\| \\
\le&|\bar{\lambda}_{s_r}|^{-(a+b-1)} + |\bar{\lambda}_{s_r}|^{-(a+b)}\|\mathcal{Z}\| + |\bar{\lambda}_{s_r}|^{-(a+b)}\|\mathcal{Z}\| + |\bar{\lambda}_{s_r}|^{-(a+b+1)}\|\mathcal{Z}\|^2 \\
=&|\bar{\lambda}_{s_r}|^{-(a+b-1)} \left(1 + \frac{2\|\mathcal{Z}\|}{\bar{\lambda}_{s_r}} + \left(\frac{\|\mathcal{Z}\|}{|\bar{\lambda}_{s_r}|}\right)^2\right) = |\bar{\lambda}_{s_r}|^{-(a+b-1)} \left(1 + \frac{\|\mathcal{Z}\|}{|\bar{\lambda}_{s_r}|}\right)^2 \\
\le&|\bar{\lambda}_{s_r}|^{-(a+b-1)} \left(\frac{1}{1-\tau}\right)^2 \le \left(\frac{1}{1-\tau}\right)^2 ((1-\tau)\bar{\sigma}_{s_r})^{-(a+b-1)},
\end{aligned}
$$

where the last inequality follows from $\frac{\|\mathcal{Z}\|}{|\bar{\lambda}_{rq}|} \le \frac{\tau}{1-\tau}$, and the last inequality follows from Lemma D.16. Together with Lemma D.15, we have

$$
\begin{aligned}
&\sum_{a+b>0} Y_{ab} = \sum_{a+b>0} \frac{\mu s_r}{nq} \left(\frac{1}{1-\tau}\right)^2 v\gamma^k \bar{\sigma}_{s_r} \left(\frac{v\gamma^k \bar{\sigma}_{s_r}}{(1-\tau)\bar{\sigma}_{s_r}}\right)^{a+b-1} \\
\le&\frac{\mu s_r}{nq} \left(\frac{1}{1-\tau}\right)^2 v\gamma^k \bar{\sigma}_1 \sum_{a+b>0} \left(\frac{v}{1-\tau}\right)^{a+b-1} \le \frac{\mu s_r}{nq} \left(\frac{1}{1-\tau}\right)^2 v\gamma^k \bar{\sigma}_1 \left(\frac{1}{1-\frac{v}{1-\tau}}\right)^2 \\
\le&\frac{\mu s_r}{nq} \left(\frac{1}{1-\tau-v}\right)^2 v\gamma^k \bar{\sigma}_1.
\end{aligned}
$$

Finally, combining them together gives

$$
\|\mathcal{L}_{k+1} - \mathcal{L}\|_\infty = Y_0 + \sum_{a+b>0} Y_{ab} = \frac{\mu s_r}{nq} 5\tau\gamma^{k+1}\bar{\sigma}_1 + \frac{\mu s_r}{nq} \left(\frac{1}{1-\tau-v}\right)^2 v\gamma^k \bar{\sigma}_1 \le \left(\frac{1}{2} - \tau\right) \frac{\mu s_r}{nq} \gamma^{k+1} \bar{\sigma}_1,
$$

where the last inequality follows from $\gamma \ge \frac{2v}{(1-12\tau)(1-\tau-v)^2}$. □

Lemma D.19. *Let $\mathcal{L} \in \mathbb{R}^{n\times n\times q}$ and $\mathcal{S} \in \mathbb{R}^{n\times n\times q}$ be two symmetric tensors satisfying Assumptions 5.1 and 5.2, respectively. Let $\tilde{\mathcal{L}}_k \in \mathbb{R}^{n\times n\times q}$ be the trim output of $\mathcal{L}_k$. Recall that $\beta = \frac{\mu s_r}{2nq}$. If*

$$
\|\mathcal{L} - \mathcal{L}_k\| \le 8\alpha\mu\gamma^k\bar{\sigma}_1, \ \|\mathcal{S} - \mathcal{S}_k\|_\infty \le \frac{\mu s_r}{nq}\gamma^k\bar{\sigma}_1, \ \textit{and } supp(\mathcal{S}_k) \subset \Omega,
$$

then we have

$$\textit{supp}(\mathcal{S}_{k+1}) \subset \Omega \textit{ and } \|\mathcal{S} - \mathcal{S}_{k+1}\|_\infty \le \frac{\mu s_r}{nq}\gamma^{k+1}\bar{\sigma}_1$$

provided $1 > \gamma \ge \max\left\{512\tau s_r\kappa^2 + \frac{1}{\sqrt{12}}, \frac{2v}{(1-12\tau)(1-\tau-v)^2}\right\}$ *and* $\tau < \frac{1}{12}$.

Proof. We first notice that

$$[\mathcal{S}]_{ijk} = [T_{\zeta_{k+1}}(\mathcal{D} - \mathcal{L}_{k+1})]_{ijk} = [T_{\zeta_{k+1}}(\mathcal{S} + \mathcal{L} - \mathcal{L}_{k+1})]_{ijk} = \begin{cases} T_{\zeta_{k+1}}([\mathcal{S} + \mathcal{L} - \mathcal{L}_{k+1}]_{ijk}) & (i,j,k) \in \Omega \\ T_{\zeta_{k+1}}([\mathcal{L} - \mathcal{L}_{k+1}]_{ijk}) & (i,j,k) \in \Omega^c \end{cases}.$$

Let $|\bar{\lambda}_i|$ denote the i^{th} singular value of $P_{\bar{T}_k}(\bar{\mathcal{D}} - \bar{\mathcal{S}}_k)$. By Lemmas D.16 and D.18, we have

$$|[\mathcal{L} - \mathcal{L}_{k+1}]_{ij}| \le \|\mathcal{L} - \mathcal{L}_{k+1}\|_\infty \le \left(\frac{1}{2} - \tau\right)\frac{\mu s_r}{nq}\gamma^{k+1}\bar{\sigma}_1 \le \left(\frac{1}{2} - \tau\right)\frac{\mu s_r}{nq}\frac{1}{1-2\tau}\left(|\bar{\lambda}_{s_r+1}^{(k)}| + \gamma^{k+1}|\bar{\lambda}_1^{(k)}|\right) = \zeta_{k+1}$$

for any entry of $\mathcal{L} - \mathcal{L}_{k+1}$. Hence, $[\mathcal{S}_{k+1}]_{ijk} = 0$ for all $(i,j,k) \in \Omega^c$, i.e., supp$(\mathcal{S}_{k+1}) \subset \Omega$.
Denote $\Omega_{k+1} := \text{supp}(\mathcal{S}_{k+1}) = \{(i,j,k) | [\mathcal{D} - \mathcal{L}_{k+1}]_{ijk} > \zeta_k\}$. Then for any entry of $\mathcal{S} - \mathcal{S}_{k+1}$, there hold

$$[\mathcal{S} - \mathcal{S}_{k+1}]_{ijk} = \begin{cases} 0 \\ [\mathcal{L}_{k+1} - \mathcal{L}]_{ijk} \\ [\mathcal{S}]_{ijk} \end{cases} \le \begin{cases} 0 \\ \|\mathcal{L} - \mathcal{L}_{k+1}\|_\infty \\ \|\mathcal{L} - \mathcal{L}_{k+1}\|_\infty + \zeta_{k+1} \end{cases} \le \begin{cases} 0 & (i,j,k) \in \Omega^c \\ \left(\frac{1}{2} - \tau\right)\frac{\mu s_r}{nq}\gamma^{k+1}\bar{\sigma}_1 & (i,j,k) \in \Omega_{k+1} \\ \frac{\mu s_r}{nq}\gamma^{k+1}\bar{\sigma}_1 & (i,j,k) \in \Omega \backslash \Omega_{k+1}. \end{cases}$$

Here the last step follows from Lemma D.16 which implies $\zeta_{k+1} = \frac{\mu s_r}{2nq}(|\bar{\lambda}_{s_r+1}^{(k)}| + \gamma^{k+1}|\bar{\lambda}_1^{(k)}|) \le \left(\frac{1}{2} + \tau\right)\frac{\mu s_r}{nq}\gamma^{k+1}\bar{\sigma}_1$. Therefore, $\|\mathcal{S} - \mathcal{S}_{k+1}\|_\infty \le \frac{\mu s_r}{nq}\gamma^{k+1}\bar{\sigma}_1$. □

Theorem D.20 (Local Convergence)**.** *With the condition of Theorem 5.5, let* $\mathcal{L} \in \mathbb{R}^{n\times n\times q}$ *and* $\mathcal{S} \in \mathbb{R}^{n\times n\times q}$ *be two symmetric tensors satisfying Assumptions 5.1 and 5.2. If the initial guess* $\mathcal{L}_0$ *and* $\mathcal{S}_0$ *obey the following conditions:*

$$\|\mathcal{L} - \mathcal{L}_0\| \le 8\alpha\mu s_r\bar{\sigma}_1, \; \|\mathcal{S} - \mathcal{S}_0\|_\infty \le \frac{\mu s_r}{nq}\bar{\sigma}_1, \textit{ and supp}(\mathcal{S}_0) \subset \Omega,$$

then the iterates of Algorithm 2 with parameters $\beta = \frac{\mu s_r}{2nq}$ *and* $\gamma \in (\frac{1}{\sqrt{12}}, 1)$ *satisfy*

$$\|\mathcal{L} - \mathcal{L}_k\| \le 8\alpha\mu s_r\gamma^k\bar{\sigma}_1, \; \|\mathcal{S} - \mathcal{S}_k\|_\infty \le \frac{\mu s_r}{nq}\gamma^k\bar{\sigma}_1, \textit{ and supp}(\mathcal{S}_k) \subset \Omega.$$

Proof. This theorem will be proved by mathematical induction.
Base Case: When $k = 0$, the base case is satisfied by the assumption on the initialization.
Induction Step: Assume we have

$$\|\mathcal{L} - \mathcal{L}_k\| \le 8\alpha\mu s_r\gamma^k\bar{\sigma}_1, \; \|\mathcal{S} - \mathcal{S}_k\|_\infty \le \frac{\mu s_r}{nq}\gamma^k\bar{\sigma}_1, \text{ and supp}(\mathcal{S}_k) \subset \Omega$$

at the k^{th} iteration. At the $(k+1)^{th}$ iteration. It follows directly from Lemmas D.17 and D.19 that

$$\|\mathcal{L} - \mathcal{L}_{k+1}\| \le 8\alpha\mu s_r\gamma^{k+1}\bar{\sigma}_1, \; \|\mathcal{S} - \mathcal{S}_{k+1}\|_\infty \le \frac{\mu s_r}{nq}\gamma^{k+1}\bar{\sigma}_1, \text{ and supp}(\mathcal{S}_{k+1}) \subset \Omega.$$

Additionally, notice that we overall require $1 > \gamma \ge 512\tau s_r\kappa^2 + \frac{1}{\sqrt{12}}$. By the definition of τ and v, one can easily see that the lower bound approaches $\frac{1}{\sqrt{12}}$ when the constant hidden in bound of α in Theorem 5.5 is sufficiently large. Therefore, the theorem can be proved for any $\gamma \in \left(\frac{1}{\sqrt{12}}, 1\right)$. □

D.2.4. Local Convergence of Alternating Projection in APT

Lemma D.21. *With the condition of Theorem 5.4, let $\mathcal{L}, \mathcal{S}$ be the symmetric tensors and satisfy Assumptions 5.2 and 5.1. Let $\mathcal{S}_k$ bethe k-th iteration of our algorithm. Suppose (1) $\|\mathcal{S}-\mathcal{S}_k\|_\infty \le \frac{4\mu s_r}{nq}\gamma^{k-1}\bar{\sigma}_1$, (2) supp$(\mathcal{S}-\mathcal{S}_k) \subset$ supp$(\mathcal{S})$. Then $\frac{127}{128}\bar{\sigma}_1 \le |\bar{\lambda}_1| \le \frac{129}{128}\bar{\sigma}_1$.*

Proof. Note that $\mathcal{D}-\mathcal{S}_k = \mathcal{L}+(\mathcal{S}-\mathcal{S}_k)$. With Lemmas D.3 and D.4, we have

$$\left|\bar{\lambda}_{i,c}-\bar{\sigma}_{i,c}\right| \le \|\mathcal{S}-\mathcal{S}_k\|_2 \le 4\alpha\mu s_r\gamma^{k-1}\bar{\sigma}_1 \le \frac{1}{128}\bar{\sigma}_1 \tag{18}$$

where the last inequality follows from the bound of α (hidden constant in Theorem 5.4 is $\frac{1}{512}$).

If $\bar{\sigma}_{i,c} = \bar{\sigma}_1$, we have $\frac{127}{128}\bar{\sigma}_1 \le |\bar{\lambda}_{i,c}| \le |\bar{\lambda}_1|$.

if $\bar{\lambda}_{i,c} = \bar{\lambda}_1$, we have $|\bar{\lambda}_1| = |\bar{\lambda}_{i,c}| \le \frac{4}{512}\bar{\sigma}_1 + \bar{\sigma}_{i,c} \le \frac{129}{128}\bar{\sigma}_1$. □

Lemma D.22. *Suppose $\gamma \ge \frac{1}{16}$, (1). $\|\mathcal{S}-\mathcal{S}_k\|_\infty \le \frac{4\mu s_r}{nq}\gamma^{k-1}\bar{\sigma}_1$, (2). supp$(\mathcal{S}-\mathcal{S}_k) \subset$ supp$(\mathcal{S})$. Then we have $\|\mathcal{L}_{k+1}-\mathcal{L}\|_\infty \le \frac{\mu s_r}{nq}\gamma^k\bar{\sigma}_1$.*

Proof. Let $\mathcal{D}-\mathcal{S}_k = \left[\mathcal{U}_{k+1}, \ddot{\mathcal{U}}_{k+1}\right] \diamond_\Phi \begin{bmatrix}\Lambda & 0\\ 0 & \ddot{\Lambda}\end{bmatrix} \diamond_\Phi \left[\mathcal{U}_{k+1}, \ddot{\mathcal{U}}_{k+1}\right]^H = \mathcal{U}_{k+1}\diamond_\Phi \Lambda \diamond_\Phi \mathcal{U}_{k+1}^H + \ddot{\mathcal{U}}_{k+1}\diamond_\Phi \ddot{\Lambda}\diamond_\Phi \ddot{\mathcal{U}}_{k+1}^H$ be its eigenvalue decomposition, where Λ has the multi-rank $\boldsymbol{r}$ eigenvalues in magnitude and $\ddot{\Lambda}$ contains the rest eigenvalues. Also, $\mathcal{U}_{k+1}$ contains the first $\boldsymbol{r}$ eigenvectors, and $\ddot{\mathcal{U}}_{k+1}$ has the rest. Notice $\mathcal{L}_{k+1} = P_{\mathbb{L}}(\mathcal{D}-\mathcal{S}_{-1}) = \mathcal{U}_{k+1}\diamond_\Phi \Lambda\diamond_\Phi \mathcal{U}_{k+1}^H$ due to the symmetric setting. Also, denote $\mathcal{S}-\mathcal{S}_k$ as $\mathcal{Z}$. Let $u_i = \mathcal{U}_{k+1}(:,i,:)$ be the i^{th} eigenvector of $\mathcal{D}-\mathcal{S}_k = \mathcal{L}+\mathcal{Z}$ and $\lambda_s = \Lambda(s,s,:)$ be the s^{th} eigenvalue of $\mathcal{D}-\mathcal{S}_k = \mathcal{L}+\mathcal{Z}$. Then recall that $\mathcal{D}-\mathcal{S}_k = \mathcal{L}+\mathcal{Z}$. We denote $\mathcal{D}-\mathcal{S}_k$ as $\mathcal{M}$, then

$$\bar{\mathcal{M}}\bar{u}_i = \bar{u}_i\bar{\lambda}_i,$$

which is

$$\begin{bmatrix}\bar{\mathcal{M}}^{(1)} & 0 & \cdots & 0\\ 0 & \bar{\mathcal{M}}^{(2)} & \cdots & 0\\ 0 & 0 & \cdots & 0\\ 0 & 0 & \cdots & \bar{\mathcal{M}}^{(q)}\end{bmatrix}\begin{bmatrix}\bar{u}_i^{(1)} & 0 & \cdots & 0\\ 0 & \bar{u}_i^{(2)} & \cdots & 0\\ 0 & 0 & \cdots & 0\\ 0 & 0 & \cdots & \bar{u}_i^{(q)}\end{bmatrix} = \begin{bmatrix}\bar{u}_i^{(1)} & 0 & \cdots & 0\\ 0 & \bar{u}_i^{(2)} & \cdots & 0\\ 0 & 0 & \cdots & 0\\ 0 & 0 & \cdots & \bar{u}_i^{(q)}\end{bmatrix}\begin{bmatrix}\bar{\lambda}_i^{(1)} & 0 & \cdots & 0\\ 0 & \bar{\lambda}_i^{(2)} & \cdots & 0\\ 0 & 0 & \cdots & 0\\ 0 & 0 & \cdots & \bar{\lambda}_i^{(q)}\end{bmatrix}.$$

This means $\bar{\mathcal{M}}^{(c)}\bar{u}_i^{(c)} = \bar{\lambda}_i^{(c)}\bar{u}_i^{(c)}$. Because $\mathcal{D}^{(c)}-\mathcal{S}_k^{(c)} = \mathcal{Z}^{(c)}+\mathcal{L}^{(c)}$, we get $(\bar{\lambda}_i^{(c)}I-\bar{\mathcal{Z}}^{(c)})\bar{u}_i^{(c)} = \bar{\mathcal{L}}^{(c)}\bar{u}_i^{(c)}$. Then we have

$$\bar{u}_i^{(c)} = \left(I-\frac{\bar{\mathcal{Z}}^{(c)}}{\bar{\lambda}_i^{(c)}}\right)^{-1}\frac{\bar{\mathcal{L}}^{(c)}}{\bar{\lambda}_i^{(c)}}\bar{u}_i^{(c)} = \left(I+\frac{\bar{\mathcal{Z}}^{(c)}}{\bar{\lambda}_i^{(c)}}+\left(\frac{\bar{\mathcal{Z}}^{(c)}}{\bar{\lambda}_i^{(c)}}\right)^2+\cdots\right)\frac{\bar{\mathcal{L}}^{(c)}}{\bar{\lambda}_i^{(c)}}\bar{u}_i^{(c)}.$$

Combining all q slices, we have $\bar{u}_i = \left(I+\bar{\mathcal{E}}_i\bar{\mathcal{Z}}+(\bar{\mathcal{E}}_i\bar{\mathcal{Z}})^2+\cdots\right)\bar{\mathcal{E}}_i\bar{\mathcal{L}}\bar{u}_i$ where

$$\bar{\mathcal{E}}_i = \begin{bmatrix}\frac{1}{\bar{\lambda}_i^{(1)}}I & 0 & \cdots & 0\\ 0 & \frac{1}{\bar{\lambda}_i^{(2)}}I & \cdots & 0\\ 0 & 0 & \cdots & 0\\ 0 & 0 & \cdots & \frac{1}{\bar{\lambda}_i^{(q)}}I\end{bmatrix},\ \bar{\mathcal{Z}} = \begin{bmatrix}\bar{\mathcal{Z}}^{(1)} & 0 & \cdots & 0\\ 0 & \bar{\mathcal{Z}}^{(2)} & \cdots & 0\\ 0 & 0 & \cdots & 0\\ 0 & 0 & \cdots & \bar{\mathcal{Z}}^{(q)}\end{bmatrix}.$$

The above inequality is valid because of $\|\bar{\mathcal{E}}_i\bar{\mathcal{Z}}\| \le \|\bar{\mathcal{Z}}\|\frac{1}{\bar{\lambda}_i^{(q)}} < 1$. To prove the above inequality, by Lemma D.3, we have

$$\begin{aligned}&|\bar{\lambda}_{i,c}-\bar{\sigma}_{i,c}| \le \|\mathcal{Z}\| \Leftrightarrow \bar{\sigma}_{i,c}-\|\mathcal{Z}\| \le \bar{\lambda}_{i,c} \le \bar{\sigma}_{i,c}+\|\mathcal{Z}\|\\ \Leftrightarrow &\frac{1}{\bar{\sigma}_{i,c}+\|\mathcal{Z}\|} \le \frac{1}{\bar{\lambda}_{i,c}} \le \frac{1}{\bar{\sigma}_{i,c}-\|\mathcal{Z}\|} \Leftrightarrow \frac{\|\mathcal{Z}\|}{\bar{\sigma}_{i,c}+\|\mathcal{Z}\|} \le \frac{\|\mathcal{Z}\|}{\bar{\lambda}_{i,c}} \le \frac{\|\mathcal{Z}\|}{\bar{\sigma}_{i,c}-\|\mathcal{Z}\|}.\end{aligned}$$

Then, we have

$$\frac{\|\mathcal{Z}\|}{\bar{\lambda}_{i,c}} \leq \frac{\|\mathcal{Z}\|}{\bar{\sigma}_{i,c} - \|\mathcal{Z}\|} \leq \frac{4\alpha\mu s_r \bar{\sigma}_1}{\bar{\sigma}_{s_r} - 4\alpha\mu s_r \bar{\sigma}_1} < 1 \Leftarrow 8\alpha\mu s_r \bar{\sigma}_1 < \bar{\sigma}_{s_r} \Leftarrow \alpha \leq \frac{1}{512\alpha\mu s_r \kappa}.$$

Then, we get $u_i = \left(\mathcal{I} + \mathcal{E}_i \diamond_\Phi \mathcal{Z} + (\mathcal{E}_i \diamond_\Phi \mathcal{Z})^2 + \cdots\right) \diamond_\Phi \mathcal{E}_i \diamond_\Phi \mathcal{L} \diamond_\Phi u_i$ for each u_i which implies

$$\begin{aligned}
&\mathcal{U}_{k+1} \diamond_\Phi \Lambda \diamond_\Phi \mathcal{U}_{k+1}^H = \sum_{i=1}^{r} u_i \diamond_\Phi \lambda_i \diamond_\Phi u_i^H \\
=&\sum_{i=1}^{r} \left(\sum_{a \geq 0} (\mathcal{E}_i \diamond_\Phi \mathcal{Z})^a \diamond_\Phi \mathcal{E}_i \diamond_\Phi \mathcal{L} \right) \diamond_\Phi u_i \diamond_\Phi \lambda_i \diamond_\Phi u_i^H \left(\sum_{b \geq 0} (\mathcal{E}_i \diamond_\Phi \mathcal{Z})^b \diamond_\Phi \mathcal{E}_i \diamond_\Phi \mathcal{L} \right)^H \\
=&\sum_{a \geq 0} \mathcal{Z}^a \diamond_\Phi \mathcal{L} \diamond_\Phi \sum_{i=1}^{r} \left(\mathcal{E}_i^{a+1} \diamond_\Phi u_i \diamond_\Phi \lambda_i \diamond_\Phi u_i^H \diamond_\Phi \mathcal{E}_i^{b+1} \right) \diamond_\Phi \mathcal{L} \diamond_\Phi \sum_{b \geq 0} \mathcal{Z}^b
\end{aligned}$$

To simplify the above formula, we have

$$\begin{aligned}
\bar{\mathcal{E}}_i \bar{u}_i &= \begin{bmatrix} \frac{1}{\bar{\lambda}_i^{(1)}} I & 0 & \cdots & 0 \\ 0 & \frac{1}{\bar{\lambda}_i^{(2)}} I & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & \frac{1}{\bar{\lambda}_i^{(q)}} I \end{bmatrix} \begin{bmatrix} \bar{u}_i^{(1)} & 0 & \cdots & 0 \\ 0 & \bar{u}_i^{(2)} & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & \bar{u}_i^{(q)} \end{bmatrix} \\
&= \begin{bmatrix} \bar{u}_i^{(1)} & 0 & \cdots & 0 \\ 0 & \bar{u}_i^{(2)} & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & \bar{u}_i^{(q)} \end{bmatrix} \begin{bmatrix} \frac{1}{\bar{\lambda}_i^{(1)}} & 0 & \cdots & 0 \\ 0 & \frac{1}{\bar{\lambda}_i^{(2)}} & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & \frac{1}{\bar{\lambda}_i^{(q)}} \end{bmatrix} = \bar{u}_i \bar{\mathcal{E}}_i.
\end{aligned}$$

Then we further have

$$\begin{aligned}
&\bar{\mathcal{E}}_i^{a+1} \bar{u}_i \bar{\lambda}_i \bar{u}_i^H \bar{\mathcal{E}}_i^{b+1} \\
=&\bar{u}_i \begin{bmatrix} \left(\frac{1}{\bar{\lambda}_i^{(1)}}\right)^{a+1} & 0 & \cdots & 0 \\ 0 & \left(\frac{1}{\bar{\lambda}_i^{(2)}}\right)^{a+1} & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & \left(\frac{1}{\bar{\lambda}_i^{(q)}}\right)^{a+1} \end{bmatrix} \begin{bmatrix} \bar{\lambda}_i^{(1)} & 0 & \cdots & 0 \\ 0 & \bar{\lambda}_i^{(2)} & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & \bar{\lambda}_i^{(q)} \end{bmatrix}. \\
&\begin{bmatrix} \left(\frac{1}{\bar{\lambda}_i^{(1)}}\right)^{b+1} & 0 & \cdots & 0 \\ 0 & \left(\frac{1}{\bar{\lambda}_i^{(2)}}\right)^{b+1} & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & \left(\frac{1}{\bar{\lambda}_i^{(q)}}\right)^{b+1} \end{bmatrix} \bar{u}_i^H \\
=&\bar{u}_i \begin{bmatrix} \left(\frac{1}{\bar{\lambda}_i^{(1)}}\right)^{a+b+1} & 0 & \cdots & 0 \\ 0 & \left(\frac{1}{\bar{\lambda}_i^{(2)}}\right)^{a+b+1} & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & \left(\frac{1}{\bar{\lambda}_i^{(q)}}\right)^{a+b+1} \end{bmatrix} \bar{u}_i^H = \bar{u}_i \bar{\Gamma}_i^{(a+b+1)} \bar{u}_i^H,
\end{aligned}$$

where we introduce new notation $\bar{\Gamma}_i^{(a+b+1)}$. Hence we have $\mathcal{E}_i^{a+1} \diamond_\Phi u_i \diamond_\Phi \lambda_i \diamond_\Phi u_i^H \diamond_\Phi \mathcal{E}_i^{b+1} = u_i \diamond_\Phi \Gamma_i^{(a+b+1)} \diamond_\Phi u_i^H$. Now, we get

$$\begin{aligned}\mathcal{U}_{k+1} \diamond_\Phi \Lambda \diamond_\Phi \mathcal{U}_{k+1}^H &= \sum_{a\geq 0} \mathcal{Z}^a \diamond_\Phi \mathcal{L} \sum_{i=1}^{r} \left(\mathcal{E}_i^{a+1} \diamond_\Phi u_i \diamond_\Phi \lambda_i \diamond_\Phi u_i^H \diamond_\Phi \mathcal{E}_i^{b+1}\right) \diamond_\Phi \mathcal{L} \diamond_\Phi \sum_{b\geq 0} \mathcal{Z}^b \\ &= \sum_{a\geq 0} \mathcal{Z}^a \diamond_\Phi \mathcal{L} \diamond_\Phi \sum_{i=1}^{r} \left(u_i \diamond_\Phi \Gamma_i^{(a+b+1)} \diamond_\Phi u_i^H\right) \diamond_\Phi \mathcal{L} \diamond_\Phi \sum_{b\geq 0} \mathcal{Z}^b \\ &= \sum_{a,b\geq 0} \mathcal{Z}^a \diamond_\Phi \mathcal{L} \diamond_\Phi \mathcal{U}_{k+1} \diamond_\Phi \Gamma^{(a+b+1)} \diamond_\Phi \mathcal{U}_{k+1}^H \diamond_\Phi \mathcal{L} \diamond_\Phi \mathcal{Z}^b.\end{aligned}$$

Thus, we have

$$\begin{aligned}&\|\mathcal{L}_{k+1} - \mathcal{L}\|_\infty = \|\mathcal{U}_{k+1} \diamond_\Phi \Lambda \diamond_\Phi \mathcal{U}_{k+1}^H - \mathcal{L}\|_\infty \\ =&\|\mathcal{L} \diamond_\Phi \mathcal{U}_{k+1} \diamond_\Phi \Gamma^{(1)} \diamond_\Phi \mathcal{U}_{k+1}^H \diamond_\Phi \mathcal{L} - \mathcal{L} + \sum_{a+b>0} \mathcal{Z}^a \diamond_\Phi \mathcal{L} \diamond_\Phi \mathcal{U}_{k+1} \diamond_\Phi \Gamma^{(a+b+1)} \diamond_\Phi \mathcal{U}_{k+1}^H \diamond_\Phi \mathcal{L} \diamond_\Phi \mathcal{Z}^b\|_\infty \\ \leq&\|\mathcal{L} \diamond_\Phi \mathcal{U}_{k+1} \diamond_\Phi \Gamma^{(1)} \diamond_\Phi \mathcal{U}_{k+1}^H \diamond_\Phi \mathcal{L} - \mathcal{L}\|_\infty + \sum_{a+b>0} \|\mathcal{Z}^a \diamond_\Phi \mathcal{L} \diamond_\Phi \mathcal{U}_{k+1} \diamond_\Phi \Gamma^{(a+b+1)} \diamond_\Phi \mathcal{U}_{k+1}^H \diamond_\Phi \mathcal{L} \diamond_\Phi \mathcal{Z}^b\|_\infty \\ :=&Y_0 + \sum_{a+b>0} Y_{ab}.\end{aligned}$$

We will handle Y_0 first. Recall that $\mathcal{L} = \mathcal{U} \diamond_\Phi \Sigma \diamond_\Phi \mathcal{V}^H$ is the t-SVD of the symmetric tensor $\mathcal{L}$ which obeys μ-incoherence, i.e., $\mathcal{U} \diamond_\Phi \mathcal{U}^H = \mathcal{V} \diamond_\Phi \mathcal{V}^H$ and $\|\dot{e}_i^H \diamond_\Phi \mathcal{U} \diamond_\Phi \mathcal{U}^H\|_F \leq \sqrt{\frac{\mu s_r}{nq}}$ for all i. So for each (i,j,k) entry of Y_0, one has

$$\begin{aligned}Y_0 &= \max_{i,j,k} \left|\left\langle \mathcal{L} \diamond_\Phi \mathcal{U}_{k+1} \diamond_\Phi \Gamma^{(1)} \diamond_\Phi \mathcal{U}_{k+1}^H \diamond_\Phi \mathcal{L} - \mathcal{L}, \mathcal{E}_{ijk}\right\rangle\right| \\ &= \max_{i,j,k} \left|\left\langle \dot{e}_i^H \diamond_\Phi \mathcal{U} \diamond_\Phi \mathcal{U}^H (\mathcal{L} \diamond_\Phi \mathcal{U}_{k+1} \diamond_\Phi \Gamma^{(1)} \diamond_\Phi \mathcal{U}_{k+1}^H \diamond_\Phi \mathcal{L} - \mathcal{L}) \diamond_\Phi \mathcal{U} \diamond_\Phi \mathcal{U}^H \diamond_\Phi \dot{e}_j, e_k\right\rangle\right| \\ &\leq \max_{i,j} \|\dot{e}_i^H \diamond_\Phi \mathcal{U} \diamond_\Phi \mathcal{U}^H (\mathcal{L} \diamond_\Phi \mathcal{U}_{k+1} \diamond_\Phi \Gamma^{(1)} \diamond_\Phi \mathcal{U}_{k+1}^H \diamond_\Phi \mathcal{L} - \mathcal{L}) \diamond_\Phi \mathcal{U} \diamond_\Phi \mathcal{U}^H \diamond_\Phi \dot{e}_j\|_F \cdot \|e_k\|_F \\ &\leq \max_{i,j} \|\dot{e}_i^H \diamond_\Phi \mathcal{U} \diamond_\Phi \mathcal{U}^H\|_F \cdot \|\mathcal{L} \diamond_\Phi \mathcal{U}_{k+1} \diamond_\Phi \Gamma^{(1)} \diamond_\Phi \mathcal{U}_{k+1}^H \diamond_\Phi \mathcal{L} - \mathcal{L}\| \cdot \|\mathcal{U} \diamond_\Phi \mathcal{U}^H \diamond_\Phi \dot{e}_j\|_F \\ &\leq \frac{\mu s_r}{nq} \|\mathcal{L} \diamond_\Phi \mathcal{U}_{k+1} \diamond_\Phi \Gamma^{(1)} \diamond_\Phi \mathcal{U}_{k+1}^H \diamond_\Phi \mathcal{L} - \mathcal{L}\|,\end{aligned}$$

where the first inequality follows from the fact $\mathcal{U} \diamond_\Phi \mathcal{U}^H \diamond_\Phi \mathcal{L} = \mathcal{L} \diamond_\Phi \mathcal{U} \diamond_\Phi \mathcal{U}^H = \mathcal{L}$. Since $\mathcal{L} = \mathcal{U}_{k+1} \diamond_\Phi \Lambda \diamond_\Phi \mathcal{U}_{k+1}^H + \ddot{\mathcal{U}}_{k+1} \diamond_\Phi \ddot{\Lambda} \diamond_\Phi \ddot{\mathcal{U}}_{k+1}^H - \mathcal{Z}$, there holds that

$$\begin{aligned}&\|\mathcal{L} \diamond_\Phi \mathcal{U}_{k+1} \diamond_\Phi \Gamma^{(1)} \diamond_\Phi \mathcal{U}_{k+1}^H \diamond_\Phi \mathcal{L} - \mathcal{L}\| \\ =&\|(\mathcal{U}_{k+1} \diamond_\Phi \Lambda \diamond_\Phi \mathcal{U}_{k+1}^H + \ddot{\mathcal{U}}_{k+1} \diamond_\Phi \ddot{\Lambda} \diamond_\Phi \ddot{\mathcal{U}}_{k+1}^H - \mathcal{Z}) \diamond_\Phi \mathcal{U}_{k+1} \diamond_\Phi \Gamma^{(1)} \diamond_\Phi \mathcal{U}_{k+1}^H \\ &\diamond_\Phi (\mathcal{U}_{k+1} \diamond_\Phi \Lambda \diamond_\Phi \mathcal{U}_{k+1}^H + \ddot{\mathcal{U}}_{k+1} \diamond_\Phi \ddot{\Lambda} \diamond_\Phi \ddot{\mathcal{U}}_{k+1}^H - \mathcal{Z}) - \mathcal{L}\| \\ =&\|\mathcal{U}_{k+1} \diamond_\Phi \Lambda \diamond_\Phi \Gamma^{(1)} \diamond_\Phi \Lambda \diamond_\Phi \mathcal{U}_{k+1}^H - \mathcal{L} - \mathcal{U}_{k+1} \diamond_\Phi \Lambda \diamond_\Phi \Gamma^{(1)} \diamond_\Phi \mathcal{U}_{k+1}^H \diamond_\Phi \mathcal{Z} \\ &- \mathcal{Z} \diamond_\Phi \mathcal{U}_{k+1} \diamond_\Phi \Gamma^{(1)} \diamond_\Phi \Lambda \diamond_\Phi \mathcal{U}_{k+1}^H + \mathcal{Z} \diamond_\Phi \mathcal{U}_{k+1} \diamond_\Phi \Gamma^{(1)} \diamond_\Phi \mathcal{U}_{k+1}^H \diamond_\Phi \mathcal{Z}\| \\ \leq&\|\mathcal{Z} - \ddot{\mathcal{U}}_{k+1} \diamond_\Phi \ddot{\Lambda} \diamond_\Phi \ddot{\mathcal{U}}_{k+1}^H\| + \|\mathcal{U}_{k+1} \diamond_\Phi \Lambda \diamond_\Phi \Gamma^{(1)} \diamond_\Phi \mathcal{U}_{k+1}^H \diamond_\Phi \mathcal{Z}\| \\ &+ \|\mathcal{Z} \diamond_\Phi \mathcal{U}_{k+1} \diamond_\Phi \Gamma^{(1)} \diamond_\Phi \Lambda \diamond_\Phi \mathcal{U}_{k+1}^H\| + \|\mathcal{Z} \diamond_\Phi \mathcal{U}_{k+1} \diamond_\Phi \Gamma^{(1)} \diamond_\Phi \mathcal{U}_{k+1}^H \diamond_\Phi \mathcal{Z}\| \\ =&\|\bar{\mathcal{Z}} - \bar{\ddot{\mathcal{U}}}_{k+1} \bar{\ddot{\Lambda}} \bar{\ddot{\mathcal{U}}}_{k+1}^H\| + \|\bar{\mathcal{U}}_{k+1} \bar{\Lambda} \bar{\Gamma}^{(1)} \bar{\mathcal{U}}_{k+1}^H \bar{\mathcal{Z}}\| + \|\bar{\mathcal{Z}} \bar{\mathcal{U}}_{k+1} \bar{\Gamma}^{(1)} \bar{\Lambda} \bar{\mathcal{U}}_{k+1}^H\| + \|\bar{\mathcal{Z}} \bar{\mathcal{U}}_{k+1} \bar{\Gamma}^{(1)} \bar{\mathcal{U}}_{k+1}^H \bar{\mathcal{Z}}\| \\ \leq&\|\bar{\mathcal{Z}} - \bar{\ddot{\mathcal{U}}}_{k+1} \bar{\ddot{\Lambda}} \bar{\ddot{\mathcal{U}}}_{k+1}^H\| + 2\|\bar{\mathcal{Z}}\| + \frac{\|Z\|^2}{|\bar{\lambda}_{s_r}|} \leq \|\bar{\ddot{\mathcal{U}}}_{k+1} \bar{\ddot{\Lambda}} \bar{\ddot{\mathcal{U}}}_{k+1}^H\| + 4\|\mathcal{Z}\| \leq |\bar{\lambda}_{s_r+1}| + 4\|\mathcal{Z}\| \leq 5\|\mathcal{Z}\|,\end{aligned}$$

where the last inequality follows from the fact $\frac{\|\mathcal{Z}\|}{|\bar{\lambda}_{s_r}|} \leq \frac{1}{7}$ and $|\bar{\lambda}_{s_r+1}| \leq \|\mathcal{Z}\|$ according to $\bar{\sigma}_{s_r+1} = 0$ and the tensor Weyl's inequality. Thus we have

$$Y_0 \leq \frac{5\mu s_r}{nq}\|\mathcal{Z}\|.$$

Next, we derive an upper bound for the rest part. Note that

$$\begin{aligned}
Y_{ab} =& \max_{i,j,k} \left\langle \mathcal{Z}^a \diamond_\Phi \mathcal{L} \diamond_\Phi \mathcal{U}_{k+1} \diamond_\Phi \Gamma^{(a+b+1)} \diamond_\Phi \mathcal{U}_{k+1}^H \diamond_\Phi \mathcal{L} \diamond_\Phi \mathcal{Z}^b, \mathcal{E}_{ijk} \right\rangle \\
=& \max_{i,j,k} \left\langle (\dot{e}_i^H \diamond_\Phi \mathcal{Z}^a \diamond_\Phi \mathcal{U} \diamond_\Phi \mathcal{U}^H) \diamond_\Phi \mathcal{L} \diamond_\Phi \mathcal{U}_{k+1} \diamond_\Phi \Gamma^{(a+b+1)} \diamond_\Phi \mathcal{U}_{k+1}^H \diamond_\Phi \mathcal{L} \diamond_\Phi (\mathcal{U} \diamond_\Phi \mathcal{U}^H \diamond_\Phi \mathcal{Z}^b \diamond_\Phi \dot{e}_j), e_k \right\rangle \\
\leq& \max_{i,j} \|(\dot{e}_i^H \diamond_\Phi \mathcal{Z}^a \diamond_\Phi \mathcal{U} \diamond_\Phi \mathcal{U}^H) \diamond_\Phi \mathcal{L} \diamond_\Phi \mathcal{U}_{k+1} \diamond_\Phi \Gamma^{(a+b+1)} \diamond_\Phi \mathcal{U}_{k+1}^H \diamond_\Phi \mathcal{L} \diamond_\Phi (\mathcal{U} \diamond_\Phi \mathcal{U}^H \diamond_\Phi \mathcal{Z}^b \diamond_\Phi \dot{e}_j)\|_F \cdot \|e_k\|_F \\
\leq& \max_{i,j} \|\dot{e}_i^H \diamond_\Phi \mathcal{Z}^a \diamond_\Phi \mathcal{U}\|_F \cdot \|\mathcal{L} \diamond_\Phi \mathcal{U}_{k+1} \diamond_\Phi \Gamma^{(a+b+1)} \diamond_\Phi \mathcal{U}_{k+1}^H \diamond_\Phi \mathcal{L}\| \cdot \|\mathcal{U}^H \diamond_\Phi \mathcal{Z}^b \diamond_\Phi \dot{e}_j\|_F \\
\leq& \max_l \frac{\mu s_r}{nq} (\alpha n q \sqrt{q} \|\mathcal{Z}\|_\infty)^{a+b} \|\mathcal{L} \diamond_\Phi \mathcal{U}_{k+1} \diamond_\Phi \Gamma^{(a+b+1)} \diamond_\Phi \mathcal{U}_{k+1}^H \diamond_\Phi \mathcal{L}\|,
\end{aligned}$$

where last inequality have used the bound of α. Furthermore, by using $\mathcal{L} = \mathcal{U}_{k+1} \diamond_\Phi \Lambda \diamond_\Phi \mathcal{U}_{k+1}^H + \ddot{\mathcal{U}}_{k+1} \diamond_\Phi \ddot{\Lambda} \diamond_\Phi \ddot{\mathcal{U}}_{k+1}^H - \mathcal{Z}$ again, we get

$$\begin{aligned}
& \|\mathcal{L} \diamond_\Phi \mathcal{U}_{k+1} \diamond_\Phi \Gamma^{(a+b+1)} \diamond_\Phi \mathcal{U}_{k+1}^H \diamond_\Phi \mathcal{L}\| \\
=& \|(\mathcal{U}_{k+1} \diamond_\Phi \Lambda \diamond_\Phi \mathcal{U}_{k+1}^H + \ddot{\mathcal{U}}_{k+1} \diamond_\Phi \ddot{\Lambda} \diamond_\Phi \ddot{\mathcal{U}}_{k+1}^H - \mathcal{Z}) \diamond_\Phi \mathcal{U}_{k+1} \diamond_\Phi \Gamma^{(a+b+1)} \diamond_\Phi \mathcal{U}_{k+1}^H \diamond_\Phi (\mathcal{U}_{k+1} \diamond_\Phi \Lambda \diamond_\Phi \mathcal{U}_{k+1}^H \\
& + \ddot{\mathcal{U}}_{k+1} \diamond_\Phi \ddot{\Lambda} \diamond_\Phi \ddot{\mathcal{U}}_{k+1}^H - \mathcal{Z})\| \\
=& \|\mathcal{U}_{k+1} \diamond_\Phi \Lambda \diamond_\Phi \Gamma^{(a+b+1)} \diamond_\Phi \Lambda \diamond_\Phi \mathcal{U}_{k+1}^H - \mathcal{U}_{k+1} \diamond_\Phi \Lambda \diamond_\Phi \Gamma^{(a+b+1)} \diamond_\Phi \mathcal{U}_{k+1}^H \diamond_\Phi \mathcal{Z} \\
& - \mathcal{Z} \diamond_\Phi \mathcal{U}_{k+1} \diamond_\Phi \Gamma^{(a+b+1)} \diamond_\Phi \Lambda \diamond_\Phi \mathcal{U}_{k+1}^H + \mathcal{Z} \diamond_\Phi \mathcal{U}_{k+1} \diamond_\Phi \Gamma^{(a+b+1)} \diamond_\Phi \mathcal{U}_{k+1}^H \diamond_\Phi \mathcal{Z}\| \\
\leq& \|\bar{\mathcal{U}}_{k+1} \bar{\Lambda} \bar{\Gamma}^{(a+b+1)} \bar{\Lambda} \bar{\mathcal{U}}_{k+1}^H - \bar{\mathcal{U}}_{k+1} \bar{\Lambda} \bar{\Gamma}^{(a+b+1)} \bar{\mathcal{U}}_{k+1}^H \bar{\mathcal{Z}} - \bar{\mathcal{Z}} \bar{\mathcal{U}}_{k+1} \bar{\Gamma}^{(a+b+1)} \bar{\Lambda} \bar{\mathcal{U}}_{k+1}^H + \bar{\mathcal{Z}} \bar{\mathcal{U}}_{k+1} \bar{\Gamma}^{(a+b+1)} \bar{\mathcal{U}}_{k+1}^H \bar{\mathcal{Z}}\| \\
\leq& |\bar{\lambda}_{s_r}|^{-(a+b-1)} + |\bar{\lambda}_{s_r}|^{-(a+b)} \|\mathcal{Z}\| + |\bar{\lambda}_{s_r}|^{-(a+b)} \|\mathcal{Z}\| + |\bar{\lambda}_{s_r}|^{-(a+b+1)} \|\mathcal{Z}\|^2 \\
=& |\bar{\lambda}_{s_r}|^{-(a+b-1)} \left(1 + \frac{2\|\mathcal{Z}\|}{\bar{\lambda}_{s_r}} + \left(\frac{\|\mathcal{Z}\|}{|\bar{\lambda}_{s_r}|}\right)^2\right) = |\bar{\lambda}_{s_r}|^{-(a+b-1)} \left(1 + \frac{\|\mathcal{Z}\|}{|\bar{\lambda}_{s_r}|}\right)^2 \leq 2|\bar{\lambda}_{s_r}|^{-(a+b-1)}.
\end{aligned}$$

Together, we have $Y_{ab} \leq 2\alpha\mu s_r \sqrt{q} \|\mathcal{Z}\|_\infty \left(\frac{\alpha n q \sqrt{q} \|\mathcal{Z}\|_\infty}{|\bar{\lambda}_{s_r}|}\right)^{a+b-1}$. Then, we have

$$\begin{aligned}
\sum_{a+b>0} Y_{ab} \leq& \sum_{a+b>0} 2\alpha\mu s_r \sqrt{q} \|\mathcal{Z}\|_\infty \left(\frac{4\alpha\mu s_r \sqrt{q} \bar{\sigma}_1}{|\bar{\lambda}_{s_r}|}\right)^{a+b-1} \leq 2\alpha\mu s_r \sqrt{q} \|\mathcal{Z}\|_\infty \sum_{a+b>0} \left(\frac{4}{508}\right)^{a+b-1} \\
\leq& 2\alpha\mu s_r \sqrt{q} \|\mathcal{Z}\|_\infty \left(\frac{1}{1 - \frac{4}{508}}\right)^2 \leq 3\alpha\mu s_r \sqrt{q} \|\mathcal{Z}\|_\infty.
\end{aligned}$$

Finally, combining them together gives

$$\|\mathcal{L}_{k+1} - \mathcal{L}\|_\infty = Y_0 + \sum_{a+b>0} Y_{ab} \leq \frac{5\mu s_r}{nq} \|\mathcal{Z}\| + 3\alpha\mu s_r \sqrt{q} \|\mathcal{Z}\|_\infty \leq 8\alpha\mu s_r \sqrt{q} \|\mathcal{Z}\|_\infty \leq \frac{\mu s_r}{nq} \gamma^k \bar{\sigma}_1,$$

where the last inequality uses the bound of α in Theorem 5.4. □

Lemma D.23. *Suppose $\|\mathcal{L} - \mathcal{L}_{k+1}\|_\infty \leq \frac{\mu s_r}{nq} \gamma^k \bar{\sigma}_1$, then we have (1). supp$(\mathcal{S} - \mathcal{S}_{k+1}) \subset$ supp$(\mathcal{S})$, (2). $\|\mathcal{S} - \mathcal{S}_{k+1}\|_\infty \leq \frac{4\mu s_r}{nq} \gamma^k \bar{\sigma}_1$.*

Proof. We shall prove the first conclusion. Recall that $\mathcal{S}_{k+1} = T_\zeta(\mathcal{D} - \mathcal{L}_{k+1}) = T_\zeta(\mathcal{L} - \mathcal{L}_{k+1} - \mathcal{S})$ where $\zeta = \frac{2\mu s_r}{nq} \gamma^k \bar{\lambda}_1$. If $\mathcal{S}_{ijt} = 0$, then $(\mathcal{S} - \mathcal{S}_{k+1})_{ijt} = (\mathcal{L} - \mathcal{L}_{k+1})_{ijt}$ when $|(\mathcal{L} - \mathcal{L}_{k+1})_{ijt}| > \zeta$. The first part of lemma is proved by using assumption that $\|\mathcal{L}^{t+1} - \mathcal{L}\|_\infty \leq \frac{\mu s_r}{nq} \gamma^k \bar{\sigma}_1 \leq \frac{2\mu s_r}{nq} \gamma^k \bar{\lambda}_1 = \zeta$.

We now come to the second conclusion

$$[\mathcal{S} - \mathcal{S}_{k+1}]_{ijk} = \begin{cases} 0 \\ [\mathcal{L}_{k+1} - \mathcal{L}]_{ijk} \\ [\mathcal{S}]_{ijk} \end{cases} \leq \begin{cases} 0 \\ \|\mathcal{L} - \mathcal{L}_{k+1}\|_\infty \\ \|\mathcal{L} - \mathcal{L}_{k+1}\|_\infty + \zeta \end{cases} \leq \begin{cases} 0 & (i,j,k) \in \Omega^c \\ \frac{\mu s_r}{nq} \gamma^k \bar{\sigma}_1 & (i,j,k) \in \Omega_{k+1} \\ \frac{4\mu s_r}{nq} \gamma^k \bar{\sigma}_1 & (i,j,k) \in \Omega \backslash \Omega_{k+1} \end{cases}.$$

where $\Omega :=$ supp$(\mathcal{S})$ and $\Omega_{k+1} :=$ supp$(\mathcal{S}_{k+1})$. So we have $\|\mathcal{S} - \mathcal{S}_{k+1}\|_\infty \leq \frac{4\mu s_r}{nq} \gamma^k \bar{\sigma}_1$. □

Theorem D.24. *With the condition of Theorem 5.4, let $\mathcal{L} \in \mathbb{R}^{n\times n\times q}$ and $\mathcal{S} \in \mathbb{R}^{n\times n\times q}$ be two symmetric tensors satisfying Assumptions 5.1 and 5.2. If the initial guess $\mathcal{L}_1$ and $\mathcal{S}_1$ obey the following conditions:*

$$\|\mathcal{L}-\mathcal{L}_1\|_\infty \leq \frac{\mu s_r}{nq}\bar{\sigma}_1,\ \|\mathcal{S}-\mathcal{S}_1\|_\infty \leq \frac{4\mu s_r}{nq}\bar{\sigma}_1,\ \textit{and supp}(\mathcal{S}_1)\subset\Omega,$$

then the iterates of Algorithm 1 with parameters $\beta = \frac{\mu s_r}{2nq}$ and $\gamma \in (\frac{1}{16}, 1)$ satisfy

$$\|\mathcal{L}-\mathcal{L}_{k+1}\|_\infty \leq \frac{\mu s_r}{nq}\gamma^k\bar{\sigma}_1,\ \|\mathcal{S}-\mathcal{S}_{k+1}\|_\infty \leq \frac{4\mu s_r}{nq}\gamma^k\bar{\sigma}_1,\ \textit{and supp}(\mathcal{S}_k)\subset\Omega.$$

Proof. This theorem will be proved by mathematical induction.
Base Case: When $k = 1$, we use the initialization in our efficient implementation and the condition is satisfied.
Induction Step: Assume we have

$$\|\mathcal{L}-\mathcal{L}_{k}\|_\infty \leq \frac{\mu s_r}{nq}\gamma^{k-1}\bar{\sigma}_1,\ \|\mathcal{S}-\mathcal{S}_{k}\|_\infty \leq \frac{4\mu s_r}{nq}\gamma^{k-1}\bar{\sigma}_1,\ \text{and supp}(\mathcal{S}_k)\subset\Omega.$$

at the k^{th} iteration. At the $(k+1)^{th}$ iteration. It follows directly from Lemmas D.22 and D.23 that

$$\|\mathcal{L}-\mathcal{L}_{k+1}\|_\infty \leq \frac{\mu s_r}{nq}\gamma^k\bar{\sigma}_1,\ \|\mathcal{S}-\mathcal{S}_{k+1}\|_\infty \leq \frac{4\mu s_r}{nq}\gamma^k\bar{\sigma}_1,\ \text{and supp}(\mathcal{S}_k)\subset\Omega.$$

□

D.2.5. Proof of Proposition 5.3

Proof. The proof of Proposition 5.3 directly follows from Theorem D.6. □

D.2.6. Proof of Theorem 5.4

Proof. The initialization $\mathcal{L}_1$ and $\mathcal{S}_1$ by Algorithm 1 satisfies

$$\|\mathcal{L}-\mathcal{L}_1\|_\infty \leq \frac{\mu s_r}{4nq}\bar{\sigma}_1 \leq \frac{\mu s_r}{nq}\bar{\sigma}_1,\ \|\mathcal{S}-\mathcal{S}_1\|_\infty \leq \frac{\mu s_r}{nq}\bar{\sigma}_1 \leq \frac{4\mu s_r}{nq}\bar{\sigma}_1,\quad \text{and supp}(\mathcal{S}_1)\subset\Omega. \tag{19}$$

By Theorem D.24, we have

$$\|\mathcal{L}-\mathcal{L}_{k}\|_\infty \leq \frac{\mu s_r}{nq}\gamma^{k-1}\bar{\sigma}_1,\ \|\mathcal{S}-\mathcal{S}_{k}\|_\infty \leq \frac{4\mu s_r}{nq}\gamma^{k-1}\bar{\sigma}_1,\ \text{and supp}(\mathcal{S}_k)\subset\Omega.$$

To get the spectral norm bound of $\mathcal{L}-\mathcal{L}_k$, we have

$$\begin{aligned}&\|\mathcal{L}-\mathcal{L}_k\| = \|\mathcal{L}-(\mathcal{D}-\mathcal{S}_{k-1})+(\mathcal{D}-\mathcal{S}_{k-1})-\mathcal{L}_k\| \leq \|\mathcal{L}-(\mathcal{D}-\mathcal{S}_{k-1})\| + \|(\mathcal{D}-\mathcal{S}_{k-1})-\mathcal{L}_k\|\\ \leq& 2\|\mathcal{L}-(\mathcal{D}-\mathcal{S}_{k-1})\| \leq 2\|\mathcal{S}-\mathcal{S}_{k-1}\| \leq 2\alpha nq\|\mathcal{S}-\mathcal{S}_{k-1}\|_\infty \leq 8\alpha\mu s_r\gamma^{k-2}\bar{\sigma}_1\end{aligned}$$

By setting $\epsilon = \mu s_r\gamma^{k-1}\bar{\sigma}_1$, we prove this theorem. □

D.2.7. Proof of Theorem 5.5

Proof. The proof of Theorem 5.5 can be directly follows from Theorem D.20 by setting $\mu s_r\bar{\sigma}_1$ as ϵ. □