Supplementary Material of Tensor Low-rank Representation for Data Recovery and Clustering

Pan Zhou, Canyi Lu, *Member, IEEE,* Jiashi Feng, Zhouchen Lin, *Fellow, IEEE,* Shuicheng Yan, *Fellow, IEEE*

1 Structure of This Document

This document is organized as follows. Firstly, it provides more experimental results on the image denoising task in Sec. 2. Then in Sec. 3 it elaborates on the optimization procedure and details of Algorithm 2 in the manuscript. Next, this document provides the proofs of Theorems 2 $\sim$ 5. Specifically, Sec. 4 introduces some additional necessary notations, definitions and properties used in this document. Sec. 5 presents the proofs of Theorem 2. Sec. 6 presents the proofs of Theorem 3. In Sec. 7, we prove Theorem 4. At last, Sec. 8 introduces the proofs of Theorem 5.

2 More Experimental Results

Here we first provide more information about the Berkeley image segmentation dataset[1] and YUV video sequences[2]. The Berkeley segmentation dataset contains total 200 color images of various natural scenes and their sizes are $321 \times 481 \times 3$. The YUV video sequences contain 26 videos whose categories, frame sizes and frame numbers are reported in Table 4.

Then we report the PSNR values on the total 200 testing images in Berkeley segmentation dataset for the case that we randomly set 20% of pixels to random values in $[0, 255]$ for each testing image. Fig. 16 summaries these PSNR values on the total 200 images. One can observe that our R-TLRR always outperforms others. Concretely, it makes at least 2.0, 1.5, 1.0 and 0.5 dBs improvement than the second best on 41, 105, 174, and 194 images, respectively. Notice, there are only two images on which our method cannot achieve the highest PSNR values. So our R-TLRR can achieve the best denoising performance on each images in most cases. These results verify the advantages and robustness of our method and are consistent with the denoising results in manuscript.

Finally, in Fig. 17 we provide more denoising results with the PSNR values when the noise ratio is 20%. As mentioned in the manuscript, R-TLRR can preserve more details and performs much better the other compared methods. E.g., it can well preserves the jaws of penuins and the stripes of the tiger, *etc.* For numerical results, R-TLRR improves by at least 2.4 dB over the second best R-TPCA on the six testing images. All these results are consistent with the denoising results in the manuscript. All results show the superiority of the proposed TLRR.

1. https://www.eecs.berkeley.edu/Research/Projects/CS/vision/bsds/
2. http://trace.eas.asu.edu/yuv/

TABLE 4: Description of the YUV video sequences. As for video size $m \times n \times l$, $m \times n$ is frame size and l is frame number.

ID	Video name	Video size	ID	Video name	Video size
1	Akiyo	$176 \times 144 \times 100$	14	Hall	$176 \times 144 \times 100$
2	Big Buck Bunny	$352 \times 288 \times 100$	15	Highway	$176 \times 144 \times 100$
3	Bridge (close)	$176 \times 144 \times 100$	16	Miss America	$176 \times 144 \times 100$
4	Bridge (far)	$176 \times 144 \times 100$	17	Mobile	$176 \times 144 \times 100$
5	Bus	$352 \times 288 \times 100$	18	Mother Daughter	$176 \times 144 \times 100$
6	Carphone	$176 \times 144 \times 100$	19	News	$176 \times 144 \times 100$
7	Claire	$176 \times 144 \times 100$	20	Paris	$352 \times 288 \times 100$
8	Coastguard	$176 \times 144 \times 100$	21	Salesman	$176 \times 144 \times 100$
9	Container	$176 \times 144 \times 100$	22	Silent	$176 \times 144 \times 100$
10	Elephants Dream	$352 \times 288 \times 100$	23	Stefan	$352 \times 288 \times 90$
11	Flower	$176 \times 144 \times 100$	24	Suzie	$176 \times 144 \times 100$
12	Foreman	$176 \times 144 \times 100$	25	Tempete	$352 \times 288 \times 100$
13	Grandma	$176 \times 144 \times 100$	26	Waterfall	$352 \times 288 \times 100$

3 Optimization Details of Algorithm 2

Here we elaborate on the optimization to problem (14) in the manuscript, namely the following problem:

$$\min_{\mathcal{J},\mathcal{Z}',\mathcal{E}} \|\mathcal{Z}'\|_* + \lambda\|\mathcal{E}\|_1, \ \text{s.t.}\ \mathcal{Z}' = \mathcal{J}, \mathcal{X} = \mathcal{D} * \mathcal{J} + \mathcal{E}. \tag{18}$$

To tackle the hard constraints, we resort to augmented Lagrangian multiplier method and solve the following problem instead:

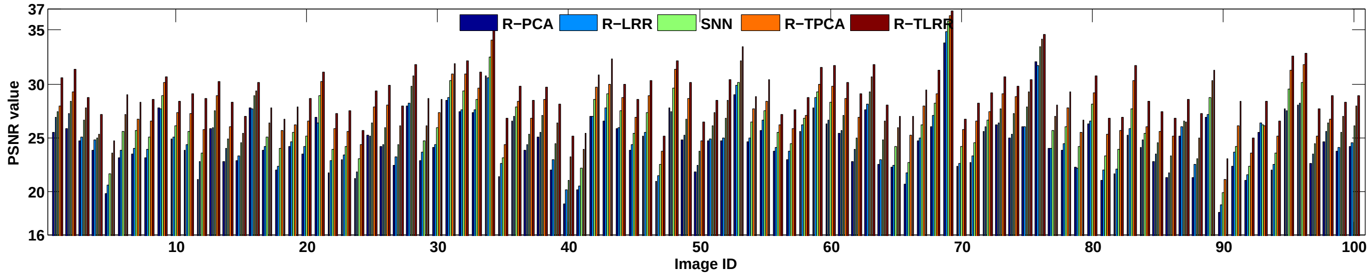
$$H(\mathcal{J},\mathcal{Z}',\mathcal{E},\mathcal{Y}^1,\mathcal{Y}^2) = \|\mathcal{Z}'\|_* + \lambda\|\mathcal{E}\|_1 + \left\langle \mathcal{Y}^1, \mathcal{Z}' - \mathcal{J} \right\rangle + \frac{\beta}{2}\|\mathcal{Z}' - \mathcal{J}\|_F^2 + \left\langle \mathcal{Y}^2, \mathcal{X} - \mathcal{D}*\mathcal{J} - \mathcal{E} \right\rangle + \frac{\beta}{2}\|\mathcal{X} - \mathcal{D}*\mathcal{J} - \mathcal{E}\|_F^2,$$

where $\mathcal{Y}^1$ and $\mathcal{Y}^2$ are the Lagrange multipliers introduced for the two constraints respectively, and β is an auto-adjusted penalty parameter. Then we solve the problem through alternately updating $\mathcal{J}$, $\mathcal{Z}'$ and $\mathcal{E}$ in each iteration to minimize $H(\mathcal{J},\mathcal{Z}',\mathcal{E},\mathcal{Y}_1,\mathcal{Y}_2)$ with other variables fixed. Details on the update of each variable are provided as follows.

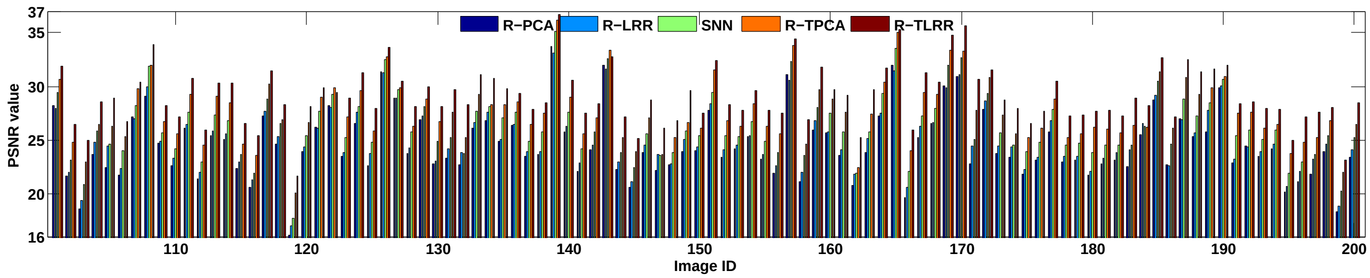
Updating $\mathcal{J}$: We minimize the following problem

$$\begin{aligned} \mathcal{J}_{k+1} &= \operatorname*{argmin}_{\mathcal{J}} \left\| \mathcal{Q}_k^1 - \mathcal{J} \right\|_F^2 + \left\| \mathcal{Q}_k^2 - \mathcal{D} * \mathcal{J} \right\|_F^2 \\ &= (\mathcal{D}^* * \mathcal{D} + \mathcal{I})^{-1} * \left(\mathcal{Q}_k^1 + \mathcal{D}^* * \mathcal{Q}_k^2 \right), \end{aligned} \tag{19}$$

where $\mathcal{Q}_k^1 = \mathcal{Z}'_k + \mathcal{Y}_k^1/\beta_k$ and $\mathcal{Q}_k^2 = \mathcal{X} - \mathcal{E}_k + \mathcal{Y}_k^2/\beta_k$.

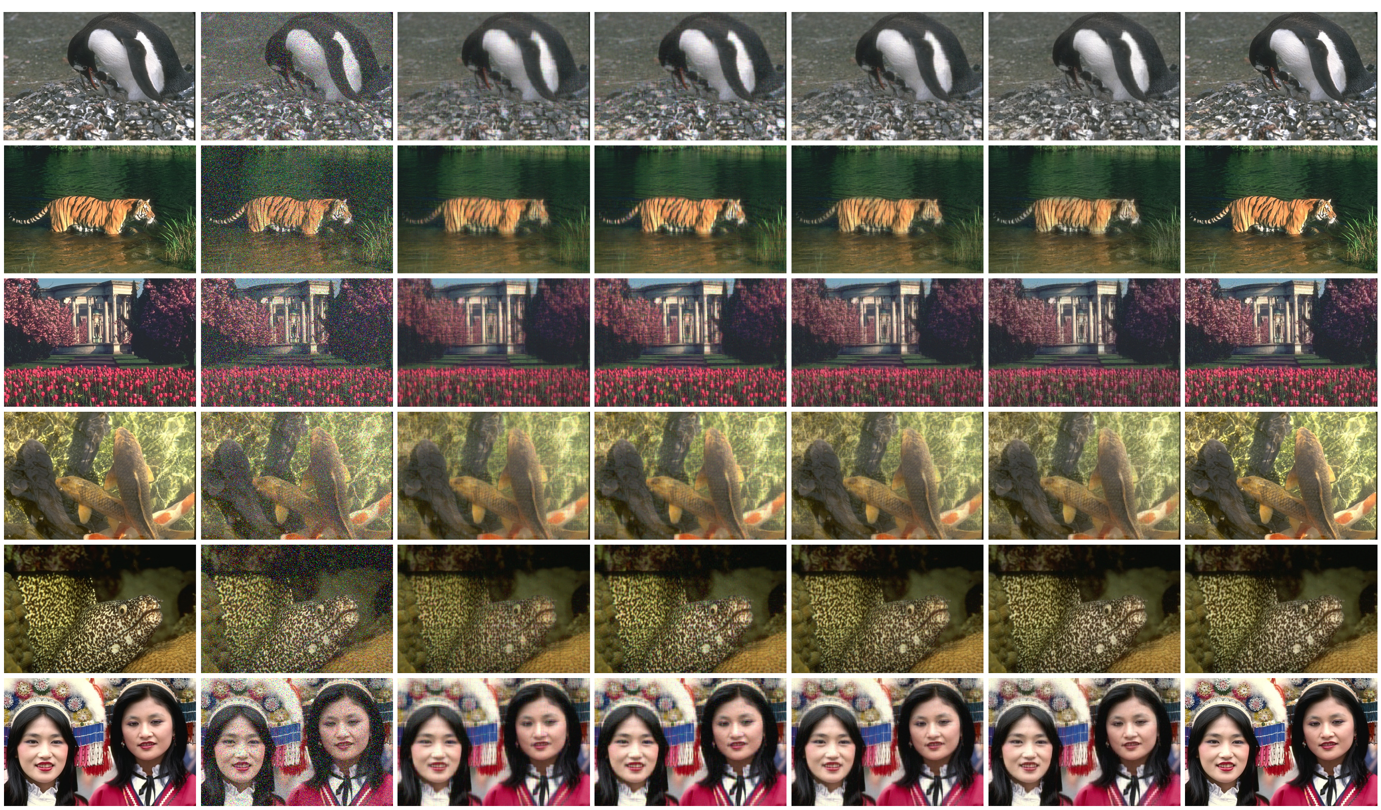


(a)



(b)

Fig. 16: Comparison of the PSNR values achieved by the compared methods on the 200 images in the Berkeley segmentation dataset. We randomly set 20% of pixels to random values in each image. (a) The PSNR values on first 100 images. (b) The PSNR values on the remaining 100 images. **Best viewed in $\times 2$ sized color pdf file.**



(a) Original image (b) Corrupted image (c) R-PCA (d) R-LRR (e) SNN (f) R-TPCA (g) R-TLRR

Image	R-PCA	R-LRR	SNN	R-TPCA	R-TLRR	Image	R-PCA	R-LRR	SNN	R-TPCA	R-TLRR	Image	R-PCA	R-LRR	SNN	R-TPCA	R-TLRR
Penuins	25.6	27.0	27.5	27.9	**30.5**	Tiger	22.4	24.5	24.5	26.3	**28.9**	Flowers	20.8	21.8	22.0	22.5	**25.2**
Fishes	21.0	22.7	23.5	25.6	**28.6**	Moray	21.0	22.7	23.5	25.6	**28.6**	Japanese man	22.2	23.7	23.6	23.7	**26.1**

(h) PSNR values achieved by the compared methods on the above six images.

Fig. 17: Examples of image denoising. We randomly set 20% of pixels to random values in each image. **Best viewed in $\times 2$ sized color pdf file.**

The property $\|\boldsymbol{\mathcal{B}}\|_F^2 = \frac{1}{n_3}\|\bar{\boldsymbol{B}}\|_F^2$ in Lemma 1 and the block-diagonal structure of $\bar{\boldsymbol{B}}$ imply that solving problem (17) is equivalent to computing $\bar{\boldsymbol{\mathcal{J}}}^{k+1}$ first:

$$\bar{\boldsymbol{J}}_{k+1}^{(i)} = \bar{\boldsymbol{G}}^{(i)}\left((\bar{\boldsymbol{Q}}_k^1)^{(i)} + (\bar{\boldsymbol{D}}^{(i)})^*(\bar{\boldsymbol{Q}}_k^2)^{(i)}\right), \ \forall i = 1, \cdots, n_3, \tag{20}$$

where $\boldsymbol{\mathcal{G}} = (\boldsymbol{\mathcal{D}}^* * \boldsymbol{\mathcal{D}} + \boldsymbol{\mathcal{I}})^{-1}$. Then we can obtain $\boldsymbol{\mathcal{J}}_{k+1} = \texttt{ifft}(\bar{\boldsymbol{\mathcal{J}}}_{k+1}, [], 3)$.

Updating the block $(\boldsymbol{\mathcal{Z}}', \boldsymbol{\mathcal{E}})$**:** For solving $\boldsymbol{\mathcal{Z}}'$ and $\boldsymbol{\mathcal{E}}$, we put $(\boldsymbol{\mathcal{Z}}', \boldsymbol{\mathcal{E}})$ into a large block of variables.

$$\begin{aligned}(\boldsymbol{\mathcal{Z}}'_{k+1}, \boldsymbol{\mathcal{E}}_{k+1}) = \operatorname*{argmin}_{\boldsymbol{\mathcal{Z}}', \boldsymbol{\mathcal{E}}} \|\boldsymbol{\mathcal{Z}}'\|_* + \lambda\|\boldsymbol{\mathcal{E}}\|_1 + \frac{\beta_k}{2}\left\|\boldsymbol{\mathcal{Z}}' - \boldsymbol{\mathcal{R}}_k^1\right\|_F^2 \\ + \frac{\beta_k}{2}\left\|\boldsymbol{\mathcal{R}}_k^2 - \boldsymbol{\mathcal{E}}\right\|_F^2,\end{aligned} \tag{21}$$

where $\boldsymbol{\mathcal{R}}_k^1 = \boldsymbol{\mathcal{J}}_{k+1} - \boldsymbol{\mathcal{Y}}_k^1/\beta_k$ and $\boldsymbol{\mathcal{R}}_k^2 = \boldsymbol{\mathcal{X}} - \boldsymbol{\mathcal{D}} * \boldsymbol{\mathcal{J}}_{k+1} + \boldsymbol{\mathcal{Y}}_k^2/\beta_k$. Problem (19) can be split into subproblems for $\boldsymbol{\mathcal{Z}}'$ and $\boldsymbol{\mathcal{E}}$ as these two variables are independent in this minimization problem. Accordingly, we update the variable $\boldsymbol{\mathcal{Z}}'$ as follows:

$$\boldsymbol{\mathcal{Z}}'_{k+1} = \operatorname*{argmin}_{\boldsymbol{\mathcal{Z}}'} \|\boldsymbol{\mathcal{Z}}'\|_* + \frac{\beta_k}{2}\left\|\boldsymbol{\mathcal{Z}}' - \boldsymbol{\mathcal{R}}_k^1\right\|_F^2.$$

Thus, by the properties $\|\boldsymbol{\mathcal{B}}\|_F^2 = \frac{1}{n_3}\|\bar{\boldsymbol{B}}\|_F^2$ and $\|\boldsymbol{\mathcal{B}}\|_* = \frac{1}{n_3}\|\bar{\boldsymbol{B}}\|_*$ in Lemma 1 and Definition 4, respectively, we can optimize its equivalent problem:

$$\bar{\boldsymbol{Z}}'_{k+1} = \operatorname*{argmin}_{\bar{\boldsymbol{Z}}'} \frac{1}{n_3}\left(\|\bar{\boldsymbol{Z}}'\|_* + \frac{\beta_k}{2}\left\|\bar{\boldsymbol{R}}_k^1 - \bar{\boldsymbol{Z}}'\right\|_F^2\right).$$

Since $\bar{\boldsymbol{Z}}'$ is a block-diagonal matrix, we only need to update all the block matrices $(\bar{\boldsymbol{Z}}')^{(i)}$ $(i = 1, \cdots, n_3)$ along the diagonal by following closed-form solution:

$$(\bar{\boldsymbol{Z}}'_{k+1})^{(i)} = \mathcal{S}_{\frac{1}{\beta_k}}\left((\bar{\boldsymbol{R}}_k^1)^{(i)}\right), \ i = 1, \cdots, n_3. \tag{22}$$

Here $\mathcal{S}_{1/\beta_k}(\cdot)$ is the singular value thresholding (SVT) operator [1]. Finally, we can compute $\boldsymbol{\mathcal{Z}}'_{k+1} = \texttt{ifft}(\bar{\boldsymbol{\mathcal{Z}}}'_{k+1}, [\,], 3)$. As for $\boldsymbol{\mathcal{E}}$, we can update it by solving

$$\boldsymbol{\mathcal{E}}_{k+1} = \operatorname*{argmin}_{\boldsymbol{\mathcal{E}}} \lambda\|\boldsymbol{\mathcal{E}}\|_1 + \frac{\beta_k}{2}\left\|\boldsymbol{\mathcal{R}}_k^2 - \boldsymbol{\mathcal{E}}\right\|_F^2.$$

Hence, we can obtain its closed-form solution:

$$\boldsymbol{\mathcal{E}}_{k+1} = \Psi_{\lambda/\beta_k}\left(\boldsymbol{\mathcal{R}}_k^2\right), \tag{23}$$

where $\Psi_{\lambda/\beta_k}(\cdot)$ is the soft thresholding [2]. The optimization details are summarized in Algorithm 2.

Complexity Analysis. At each iteration, when updating $\boldsymbol{\mathcal{J}}_{k+1}$ by Eqn. (18), the computational cost for the matrix product and inverse DFT is $\mathcal{O}\left(r_{\boldsymbol{\mathcal{A}}} n_1 n_2 n_3 + r_{\boldsymbol{\mathcal{A}}}(n_1+n_2)n_3\log(n_3)\right)$. The major cost of updating $\boldsymbol{\mathcal{Z}}_{k+1}$ by (20) includes n_3 SVD on $r_{\boldsymbol{\mathcal{A}}} \times n_2$ matrices at the cost of $\mathcal{O}(r_{\boldsymbol{\mathcal{A}}}^2 n_2 n_3)$ and computing the inverse DFT at the cost of $\mathcal{O}(r_{\boldsymbol{\mathcal{A}}} n_2 n_3 \log(n_3))$. When updating $\boldsymbol{\mathcal{E}}_{k+1}$ by (21), the step of tensor product $\boldsymbol{\mathcal{D}} * \boldsymbol{\mathcal{J}}_{k+1}$ costs $\mathcal{O}\left(r_{\boldsymbol{\mathcal{A}}} n_1 n_2 n_3 + r_{\boldsymbol{\mathcal{A}}}(n_1+n_2)n_3\log(n_3)\right)$. So the cost of Algorithm 2 is $\mathcal{O}\left(r_{\boldsymbol{\mathcal{A}}} n_1 n_2 n_3 + r_{\boldsymbol{\mathcal{A}}}(n_1+n_2)n_3\log(n_3)\right)$ for each iteration.

Note, our optimization method can be implemented in parallel, as at each iteration all frontal slices $\bar{\boldsymbol{J}}_{k+1}^{(i)}$ $(i{=}1, \cdots, n_3)$ of $\bar{\boldsymbol{\mathcal{J}}}$ can be parallelly updated which is the main computation cost when updating $\boldsymbol{\mathcal{J}}_{k+1}$. Similarly, when updating $\boldsymbol{\mathcal{Z}}'_{k+1}$, all frontal slices $(\bar{\boldsymbol{Z}}'_{k+1})^{(i)}$ $(i = 1, \cdots, n_3)$ of $\bar{\boldsymbol{\mathcal{Z}}}'$ can be parallelly computed. The tensor product required for updating $\boldsymbol{\mathcal{E}}_k$ can be also divided into n_3 matrix products in Fourier domain and thus can be parallelly updated. DFT and inverse DFT can be parallelly conducted on each tubes. So our optimization can be highly parallel. But for fairness, we adopt the serial updating scheme in our implementation, which is also very fast (see experimental results in Sec. 9 in manuscript).

4 Notations and Preliminaries

Besides the notations introduced in the main text, we introduce some additional necessary notations used in this document. Then we introduce two important properties about DFT on tensors, which are commonly used later.

4.1 Notations

The tensor spectral (or operator) norm of $\boldsymbol{\mathcal{B}}$ is defined as $\|\boldsymbol{\mathcal{B}}\| = \|\bar{\boldsymbol{\mathcal{B}}}\|$ [3]. The operator norm of an operator on tensor is defined as $\|\mathfrak{L}\| = \sup_{\|\boldsymbol{\mathcal{B}}\|_F=1}\|\mathfrak{L}(\boldsymbol{\mathcal{B}})\|_F$. The $\ell_{2,\infty}$ norm of $\boldsymbol{\mathcal{B}}$ is defined as $\|\boldsymbol{\mathcal{B}}\|_{2,\infty} = \max_i \|\boldsymbol{\mathcal{B}}(:, i, :)\|_F$.

We use $\boldsymbol{\mathcal{B}}^\dagger$ to denote the pseudo-inverse of $\boldsymbol{\mathcal{B}}$. We first compute the pseudo-inverse $(\bar{\boldsymbol{B}}^{(i)})^\dagger$ $(i = 1, \cdots, n_3)$ of the matrix $\bar{\boldsymbol{B}}^{(i)}$ which is the i-th frontal slice of the DFT result $\bar{\boldsymbol{\mathcal{B}}}$ of $\boldsymbol{\mathcal{B}}$. Then we can compute $\boldsymbol{\mathcal{B}}^\dagger = \texttt{ifft}(\bar{\boldsymbol{\mathcal{B}}}^\dagger, [\,], 3)$, where $(\bar{\boldsymbol{B}}^{(i)})^\dagger$ is the i-th frontal slice of $\bar{\boldsymbol{\mathcal{B}}}^\dagger$. Actually, we can obtain $\boldsymbol{\mathcal{B}}^\dagger = \boldsymbol{\mathcal{V}} * \boldsymbol{\mathcal{S}}^\dagger * \boldsymbol{\mathcal{U}}^*$, where $\boldsymbol{\mathcal{B}} = \boldsymbol{\mathcal{U}} * \boldsymbol{\mathcal{S}} * \boldsymbol{\mathcal{V}}^*$ is the t-SVD of $\boldsymbol{\mathcal{B}}$. It is easy to check that the computed $\boldsymbol{\mathcal{B}}^\dagger$ by the above method obeys the conditions in Definition (5).

Assume that $\boldsymbol{\mathcal{U}}_{\boldsymbol{\mathcal{A}}} * \boldsymbol{\mathcal{S}}_{\boldsymbol{\mathcal{A}}} * \boldsymbol{\mathcal{V}}_{\boldsymbol{\mathcal{A}}}^*$, $\boldsymbol{\mathcal{U}}_0 * \boldsymbol{\mathcal{S}}_0 * \boldsymbol{\mathcal{V}}_0^*$, and $\boldsymbol{\mathcal{U}} * \boldsymbol{\mathcal{S}} * \boldsymbol{\mathcal{V}}^*$ are the skinny t-SVDs of $\boldsymbol{\mathcal{A}} \in \mathbb{R}^{n_1 \times n_4 \times n_3}$, $\boldsymbol{\mathcal{L}}_0 \in \mathbb{R}^{n_1 \times n_2 \times n_3}$, and $\boldsymbol{\mathcal{A}}^\dagger * \boldsymbol{\mathcal{L}}_0 \in \mathbb{R}^{n_4 \times n_2 \times n_3}$, respectively. That is,

$$\begin{aligned}\boldsymbol{\mathcal{A}} &= \boldsymbol{\mathcal{U}}_{\boldsymbol{\mathcal{A}}} * \boldsymbol{\mathcal{S}}_{\boldsymbol{\mathcal{A}}} * \boldsymbol{\mathcal{V}}_{\boldsymbol{\mathcal{A}}}^*, \\ \boldsymbol{\mathcal{L}}_0 &= \boldsymbol{\mathcal{U}}_0 * \boldsymbol{\mathcal{S}}_0 * \boldsymbol{\mathcal{V}}_0^*, \\ \boldsymbol{\mathcal{A}}^\dagger * \boldsymbol{\mathcal{L}}_0 &= \boldsymbol{\mathcal{U}} * \boldsymbol{\mathcal{S}} * \boldsymbol{\mathcal{V}}^*.\end{aligned}$$

It also should be pointed out that for an arbitrary tensor $\boldsymbol{\mathcal{M}}$, if $\boldsymbol{\mathcal{M}} = \boldsymbol{\mathcal{U}}_{\boldsymbol{\mathcal{M}}} * \boldsymbol{\mathcal{S}}_{\boldsymbol{\mathcal{M}}} * \boldsymbol{\mathcal{V}}_{\boldsymbol{\mathcal{M}}}^*$ is its skinny t-SVD, then $\boldsymbol{\mathcal{U}}_{\boldsymbol{\mathcal{M}}}$ and $\boldsymbol{\mathcal{V}}_{\boldsymbol{\mathcal{M}}}$ obey

$$\begin{aligned}\boldsymbol{\mathcal{U}}_{\boldsymbol{\mathcal{M}}}^* * \boldsymbol{\mathcal{U}}_{\boldsymbol{\mathcal{M}}} &= \boldsymbol{\mathcal{I}}, \\ \boldsymbol{\mathcal{V}}_{\boldsymbol{\mathcal{M}}}^* * \boldsymbol{\mathcal{V}}_{\boldsymbol{\mathcal{M}}} &= \boldsymbol{\mathcal{I}}.\end{aligned}$$

Suppose that for $\forall i \in \{1, \cdots, n_3\}$, $\widehat{\boldsymbol{U}}_{\boldsymbol{\mathcal{A}}}^{(i)}\widehat{\boldsymbol{S}}_{\boldsymbol{\mathcal{A}}}^{(i)}(\widehat{\boldsymbol{V}}_{\boldsymbol{\mathcal{A}}}^{(i)})^*$, $\widehat{\boldsymbol{U}}_{\boldsymbol{\mathcal{L}}_0}^{(i)}\widehat{\boldsymbol{S}}_{\boldsymbol{\mathcal{L}}_0}^{(i)}(\widehat{\boldsymbol{V}}_{\boldsymbol{\mathcal{L}}_0}^{(i)})^*$, and $\widehat{\boldsymbol{U}}^{(i)}\widehat{\boldsymbol{S}}^{(i)}(\widehat{\boldsymbol{V}}^{(i)})^*$ are the skinny SVDs of matrices $\bar{\boldsymbol{A}}^{(i)}$, $\bar{\boldsymbol{L}}_0^{(i)}$, and $(\bar{\boldsymbol{A}}^{(i)})^\dagger\bar{\boldsymbol{L}}_0^{(i)}$, respectively. That is,

$$\begin{aligned}\bar{\boldsymbol{A}}^{(i)} &= \widehat{\boldsymbol{U}}_{\boldsymbol{\mathcal{A}}}^{(i)}\widehat{\boldsymbol{S}}_{\boldsymbol{\mathcal{A}}}^{(i)}(\widehat{\boldsymbol{V}}_{\boldsymbol{\mathcal{A}}}^{(i)})^*, \\ \bar{\boldsymbol{L}}_0^{(i)} &= \widehat{\boldsymbol{U}}_{\boldsymbol{\mathcal{L}}_0}^{(i)}\widehat{\boldsymbol{S}}_{\boldsymbol{\mathcal{L}}_0}^{(i)}(\widehat{\boldsymbol{V}}_{\boldsymbol{\mathcal{L}}_0}^{(i)})^*, \\ (\bar{\boldsymbol{A}}^{(i)})^\dagger\bar{\boldsymbol{L}}_0^{(i)} &= \widehat{\boldsymbol{U}}^{(i)}\widehat{\boldsymbol{S}}^{(i)}(\widehat{\boldsymbol{V}}^{(i)})^*.\end{aligned}$$

Next, we define the commonly used operators in this document. Define that $\mathcal{P}_{\boldsymbol{\mathcal{V}}}(\boldsymbol{\mathcal{B}}) = \boldsymbol{\mathcal{B}} * \boldsymbol{\mathcal{V}} * \boldsymbol{\mathcal{V}}^*$, $\mathcal{P}_{\boldsymbol{\mathcal{U}}}\mathcal{P}_{\boldsymbol{\mathcal{V}}}(\boldsymbol{\mathcal{B}}) = \boldsymbol{\mathcal{U}} * \boldsymbol{\mathcal{U}}^* * \boldsymbol{\mathcal{B}} * \boldsymbol{\mathcal{V}} * \boldsymbol{\mathcal{V}}^*$, $\mathcal{P}_{\boldsymbol{\mathcal{T}}}(\boldsymbol{\mathcal{B}}) = \mathcal{P}_{\boldsymbol{\mathcal{U}}}(\boldsymbol{\mathcal{B}}) + \mathcal{P}_{\boldsymbol{\mathcal{V}}}(\boldsymbol{\mathcal{B}}) - \mathcal{P}_{\boldsymbol{\mathcal{U}}}\mathcal{P}_{\boldsymbol{\mathcal{V}}}(\boldsymbol{\mathcal{B}})$, $\mathcal{P}_{\boldsymbol{\mathcal{T}}_0}(\boldsymbol{\mathcal{B}}) = \mathcal{P}_{\boldsymbol{\mathcal{U}}_0}(\boldsymbol{\mathcal{B}}) + \mathcal{P}_{\boldsymbol{\mathcal{V}}_0}(\boldsymbol{\mathcal{B}}) - \mathcal{P}_{\boldsymbol{\mathcal{U}}_0}\mathcal{P}_{\boldsymbol{\mathcal{V}}_0}(\boldsymbol{\mathcal{B}})$, and $\mathcal{P}_{\boldsymbol{\mathcal{T}}_0^{\boldsymbol{\mathcal{A}}}}(\boldsymbol{\mathcal{B}}) = \mathcal{P}_{\boldsymbol{\mathcal{U}}_{\boldsymbol{\mathcal{A}}}}\mathcal{P}_{\boldsymbol{\mathcal{T}}_0}(\boldsymbol{\mathcal{B}}) = \mathcal{P}_{\boldsymbol{\mathcal{U}}_0}(\boldsymbol{\mathcal{B}}) + \mathcal{P}_{\boldsymbol{\mathcal{U}}_{\boldsymbol{\mathcal{A}}}}\mathcal{P}_{\boldsymbol{\mathcal{V}}_0}(\boldsymbol{\mathcal{B}}) - \mathcal{P}_{\boldsymbol{\mathcal{U}}_0}\mathcal{P}_{\boldsymbol{\mathcal{V}}_0}(\boldsymbol{\mathcal{B}})$.

Now we introduce standard tensor basis defined in Definition 8, which is commonly to operate the tensors in the proofs.

Definition 8. ***(Standard tensor basis)*** *[3] For an arbitrary tensor $\mathcal{B} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$, its column basis is $\mathring{\mathfrak{e}}_i$ of size $n_1 \times 1 \times n_3$ with the $(i,1,1)$-th entry equaling to 1 and the rest equaling to 0. Similarly, the row basis is $\mathring{\mathfrak{e}}_j^*$ of size $1 \times n_2 \times n_3$ with the $(1,j,1)$-th entry equaling to 1 and the rest equaling to 0. The tube basis is $\dot{\mathfrak{e}}_k$ of size $1 \times 1 \times n_3$ with the $(1,1,k)$-th entry equaling to 1 and the rest equaling to 0.*

Based on the tensor basis, for brevity, we further define $\mathfrak{e}_{ijk} = \mathring{\mathfrak{e}}_i * \dot{\mathfrak{e}}_k * \mathring{\mathfrak{e}}_j^*$. Then for any $\mathcal{B} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$, we have $\mathcal{B} = \sum_{ijk} \langle \mathfrak{e}_{ijk}, \mathcal{B} \rangle \mathfrak{e}_{ijk} = \sum_{ijk} \mathcal{B}_{ijk} \mathfrak{e}_{ijk}$.

Finally, we give the third coherence parameter of $\mathcal{L}_0$ which is related to the dictionary $\mathcal{A}$.

Definition 9. *For a low-rank tensor $\mathcal{L}_0 \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ with a dictionary $\mathcal{A}$, assume that the skinny t-SVDs of $\mathcal{L}_0$ and $\mathcal{A}^\dagger * \mathcal{L}_0$ are respectively $\mathcal{U}_0 * \mathcal{S}_0 * \mathcal{V}_0^*$ and $\mathcal{U} * \mathcal{S} * \mathcal{V}^*$, and the dictionary $\mathcal{A}$ obeys $\mathcal{P}_{\mathcal{U}_{\mathcal{A}}}(\mathcal{U}_0) = \mathcal{U}_0$. Then the third parameter of $\mathcal{L}_0$ is defined as*

$$\mu_3^{\mathcal{A}} = \frac{n_3 \max(n_1, n_2)}{r(\kappa^{\mathcal{A}})^2 \log\left(n_3 \max(n_1, n_2)\right)} \left\| (\mathcal{A}^*)^\dagger * \mathcal{U} * \mathcal{V}^* \right\|_{2,\infty}^2,$$

where $r = \text{rank}_t(\mathcal{L}_0)$ and $\kappa^{\mathcal{A}} = \|\mathcal{A}\| \|\mathcal{A}^\dagger\|$.

Note that $\|\mathcal{A}\| = \|\bar{\mathcal{A}}\|$ is the spectral (or operator) norm which is defined in [3]. In LRR [4], Liu *et al.* also define a coherence parameter $\mu_3^{\boldsymbol{A}}$ which is related to the dictionary $\boldsymbol{A}$ used in their paper and is very similar to the definition of $\mu_3^{\mathcal{A}}$. They empirically find that $\mu_3^{\boldsymbol{A}}$ is a small value around a small constant (though the size of random matrix varies) and use this observe to simplify their proofs. Similar to LRR [3], as shown in Fig. 18 we also find that $\mu_3^{\mathcal{A}}$ is a small value and nearly invariant to the size of the testing tensor. We adopt the following experiment to investigate the value of $\mu_3^{\mathcal{A}}(\mathcal{L}_0)$. We produce the random testing tensors $\mathcal{A} \in \mathbb{R}^{n_1 \times n_4 \times n_3}$ with $\text{rank}_t(\mathcal{A}) = \lfloor 0.4 \max(n_1, n_4) \rfloor$ and $\mathcal{L}_0 \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ with $\text{rank}_t(\mathcal{L}_0) = \lfloor 0.2 \max(n_1, n_2) / \log(\max(n_1, n_2)) \rfloor$ as follows. For more simplicity, let $n_3 = 50$ and $n_4 = n_2 = n_1 = n$ which varies in $[100, 2400]$. We first produce 5 small tensors $\mathcal{B}_i \in \mathbb{R}^{n_1 \times 0.08 n_2 \times n_3}$ $(i = 1, \cdots, 5)$ and another 5 tensors $\mathcal{C}_i \in \mathbb{R}^{0.08 n_2 \times 0.2 n_2 \times n_3}$ $(i = 1, \cdots, 5)$. Note that the entries in $\mathcal{B}_i$ and $\mathcal{C}_i$ are drawn from *i.i.d.* $\mathcal{N}(0,1)$. Then let $\mathcal{A}_i = \mathcal{B}_i * \mathcal{C}_i \in \mathbb{R}^{n_1 \times 0.2 n_2 \times n_3}$ $(i = 1, \cdots, 5)$ and we further arrange $\mathcal{A}_i$ along the 2nd dimension to construct $\mathcal{A} = [\mathcal{A}_1, \cdots, \mathcal{A}_5] \in \mathbb{R}^{n_1 \times n_2 \times n_3}$. When producing $\mathcal{L}_0$, we employ similar strategy. That is, we first produce small tensors $\mathcal{G}_i \in \mathbb{R}^{n_1 \times \lfloor 0.04 n_2 / \log(n_2) \rfloor \times n_3}$ and $\mathcal{H}_i \in \mathbb{R}^{\lfloor 0.04 n_2 / \log(n_2) \rfloor \times 0.2 n_2 \times n_3}$ $(i = 1, \cdots, 5)$. However, the lateral slices in $\mathcal{G}_i$ are randomly selected from the lateral slices in $\mathcal{B}_i$ to satisfy $\mathcal{P}_{\mathcal{U}_{\mathcal{A}}}(\mathcal{U}_0) = \mathcal{U}_0$. Then we compute $\mathcal{L}_0 = [\mathcal{G}_1 * \mathcal{H}_1, \cdots, \mathcal{G}_5 * \mathcal{H}_5] \in \mathbb{R}^{n_1 \times n_2 \times n_3}$. Finally, we normalize $\mathcal{A}$ and $\mathcal{L}_0$ such that $\|\mathcal{A}\|_\infty = 1$ and $\|\mathcal{L}_0\|_\infty = 1$. By plotting the values of $\mu_3^{\mathcal{A}}(\mathcal{L}_0)$, we can easily observe that $\mu_3^{\mathcal{A}}(\mathcal{L}_0)$ is a small value and does not vary markedly when the tensor size varies. Note that $\mu_3^{\mathcal{A}}(\mathcal{L}_0)$ is only used in step [(1) Bound $\|\mathcal{F}^1\|_\infty$] in the proof of Lemma 11.

4.2 Properties of DFT on Tensors

Since the tensor nuclear norm is defined on the Fourier domain and in the proofs we will use some important properties of Discrete Fourier transformation (DFT), we introduce it first. The Fourier transformation on $\boldsymbol{v} \in \mathbb{R}^{n_3}$ is given as

$$\bar{\boldsymbol{v}} = \boldsymbol{F}_{n_3} \boldsymbol{v} \in \mathbb{C}^{n_3},$$

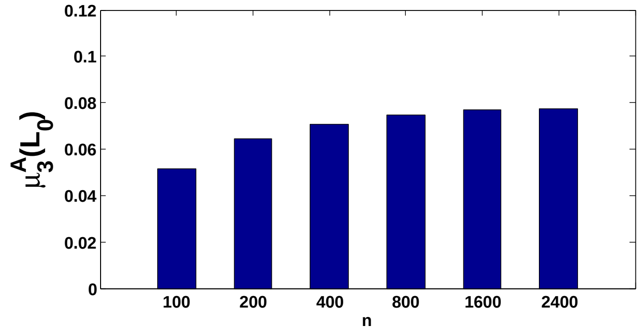


Fig. 18: Investigating the value of $\mu_3^{\mathcal{A}}$ when the size of the random testing tensor varies.

where $\boldsymbol{F}_{n_3}$ is the DFT matrix defined

$$\boldsymbol{F}_{n_3} = \begin{bmatrix} 1 & 1 & 1 & \cdots & 1 \\ 1 & \omega & \omega^2 & \cdots & \omega^{n_3-1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & \omega^{n_3-1} & \omega^{2(n_3-1)} & \cdots & \omega^{(n_3-1)(n_3-1)} \end{bmatrix} \in \mathbb{C}^{n_3 \times n_3},$$

where $\omega = e^{-\frac{2\pi \mathfrak{c}}{n_3}}$ is a primitive n_3-th root of unity in which $\mathfrak{c} = \sqrt{-1}$. Note that $\boldsymbol{F}_{n_3} / \sqrt{n_3}$ is an orthogonal matrix, *i.e.*,

$$\boldsymbol{F}_{n_3}^* \boldsymbol{F}_{n_3} = \boldsymbol{F}_{n_3} \boldsymbol{F}_{n_3}^* = n_3 \boldsymbol{I}_{n_3}. \tag{24}$$

Thus $\boldsymbol{F}_{n_3}^{-1} = \boldsymbol{F}_{n_3}^* / n_3$. When conducting DFT on a tensor $\mathcal{B} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$, it actually performs the DFT on all the tubes of $\mathcal{B}$, *i.e.* $\bar{\mathcal{B}}(i,j,:) = \boldsymbol{F}_{n_3} \mathcal{B}(i,j,:)\ \forall (i,j)$. Then, we have

$$(\boldsymbol{F}_{n_3} \otimes \boldsymbol{I}_{n_1}) \cdot \texttt{bcirc}(\mathcal{B}) \cdot (\boldsymbol{F}_{n_3}^{-1} \otimes \boldsymbol{I}_{n_2}) = \bar{\boldsymbol{B}},$$

where $\otimes$ denotes the Kronecker product and $(\boldsymbol{F}_{n_3} \otimes \boldsymbol{I}_{n_1}) / \sqrt{n_3}$ is orthogonal. By using (22), we have the following properties [5] which will be used frequently:

$$\|\mathcal{B}\|_F^2 = \frac{1}{n_3} \|\bar{\boldsymbol{B}}\|_F^2, \tag{25}$$

$$\langle \mathcal{B}, \mathcal{C} \rangle = \frac{1}{n_3} \langle \bar{\boldsymbol{B}}, \bar{\boldsymbol{C}} \rangle. \tag{26}$$

5 Proofs of Theorem 2

Proof. Let $\boldsymbol{x}_{(t)} = \texttt{vec}(\mathcal{X}_{(t)})$ and $\boldsymbol{X}_t = \texttt{squeeze}(\mathcal{X}_{(t)})$. Then, we can obtain

$$\boldsymbol{x}_{(t)} = \begin{bmatrix} \boldsymbol{X}_t \boldsymbol{e}_1 \\ \vdots \\ \boldsymbol{X}_t \boldsymbol{e}_{n_3} \end{bmatrix},$$

where $\boldsymbol{e}_j \in \mathbb{R}^{n_3}$ is the matrix column basis whose j-th entry is 1 and the rest is $\boldsymbol{0}$. On the other hand, we define a tensor $\mathcal{A} \in \mathbb{R}^{n_1 \times p' \times n_3}$ where $\mathcal{A}_{(t)} = \texttt{ivec}(\boldsymbol{A}(:,t))$ $(t = 1, \cdots, p')$. Also let $\boldsymbol{A}_t = \texttt{squeeze}(\mathcal{A}_{(t)}) \in \mathbb{R}^{n_1 \times n_3}$ $(t = 1, \cdots, p')$. Then we can rewrite $\boldsymbol{x}_{(t)} = \boldsymbol{A} \boldsymbol{z}_{(t)}$ as its equivalent form:

$$\begin{bmatrix} \boldsymbol{X}_t \boldsymbol{e}_1 \\ \vdots \\ \boldsymbol{X}_t \boldsymbol{e}_{n_3} \end{bmatrix} = \begin{bmatrix} \boldsymbol{A}_1 \boldsymbol{e}_1 & \cdots & \boldsymbol{A}_{p'} \boldsymbol{e}_1 \\ \vdots & \ddots & \vdots \\ \boldsymbol{A}_1 \boldsymbol{e}_{n_3} & \cdots & \boldsymbol{A}_{p'} \boldsymbol{e}_{n_3} \end{bmatrix} \boldsymbol{z}_{(t)}. \tag{27}$$

Since the i-th Fourier basis $\boldsymbol{f}_j = \sum_{s=1}^{n_3} \omega^{(s-1)(j-1)} \boldsymbol{e}_s$ and $\boldsymbol{X}_t\boldsymbol{e}_s = [\boldsymbol{A}_1\boldsymbol{e}_s, \cdots, \boldsymbol{A}_{p'}\boldsymbol{e}_s]\boldsymbol{z}_{(t)}$ from Eqn. (25), we have

$$\begin{aligned}
\boldsymbol{X}_t\boldsymbol{f}_j &= \sum_{s=1}^{n_3} \omega^{(s-1)(j-1)} \boldsymbol{X}_t\boldsymbol{e}_s \\
&= \sum_{s=1}^{n_3} \omega^{(s-1)(j-1)} [\boldsymbol{A}_1\boldsymbol{e}_s, \cdots, \boldsymbol{A}_{p'}\boldsymbol{e}_s]\boldsymbol{z}_{(t)} \\
&= [\boldsymbol{A}_1, \cdots, \boldsymbol{A}_{p'}] \sum_{s=1}^{n_3} \omega^{(s-1)(j-1)} \begin{bmatrix} \boldsymbol{e}_s & & \\ & \ddots & \\ & & \boldsymbol{e}_s \end{bmatrix} \boldsymbol{z}_{(t)} \\
&= [\boldsymbol{A}_1, \cdots, \boldsymbol{A}_{p'}] \begin{bmatrix} \boldsymbol{f}_j & & \\ & \ddots & \\ & & \boldsymbol{f}_j \end{bmatrix} \boldsymbol{z}_{(t)} \\
&= [\boldsymbol{A}_1\boldsymbol{f}_j, \cdots, \boldsymbol{A}_{p'}\boldsymbol{f}_j]\boldsymbol{z}_{(t)}.
\end{aligned} \tag{28}$$

For $j = 1, \cdots, n_1$, Eqn. (26) holds. We further define

$$\boldsymbol{A}_{f_i} = [\boldsymbol{A}_1\boldsymbol{f}_i, \cdots, \boldsymbol{A}_{p'}\boldsymbol{f}_i] \in \mathbb{R}^{n_1 \times n_3}$$

and a tensor $\bar{\boldsymbol{\mathcal{Z}}} \in \mathbb{R}^{p' \times n_2 \times n_3}$ where

$$\bar{\boldsymbol{\mathcal{Z}}}(:, t, j) = \boldsymbol{z}_{(t)} \in \mathbb{R}^{p'} \ (j = 1, \cdots, n_3). \tag{29}$$

Therefore, we can compute a tensor $\boldsymbol{\mathcal{Z}} = \texttt{ifft}(\bar{\boldsymbol{\mathcal{Z}}}, [], 3)$. Note that when conducting DFT on $\boldsymbol{\mathcal{A}}$, we have

$$\bar{\boldsymbol{A}} = \begin{bmatrix} \boldsymbol{A}_{f_1} & & \\ & \ddots & \\ & & \boldsymbol{A}_{f_{n_3}} \end{bmatrix}.$$

Therefore, as for the DFT result $\overline{\boldsymbol{\mathcal{X}}(:, t, :)}$ of $\boldsymbol{\mathcal{X}}(:, t, :)$, we can establish

$$\begin{aligned}
\overline{\boldsymbol{\mathcal{X}}(:, t, :)} &= \begin{bmatrix} \boldsymbol{X}_t\boldsymbol{f}_1 & & \\ & \ddots & \\ & & \boldsymbol{X}_t\boldsymbol{f}_{n_3} \end{bmatrix} \\
&= \begin{bmatrix} \boldsymbol{A}_{f_1} & & \\ & \ddots & \\ & & \boldsymbol{A}_{f_{n_3}} \end{bmatrix} \begin{bmatrix} \boldsymbol{z}_{(t)} & & \\ & \ddots & \\ & & \boldsymbol{z}_{(t)} \end{bmatrix} \\
&= \bar{\boldsymbol{A}} \overline{\boldsymbol{\mathcal{Z}}(:, t, :)}.
\end{aligned} \tag{30}$$

So for any $\boldsymbol{z}_{(t)}$, we have

$$\boldsymbol{\mathcal{X}}_{(t)} = \boldsymbol{\mathcal{A}} * \boldsymbol{\mathcal{Z}}_{(t)}, \ (t = 1, \cdots, n_2). \tag{31}$$

Conversely, for any tensor linear representation (29), $\boldsymbol{\mathcal{Z}}_{(t)}$ may not satisfy Eqn. (27). This means that there may not exist $\boldsymbol{z}_{(t)}$ such that $\boldsymbol{x}_{(t)} = \boldsymbol{A}\boldsymbol{z}_{(t)}$ holds. Thus, if the linear representation relationship in vector space holds, then there exists feasible solution such that the tensor linear representation also holds. Conversely, it does not holds. $\square$

6 Proofs of Theorem 3

Before proving Theorem 3, we first give two lemma which is used in the proofs.

Lemma 2. *Assume that* $\boldsymbol{\mathcal{U}}_{\boldsymbol{\mathcal{A}}} * \boldsymbol{\mathcal{S}}_{\boldsymbol{\mathcal{A}}} * \boldsymbol{\mathcal{V}}_{\boldsymbol{\mathcal{A}}}^*$ *and* $\boldsymbol{\mathcal{U}}_0 * \boldsymbol{\mathcal{S}}_0 * \boldsymbol{\mathcal{V}}_0^*$ *is the skinny t-SVDs of* $\boldsymbol{\mathcal{A}}$ *and* $\boldsymbol{\mathcal{L}}_0$*, respectively. Suppose that* $\mathcal{P}_{\boldsymbol{\mathcal{U}}_{\boldsymbol{\mathcal{A}}}}(\boldsymbol{\mathcal{U}}_0) = \boldsymbol{\mathcal{U}}_0$*. Then,*

$$\boldsymbol{\mathcal{Z}}_\star = \boldsymbol{\mathcal{A}}^\dagger * \boldsymbol{\mathcal{L}}_0,$$

is the minimizer to the noiseless TLRR (9) (in manuscript).

We defer the proof of Lemma 2 to Sec. 6.2. By Lemma 2, we can obtain the closed-form solution to problem (9).

Lemma 3. *Assume that* $\boldsymbol{\mathcal{C}} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$, $\boldsymbol{\mathcal{D}} \in \mathbb{R}^{n_1' \times n_2' \times n_3}$, $\boldsymbol{\mathcal{Q}} \in \mathbb{R}^{n_1' \times n_2 \times n_3}$*, and* $\boldsymbol{\mathcal{R}} \in \mathbb{R}^{n_1 \times n_2' \times n_3}$ *are four arbitrary tensors. Let*

$$\boldsymbol{\mathcal{H}} = \begin{bmatrix} \boldsymbol{\mathcal{C}} & \boldsymbol{\mathcal{R}} \\ \boldsymbol{\mathcal{Q}} & \boldsymbol{\mathcal{D}} \end{bmatrix} \in \mathbb{R}^{(n_1+n_1') \times (n_2+n_2') \times n_3}$$

and

$$\boldsymbol{\mathcal{F}} = \begin{bmatrix} \boldsymbol{\mathcal{C}} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{\mathcal{D}} \end{bmatrix} \in \mathbb{R}^{(n_1+n_1') \times (n_2+n_2') \times n_3}.$$

Then we have

$$\|\boldsymbol{\mathcal{H}}\|_* \geq \|\boldsymbol{\mathcal{F}}\|_* = \|\boldsymbol{\mathcal{C}}\|_* + \|\boldsymbol{\mathcal{D}}\|_*.$$

See its proof in Sec. 6.2. Actually this conclusion also holds when the 3rd dimension is one, *i.e.* $n_3 = 1$.

6.1 Proofs of Theorem 3

Now we are ready to prove Theorem 3.

Proof. Assume that $\boldsymbol{\mathcal{Z}}$ is an arbitrary optimal solution of noiselessness TLRR (model (9) in manuscript). Then we decompose $\boldsymbol{\mathcal{Z}}$ as $\boldsymbol{\mathcal{C}} + \boldsymbol{\mathcal{H}}$, where $\boldsymbol{\mathcal{C}}$ is a block-diagonal tensor and defined as

$$\boldsymbol{\mathcal{C}}(i, j, :) = \begin{cases} \boldsymbol{\mathcal{Z}}(i, j, :), & \text{if } \boldsymbol{\mathcal{A}}(:, i, :) \text{ and } \boldsymbol{\mathcal{X}}(:, j, :) \text{ are drawn} \\ & \text{from the same tensor subspace;} \\ \boldsymbol{0}, & \text{otherwise.} \end{cases}$$

Now suppose that $\boldsymbol{\mathcal{X}}(:, i, :)$ belongs to the p-th tensor subspace $\boldsymbol{\mathcal{K}}_p$. We further denote $\boldsymbol{\mathcal{K}}_p' = \bigoplus_{i \neq p} \boldsymbol{\mathcal{K}}_i$. Then by the definition of $\boldsymbol{\mathcal{C}}$ and $\boldsymbol{\mathcal{H}}$, we know that $\boldsymbol{\mathcal{A}} * \boldsymbol{\mathcal{C}}(:, i, :) \in \boldsymbol{\mathcal{K}}^p$, $\boldsymbol{\mathcal{A}} * \boldsymbol{\mathcal{H}}(:, i, :) \in \boldsymbol{\mathcal{K}}_p'$ and $\boldsymbol{\mathcal{A}} * \boldsymbol{\mathcal{H}}(:, i, :) \notin \boldsymbol{\mathcal{K}}^p$. However, we have $\boldsymbol{\mathcal{A}} * \boldsymbol{\mathcal{H}}(:, i, :) = \boldsymbol{\mathcal{A}} * \boldsymbol{\mathcal{Z}}(:, i, :) - \boldsymbol{\mathcal{A}} * \boldsymbol{\mathcal{C}}(:, i, :) = \boldsymbol{\mathcal{X}}(:, i, :) - \boldsymbol{\mathcal{A}} * \boldsymbol{\mathcal{C}}(:, i, :) \in \boldsymbol{\mathcal{K}}_p$. On the other hand, since $\boldsymbol{\mathcal{K}}_1, \cdots, \boldsymbol{\mathcal{K}}_k$ are independent to each other, $\boldsymbol{\mathcal{K}}_p \bigcap \left(\bigoplus_{i \neq p} \boldsymbol{\mathcal{K}}_i\right) = \boldsymbol{0}$. Thus, we have $\boldsymbol{\mathcal{A}} * \boldsymbol{\mathcal{H}}(:, i, :) = \boldsymbol{0}$.

Accordingly, we have $\boldsymbol{\mathcal{A}} * \boldsymbol{\mathcal{C}} = \boldsymbol{\mathcal{X}}$. So $\boldsymbol{\mathcal{C}}$ is also a feasible solution to noiselessness TLRR. On the other hand, by Lemma 3, we have $\|\boldsymbol{\mathcal{C}}\|_* \leq \|\boldsymbol{\mathcal{Z}}\|_*$. By Lemma 2, the optimal solution is unique and hence has block-diagonal structure. Moreover, conducting DFT would not destroy the structure, and thus each frontal slice $\bar{\boldsymbol{Z}}^{(i)}$ of $\bar{\boldsymbol{\mathcal{Z}}}$ is block-diagonal. $\square$

6.2 Proofs of Some Lemmas

6.2.1 *Proofs of Lemma 2*

Before our proof, we first introduce a lemma, which will be used later. Consider the following matrix LRR [6]:

$$\min_{\boldsymbol{Z}} \|\boldsymbol{Z}\|_*, \quad \text{s.t. } \boldsymbol{L} = \boldsymbol{A}\boldsymbol{Z}, \tag{32}$$

where $\boldsymbol{L}$ is the clean data, $\boldsymbol{A}$ is a dictionary, and $\boldsymbol{Z}$ is the representation coefficient matrix of $\boldsymbol{L}$ under dictionary $\boldsymbol{A}$. Then the following lemma analyzes the solution to problem (30).

Lemma 4. *[6] Assume* $\boldsymbol{A} \neq \boldsymbol{0}$ *and* $\boldsymbol{L} = \boldsymbol{A}\boldsymbol{Z}$ *have feasible solution(s), i.e.,* $\mathcal{P}_{U_{\boldsymbol{A}}}(U_{\boldsymbol{Z}}) = U_{\boldsymbol{Z}}$*, where* $U_{\boldsymbol{A}}$ *and* $U_{\boldsymbol{Z}}$ *are the column subspaces of* $\boldsymbol{A}$ *and* $\boldsymbol{Z}$*, respectively. Then,*

$$\boldsymbol{Z} = \boldsymbol{A}^\dagger \boldsymbol{L},$$

is the unique minimizer to problem (30), where $\boldsymbol{A}^\dagger$ *is the pseudo inverse of* $\boldsymbol{A}$*.*

Now we begin to prove Lemma 2.

Proof. Firstly, problem (31)

$$\min_{\boldsymbol{\mathcal{Z}}} \|\boldsymbol{\mathcal{Z}}\|_*, \quad \text{s.t. } \boldsymbol{\mathcal{L}} = \boldsymbol{\mathcal{A}} * \boldsymbol{\mathcal{Z}}, \tag{33}$$

is equivalent to

$$\min_{\bar{\boldsymbol{Z}}} \frac{1}{n_3}\|\bar{\boldsymbol{Z}}\|_*, \quad \text{s.t. } \bar{\boldsymbol{L}} = \bar{\boldsymbol{A}}\bar{\boldsymbol{Z}}. \tag{34}$$

Since $\bar{\boldsymbol{L}}$, $\bar{\boldsymbol{A}}$, and $\bar{\boldsymbol{Z}}$ are three block-diagonal matrices, problem (32) can be divided into n_3 smaller problems.

$$\min_{\bar{\boldsymbol{Z}}^{(i)}} \frac{1}{n_3}\|\bar{\boldsymbol{Z}}^{(i)}\|_*, \text{ s.t. } \bar{\boldsymbol{L}}^{(i)} = \bar{\boldsymbol{A}}^{(i)}\bar{\boldsymbol{Z}}^{(i)}, \ (i=1,\cdots,n_3). \tag{35}$$

Since $\mathcal{P}_{\boldsymbol{\mathcal{U}}_{\boldsymbol{\mathcal{A}}}}(\boldsymbol{\mathcal{U}}_0) = \boldsymbol{\mathcal{U}}_0$, we have $\mathcal{P}_{\bar{\boldsymbol{U}}^{(i)}}(\bar{\boldsymbol{U}}_0^{(i)}) = \bar{\boldsymbol{U}}_0^{(i)}$, where $\bar{\boldsymbol{U}}_{\boldsymbol{\mathcal{A}}}^{(i)}$ and $\bar{\boldsymbol{U}}_0^{(i)}$ are the column spaces of $\bar{\boldsymbol{A}}^{(i)}$ and $\bar{\boldsymbol{Z}}^{(i)}$, respectively. Thus, by Lemma 4, we know that $\bar{\boldsymbol{Z}}^{(i)} = (\bar{\boldsymbol{A}}^{(i)})^{\dagger}\bar{\boldsymbol{L}}^{(i)}$ is the unique optimal solution to problem (33). Hence, we can obtain the unique solution $\bar{\boldsymbol{Z}} = (\bar{\boldsymbol{A}})^{\dagger}\bar{\boldsymbol{L}}$ to problem (32). Furthermore, the unique solution to problem (31) is $\boldsymbol{\mathcal{Z}} = \boldsymbol{\mathcal{A}}^{\dagger} * \boldsymbol{\mathcal{L}}$. The proof is completed. □

6.2.2 Proofs of Lemma 3

To prove Lemma 3, we first introduce a well-known lemma.

Lemma 5. *[6], [7] Assume that $\boldsymbol{C}$, $\boldsymbol{D}$, $\boldsymbol{Q}$, and $\boldsymbol{R}$ are four arbitrary matrices of appropriate sizes. Then we have*

$$\left\|\begin{bmatrix}\boldsymbol{C} & \boldsymbol{R}\\ \boldsymbol{Q} & \boldsymbol{D}\end{bmatrix}\right\|_* \geq \left\|\begin{bmatrix}\boldsymbol{C} & \boldsymbol{0}\\ \boldsymbol{0} & \boldsymbol{D}\end{bmatrix}\right\|_* = \|\boldsymbol{C}\|_* + \|\boldsymbol{D}\|_*.$$

Now we start to prove Lemma 3.

Proof. According to tensor nuclear norm definition and Lemma 3, we have

$$\begin{aligned}\|\boldsymbol{\mathcal{H}}\|_* =& \frac{1}{n_3}\sum_{i=1}^{n_3}\|\bar{\boldsymbol{H}}^{(i)}\|_*\\ =& \frac{1}{n_3}\sum_{i=1}^{n_3}\left\|\begin{bmatrix}\bar{\boldsymbol{C}}^{(i)} & \bar{\boldsymbol{R}}^{(i)}\\ \bar{\boldsymbol{Q}}^{(i)} & \bar{\boldsymbol{D}}^{(i)}\end{bmatrix}\right\|_*\\ \geq& \frac{1}{n_3}\sum_{i=1}^{n_3}\left\|\begin{bmatrix}\bar{\boldsymbol{C}}^{(i)} & \boldsymbol{0}\\ \boldsymbol{0} & \bar{\boldsymbol{D}}^{(i)}\end{bmatrix}\right\|_*\\ =& \frac{1}{n_3}\sum_{i=1}^{n_3}\left\|\bar{\boldsymbol{F}}^{(i)}\right\|_*\\ =& \|\boldsymbol{\mathcal{F}}\|_*.\end{aligned}$$

Besides, we can also establish

$$\begin{aligned}\|\boldsymbol{\mathcal{F}}\|_* =& \frac{1}{n_3}\sum_{i=1}^{n_3}\left\|\begin{bmatrix}\bar{\boldsymbol{C}}^{(i)} & \boldsymbol{0}\\ \boldsymbol{0} & \bar{\boldsymbol{D}}^{(i)}\end{bmatrix}\right\|_*\\ =& \frac{1}{n_3}\sum_{i=1}^{n_3}\left(\left\|\bar{\boldsymbol{C}}^{(i)}\right\|_* + \left\|\bar{\boldsymbol{D}}^{(i)}\right\|_*\right)\\ =& \|\boldsymbol{\mathcal{C}}\|_* + \|\boldsymbol{\mathcal{D}}\|_*.\end{aligned}$$

Thus, the conclusion holds. □

7 Proofs of Theorem 4

Now we prove Theorem 4 in manuscript. Sec. 7.1 proves the dual conditions of the TLRR problem (problem (2) in the manuscript). Sec. 7.2 provides a way to construct the dual certificates such that the dual conditions hold. Sec. 7.3 gives the proofs of some lemmas which are used in Sec. 7.2.

Before we prove Theorem 4, we first give two lemmas that are commonly used in the proofs.

Lemma 6. *Based the definitions in Sec. 4, we have*

$$\begin{aligned}&(a)\quad \boldsymbol{\mathcal{V}}_{\boldsymbol{\mathcal{A}}} * \boldsymbol{\mathcal{V}}_{\boldsymbol{\mathcal{A}}}^* * \boldsymbol{\mathcal{U}} = \boldsymbol{\mathcal{U}},\\ &(b)\quad \boldsymbol{\mathcal{V}} * \boldsymbol{\mathcal{V}}^* = \boldsymbol{\mathcal{V}}_0 * \boldsymbol{\mathcal{V}}_0^*,\\ &(c)\quad \boldsymbol{\mathcal{A}} * \boldsymbol{\mathcal{U}} = \boldsymbol{\mathcal{L}}_0 * \boldsymbol{\mathcal{V}} * \boldsymbol{\mathcal{S}}^{\dagger} \in \mathcal{P}_{\boldsymbol{\mathcal{U}}_0},\\ &(d)\quad (\boldsymbol{\mathcal{A}}^*)^{\dagger} * \boldsymbol{\mathcal{U}} * \boldsymbol{\mathcal{V}}^* \in \mathcal{P}_{\mathcal{T}_0^{\boldsymbol{\mathcal{A}}}},\\ &(e)\quad \boldsymbol{\mathcal{A}} * \mathcal{P}_{\mathcal{T}}(\cdot) \in \mathcal{P}_{\mathcal{T}_0^{\boldsymbol{\mathcal{A}}}},\\ &(f)\quad \mathcal{P}_{\mathcal{T}}(\boldsymbol{\mathcal{A}}^*(\cdot)) = \mathcal{P}_{\mathcal{T}}\left(\boldsymbol{\mathcal{A}}^* * \mathcal{P}_{\mathcal{T}_0^{\boldsymbol{\mathcal{A}}}}\right).\end{aligned}$$

If $\boldsymbol{\mathcal{A}}$ further obeys that the rank of $\bar{\boldsymbol{A}}^{(i)}$ $(i = 1,\cdots,n_3)$ are equal to each other, then

$$(g)\quad \boldsymbol{\mathcal{A}} * \boldsymbol{\mathcal{A}}^{\dagger} = \boldsymbol{\mathcal{U}}_{\boldsymbol{\mathcal{A}}} * \boldsymbol{\mathcal{U}}_{\boldsymbol{\mathcal{A}}}^*.$$

We defer the proof of Lemma 6 to Sec. 7.3.

Lemma 7. *We have*

$$\|\mathcal{P}_{\mathcal{T}_0^{\boldsymbol{\mathcal{A}}}}(\mathfrak{e}_{ijk})\|_F^2 \leq \frac{2\mu^{\boldsymbol{\mathcal{A}}} r}{n_{(2)}n_3},$$

where $\mu^{\boldsymbol{\mathcal{A}}} = \max\left(\mu_2(\boldsymbol{\mathcal{L}}_0), \mu_1^{\boldsymbol{\mathcal{A}}}(\boldsymbol{\mathcal{L}}_0)\right)$.

See its proof in Sec. 7.3.

7.1 Dual Conditions

Lemma 8. *(Dual conditions for T-LRR) Assume $\mathcal{P}_{\boldsymbol{\mathcal{U}}_{\boldsymbol{\mathcal{A}}}}(\boldsymbol{\mathcal{U}}_0) = \boldsymbol{\mathcal{U}}_0$ and $\|\mathcal{P}_{\mathcal{T}^{\boldsymbol{\mathcal{A}}}}\mathcal{P}_{\Omega}\| < 1$. Then $(\boldsymbol{\mathcal{Z}}_{\star}, \boldsymbol{\mathcal{E}}_{\star}) = (\boldsymbol{\mathcal{A}}^{\dagger} * \boldsymbol{\mathcal{L}}_0, \boldsymbol{\mathcal{E}}_0)$ is the unique optimal solution to the T-LRR problem, provided that there are a pair $(\boldsymbol{\mathcal{W}}, \boldsymbol{\mathcal{F}})$ obeying*

$$\begin{cases}\boldsymbol{\mathcal{U}} * \boldsymbol{\mathcal{V}}^* + \boldsymbol{\mathcal{W}} = \lambda \boldsymbol{\mathcal{A}}^* * (\operatorname{sgn}(\boldsymbol{\mathcal{E}}_0) + \boldsymbol{\mathcal{F}}),\\ \mathcal{P}_{\mathcal{T}}(\boldsymbol{\mathcal{W}}) = \boldsymbol{0}, \quad \|\boldsymbol{\mathcal{W}}\| < 1,\\ \mathcal{P}_{\Omega}(\boldsymbol{\mathcal{F}}) = 0, \quad \|\boldsymbol{\mathcal{F}}\|_{\infty} < 1.\end{cases}$$

Proof. Assume that $(\boldsymbol{\mathcal{Z}}', \boldsymbol{\mathcal{E}}') = (\boldsymbol{\mathcal{A}}^{\dagger} * \boldsymbol{\mathcal{L}}_0 + \boldsymbol{\mathcal{H}}_1, \boldsymbol{\mathcal{E}}_0 - \boldsymbol{\mathcal{H}}_2)$ is also an optimal solution to the T-LRR problem. First, we have

$$\begin{aligned}&\boldsymbol{\mathcal{A}} * \left(\boldsymbol{\mathcal{A}}^{\dagger} * \boldsymbol{\mathcal{L}}_0 + \boldsymbol{\mathcal{H}}_1\right) + \boldsymbol{\mathcal{E}}_0 - \boldsymbol{\mathcal{H}}_2\\ =& \mathcal{P}_{\boldsymbol{\mathcal{U}}_{\boldsymbol{\mathcal{A}}}}(\boldsymbol{\mathcal{L}}_0) + \boldsymbol{\mathcal{A}} * \boldsymbol{\mathcal{H}}_1 + \boldsymbol{\mathcal{E}}_0 - \boldsymbol{\mathcal{H}}_2\\ =& \boldsymbol{\mathcal{L}}_0 + \boldsymbol{\mathcal{E}}_0 + \boldsymbol{\mathcal{A}} * \boldsymbol{\mathcal{H}}_1 - \boldsymbol{\mathcal{H}}_2.\end{aligned}$$

On the other hand, $(\boldsymbol{\mathcal{Z}}', \boldsymbol{\mathcal{E}}')$ obeys the constraint $\boldsymbol{\mathcal{Z}}' + \boldsymbol{\mathcal{E}}' = \boldsymbol{\mathcal{X}} = \boldsymbol{\mathcal{L}}_0 + \boldsymbol{\mathcal{E}}_0$. So we can obtain $\boldsymbol{\mathcal{A}} * \boldsymbol{\mathcal{H}}_1 = \boldsymbol{\mathcal{H}}_2$.

We then recall that the subgradient of tensor nuclear and ℓ_1 norms are as follows:

$$\begin{aligned}\partial_{\boldsymbol{\mathcal{Z}}_{\star}}\|\boldsymbol{\mathcal{Z}}\|_* &= \{\boldsymbol{\mathcal{U}} * \boldsymbol{\mathcal{V}}^* + \widehat{\boldsymbol{\mathcal{W}}}, \mid \mathcal{P}_{\mathcal{T}}(\widehat{\boldsymbol{\mathcal{W}}}) = \boldsymbol{0}, \ \|\widehat{\boldsymbol{\mathcal{W}}}\| \leq 1\},\\ \partial_{\boldsymbol{\mathcal{E}}_{\star}}\|\boldsymbol{\mathcal{E}}\|_1 &= \{\operatorname{sgn}(\boldsymbol{\mathcal{E}}_0) + \widehat{\boldsymbol{\mathcal{H}}}, \mid \mathcal{P}_{\Omega}(\widehat{\boldsymbol{\mathcal{H}}}) = \boldsymbol{0}, \ \|\widehat{\boldsymbol{\mathcal{H}}}\|_{\infty} \leq 1\}.\end{aligned}$$

Since the nuclear and the operator norms are dual, there exists a tensor $\widehat{\boldsymbol{\mathcal{W}}} \in \mathcal{P}_{\mathcal{T}^{\perp}}$ such that $\langle \widehat{\boldsymbol{\mathcal{W}}}, \mathcal{P}_{\mathcal{T}^{\perp}}(\boldsymbol{\mathcal{H}}_1)\rangle = \|\mathcal{P}_{\mathcal{T}^{\perp}}(\boldsymbol{\mathcal{H}}_1)\|_*$ and $\|\widehat{\boldsymbol{\mathcal{W}}}\| \leq 1$. Then similarly, thanks to the duality between the ℓ_1 and ℓ_∞ norms, we can pick a $\widehat{\mathcal{H}} \in \mathcal{P}_{\Omega^\perp}$ such that $\langle \widehat{\mathcal{H}}, \mathcal{P}_{\Omega^\perp}(\mathcal{H}_2)\rangle = -\|\mathcal{P}_{\Omega^\perp}(\mathcal{H}_2)\|_1$. Then by the convexity of tensor nuclear and ℓ_1 norms, we have

$$\begin{aligned}
&\|\mathcal{Z}'\|_* + \lambda\|\mathcal{E}'\|_1 - \|\mathcal{Z}_\star\|_* - \lambda\|\mathcal{E}_\star\|_1\\
\geq&\langle \mathcal{U} * \mathcal{V}^* + \widehat{\mathcal{W}}, \mathcal{Z}' - \mathcal{Z}_\star\rangle + \lambda\langle \operatorname{sgn}(\mathcal{E}_0) + \widehat{\mathcal{F}}, \mathcal{E}' - \mathcal{E}_\star\rangle\\
=&\langle \mathcal{U} * \mathcal{V}^* + \widehat{\mathcal{W}}, \mathcal{H}_1\rangle - \lambda\langle \operatorname{sgn}(\mathcal{E}_0) + \widehat{\mathcal{F}}, \mathcal{H}_2\rangle\\
=&\langle \mathcal{U} * \mathcal{V}^* - \lambda\mathcal{A}^* * \operatorname{sgn}(\mathcal{E}_0), \mathcal{H}_1\rangle + \langle \widehat{\mathcal{W}}, \mathcal{H}_1\rangle - \lambda\langle \widehat{\mathcal{F}}, \mathcal{H}_2\rangle\\
=&\langle \lambda\mathcal{A}^* * \mathcal{F} - \mathcal{W}, \mathcal{H}_1\rangle + \|\mathcal{P}_{\mathcal{T}^\perp}(\mathcal{H}_1)\|_* + \lambda\|\mathcal{P}_{\Omega^\perp}(\mathcal{H}_2)\|_1\\
=&\lambda\langle \mathcal{F}, \mathcal{H}_2\rangle - \langle \mathcal{W}, \mathcal{H}_1\rangle + \|\mathcal{P}_{\mathcal{T}^\perp}(\mathcal{H}_1)\|_* + \lambda\|\mathcal{P}_{\Omega^\perp}(\mathcal{H}_2)\|_1\\
\geq&\lambda(1 - \|\mathcal{F}\|_\infty)\|\mathcal{P}_{\Omega^\perp}(\mathcal{H}_2)\|_1 + (1 - \|\mathcal{W}\|)\|\mathcal{P}_{\mathcal{T}^\perp}(\mathcal{H}_1)\|_*.
\end{aligned}$$

Since both $(\mathcal{Z}_\star, \mathcal{E}_\star)$ and $(\mathcal{Z}', \mathcal{E}')$ are optimal, $\|\mathcal{Z}'\|_* + \lambda\|\mathcal{E}'\|_1 - \|\mathcal{Z}_\star\|_* - \lambda\|\mathcal{E}_\star\|_1 = 0$. Therefore, $\|\mathcal{P}_{\Omega^\perp}(\mathcal{H}_2)\|_1 = 0$ and $\|\mathcal{P}_{\mathcal{T}^\perp}(\mathcal{H}_1)\|_* = 0$. That is, we have $\mathcal{H}_2 \in \mathcal{P}_\Omega$ and $\mathcal{H}_1 \in \mathcal{P}_\mathcal{T}$. Since $\mathcal{H}_2 = \mathcal{A} * \mathcal{H}_1$, we have $\mathcal{H}_2 \in \mathcal{P}_{\mathcal{T}_0^\mathcal{A}}$. But $\|\mathcal{P}_{\mathcal{T}_0^\mathcal{A}}\mathcal{P}_\Omega\| < 1$ implies $\mathcal{P}_{\mathcal{T}_0^\mathcal{A}} \cap \mathcal{P}_\Omega = \{\mathbf{0}\}$. Thus, we have $\mathcal{H}_2 = \mathbf{0}$ and therefore $\mathcal{H}_1 = \mathbf{0}$. Thus the proof is completed. □

Lemma 8 implies that if we can find a dual certificate obeying

$$\begin{aligned}
&\text{(a)} \quad \mathcal{P}_\Omega(\mathcal{F}) = \mathbf{0},\\
&\text{(b)} \quad \mathcal{U} * \mathcal{V}^* = \lambda\mathcal{P}_\mathcal{T}\left(\mathcal{A}^* * (\operatorname{sgn}(\mathcal{E}_0) + \mathcal{F})\right),\\
&\text{(c)} \quad \|\mathcal{F}\|_\infty < 1,\\
&\text{(d)} \quad \lambda\|\mathcal{P}_{\mathcal{T}^\perp}\left(\mathcal{A}^* * (\operatorname{sgn}(\mathcal{E}_0) + \mathcal{F})\right)\| < 1,
\end{aligned} \tag{36}$$

then we can exactly recover the low-rank tensor $\mathcal{L}_0$ and sparse noise $\mathcal{E}_0$.

7.2 Dual Certification via the Least Squares

Before we construct the dual certificate $\mathcal{F}$, we first give some key lemmas which will be used in the construction process.

Lemma 9. *Assume $\Omega \sim Ber(\rho)$. Then with a probability at least $1 - 2(n_{(1)}n_3)^{-8}$,*

$$\left\|\mathcal{P}_{\mathcal{T}_0^\mathcal{A}} - \frac{1}{\rho}\mathcal{P}_{\mathcal{T}_0^\mathcal{A}}\mathcal{P}_\Omega\mathcal{P}_{\mathcal{T}_0^\mathcal{A}}\right\| \leq \epsilon,$$

provided that $\rho \geq 48\mu^\mathcal{A} r\log(n_{(1)}n_3)/(\epsilon^2 n_{(2)}n_3)$.

See its proof in Sec. 7.3.

Corollary 1. *Suppose $\Omega \sim Ber(\rho)$. Then with a probability at least $1 - 2(n_{(1)}n_3)^{-8}$,*

$$\|\mathcal{P}_\Omega\mathcal{P}_{\mathcal{T}_0^\mathcal{A}}\|^2 \leq (1-\rho)\epsilon + \rho < 1,$$

provided that $1 - \rho \geq 48\mu^\mathcal{A} r\log(n_{(1)}n_3)/(\epsilon^2 n_{(2)}n_3)$.

See its proof in Sec. 7.3.

Lemma 10. *[3] For the Bernoulli sign variable $\mathcal{M} = \operatorname{sgn}(\mathcal{E}_0)$ distributed as*

$$\mathcal{M}_{ijk} = \begin{cases} 1, & \textit{w.p. } \rho/2,\\ 0, & \textit{w.p. } 1-\rho,\\ -1, & \textit{w.p. } \rho/2, \end{cases} \tag{37}$$

there exists a monotone increasing function $\varphi(\rho)$ on $\rho \in [0,1]$, which also satisfies $\lim_{\rho\to 0^+}\varphi(\rho) = 0$, such that the following statement holds with high probability,

$$\|\mathcal{M}\| \leq \varphi(\rho)\sqrt{n_1 n_3}.$$

Lemma 11. *Suppose that the assumptions in Theorem 4 are satisfied. Then, the constructed $\mathcal{F}$ defined as*

$$\mathcal{F} = \mathcal{P}_{\Omega^\perp}\mathcal{P}_{\mathcal{T}_0^\mathcal{A}}\mathcal{G}\left(\frac{1}{\lambda}(\mathcal{A}^*)^\dagger * \mathcal{U} * \mathcal{V}^* - \mathcal{P}_{\mathcal{T}_0^\mathcal{A}}(\operatorname{sgn}(\mathcal{E}_0))\right),$$

where $\mathcal{G} = \sum_{k=0}^{+\infty}(\mathcal{P}_{\mathcal{T}_0^\mathcal{A}}\mathcal{P}_\Omega\mathcal{P}_{\mathcal{T}_0^\mathcal{A}})^k$, obeys the dual conditions (34).

Proof. Note that by Corollary 1, we have $\|\mathcal{P}_{\mathcal{T}_0^\mathcal{A}}\mathcal{P}_\Omega\mathcal{P}_{\mathcal{T}_0^\mathcal{A}}\| = \|\mathcal{P}_\Omega\mathcal{P}_{\mathcal{T}_0^\mathcal{A}}\|^2 < 1$. Thus, $\mathcal{F}$ is well defined. Now we verify the four conditions in (34) in turn.

Proof of (34) (a): It is easy to verify that $\mathcal{F} \in \mathcal{P}_{\Omega^\perp}$.

Proof of (34) (b): For brevity, we define

$$\mathcal{R} = \left(\frac{1}{\lambda}(\mathcal{A}^*)^\dagger * \mathcal{U} * \mathcal{V}^* - \mathcal{P}_{\mathcal{T}_0^\mathcal{A}}(\operatorname{sgn}(\mathcal{E}_0))\right).$$

Now we give an useful equality which will be used later:

$$\begin{aligned}
&\mathcal{P}_\mathcal{T}\left(\mathcal{A}^* * \mathcal{P}_{\mathcal{T}_0^\mathcal{A}}(\mathcal{F})\right)\\
=&\mathcal{P}_\mathcal{T}\left(\mathcal{A}^* * \mathcal{P}_{\mathcal{T}_0^\mathcal{A}}\mathcal{P}_{\Omega^\perp}\mathcal{P}_{\mathcal{T}_0^\mathcal{A}}\mathcal{G}(\mathcal{R})\right)\\
=&\mathcal{P}_\mathcal{T}\left(\mathcal{A}^* * \mathcal{P}_{\mathcal{T}_0^\mathcal{A}}(\mathcal{I} - \mathcal{P}_\Omega)\mathcal{P}_{\mathcal{T}_0^\mathcal{A}}\mathcal{G}(\mathcal{R})\right)\\
=&\mathcal{P}_\mathcal{T}\left(\mathcal{A}^* * \mathcal{P}_{\mathcal{T}_0^\mathcal{A}}(\mathcal{I} - \mathcal{P}_{\mathcal{T}_0^\mathcal{A}}\mathcal{P}_\Omega\mathcal{P}_{\mathcal{T}_0^\mathcal{A}})\mathcal{G}(\mathcal{R})\right)\\
=&\mathcal{P}_\mathcal{T}\left(\mathcal{A}^* * \mathcal{P}_{\mathcal{T}_0^\mathcal{A}}(\mathcal{G} - \mathcal{P}_{\mathcal{T}_0^\mathcal{A}}\mathcal{P}_\Omega\mathcal{P}_{\mathcal{T}_0^\mathcal{A}}\mathcal{G})(\mathcal{R})\right)\\
=&\mathcal{P}_\mathcal{T}\left(\mathcal{A}^* * \mathcal{P}_{\mathcal{T}_0^\mathcal{A}}\left(\frac{1}{\lambda}(\mathcal{A}^*)^\dagger * \mathcal{U} * \mathcal{V}^* - \mathcal{P}_{\mathcal{T}_0^\mathcal{A}}(\operatorname{sgn}(\mathcal{E}_0))\right)\right)\\
\overset{①}{=}&\mathcal{P}_\mathcal{T}\left(\frac{1}{\lambda}\mathcal{V}_\mathcal{A} * \mathcal{V}_\mathcal{A}{}^* * \mathcal{U} * \mathcal{V}^* - \mathcal{A}^* * \mathcal{P}_{\mathcal{T}_0^\mathcal{A}}(\operatorname{sgn}(\mathcal{E}_0))\right)\\
\overset{②}{=}&\frac{1}{\lambda}\mathcal{U} * \mathcal{V}^* - \mathcal{P}_\mathcal{T}\left(\mathcal{A}^* * \mathcal{P}_{\mathcal{T}_0^\mathcal{A}}(\operatorname{sgn}(\mathcal{E}_0))\right),
\end{aligned}$$

where ① holds because $(\mathcal{A}^*)^\dagger * \mathcal{U} * \mathcal{V}^* \in \mathcal{P}_{\mathcal{T}_0^\mathcal{A}}$ and ② holds because of $\mathcal{V}_\mathcal{A} * \mathcal{V}_\mathcal{A}{}^* * \mathcal{U} = \mathcal{U}$. Thus, we can further establish

$$\begin{aligned}
&\lambda\mathcal{P}_\mathcal{T}\left(\mathcal{A}^* * (\operatorname{sgn}(\mathcal{E}_0) + \mathcal{F})\right)\\
=&\lambda\mathcal{P}_\mathcal{T}\left(\mathcal{A}^* * \mathcal{P}_{\mathcal{T}_0^\mathcal{A}}(\operatorname{sgn}(\mathcal{E}_0) + \lambda\mathcal{F})\right)\\
=&\lambda\mathcal{P}_\mathcal{T}\left(\mathcal{A}^* * \mathcal{P}_{\mathcal{T}_0^\mathcal{A}}(\operatorname{sgn}(\mathcal{E}_0))\right) + \lambda\mathcal{P}_\mathcal{T}\left(\mathcal{A}^* * \mathcal{P}_{\mathcal{T}_0^\mathcal{A}}(\mathcal{F})\right)\\
=&\mathcal{U} * \mathcal{V}^*.
\end{aligned}$$

So $\mathcal{F}$ obeys the condition (34) (b).

Proof of (34) (c): Assume that $\mathcal{F}^1 = \mathcal{P}_{\Omega^\perp}\mathcal{P}_{\mathcal{T}_0^\mathcal{A}}\mathcal{G}(\operatorname{sgn}(\mathcal{E}_0))$ and $\mathcal{F}^2 = \frac{1}{\lambda}\mathcal{P}_{\Omega^\perp}\mathcal{P}_{\mathcal{T}_0^\mathcal{A}}\mathcal{G}\left((\mathcal{A}^*)^\dagger * \mathcal{U} * \mathcal{V}^*\right)$. Then we have $\mathcal{F} = \mathcal{F}^2 - \mathcal{F}^1$. Hence, if we can bound $\|\mathcal{F}^1\|_\infty$ and $\|\mathcal{F}^2\|_\infty$, then we can bound $\|\mathcal{F}\|_\infty \leq \|\mathcal{F}^1\|_\infty + \|\mathcal{F}^2\|_\infty$. Now we try to bound $\|\mathcal{F}^1\|_\infty$ and $\|\mathcal{F}^2\|_\infty$ in turn.

(1) Bound $\|\mathcal{F}^1\|_\infty$.
Observe that

$$\mathcal{F}^1 = \mathcal{P}_{\Omega^\perp}\mathcal{P}_{\mathcal{T}_0^\mathcal{A}}\mathcal{G}(\mathcal{M}).$$

where $\mathcal{M} = \operatorname{sgn}(\mathcal{E}_0)$ defined in Eqn. (35) is the Bernoulli sign variable. Now for $(i,j,k) \in \Omega^c$,

$$\mathcal{F}^1_{ijk} = \langle \mathcal{G}\mathcal{P}_{\mathcal{T}_0^\mathcal{A}}(\mathfrak{e}_{ijk}), \mathcal{M}\rangle := \langle \mathcal{Q}(i,j,k), \mathcal{M}\rangle.$$

Conditional on $\Omega = \operatorname{supp}(\mathcal{M})$, the signs of $\mathcal{M}$ are independent and identically distributed symmetric, and the Hoeffding's inequality gives

$$\mathbb{P}\left(|\mathcal{F}^1_{ijk}| \geq \epsilon|\Omega\right) \leq 2\exp\left(-\frac{\epsilon^2}{2\|\mathcal{Q}(i,j,k)\|_F^2}\right),$$

and

$$\mathbb{P}\left(\sup_{i,j,k}|\boldsymbol{\mathcal{F}}^1_{ijk}|\geq\epsilon|\Omega\right)\leq 2n_1n_2n_3\exp\left(-\frac{\epsilon^2}{2\sup_{i,j,k}\|\boldsymbol{\mathcal{Q}}(i,j,k)\|_F^2}\right).$$

By Lemma 1, we have $\|\boldsymbol{\mathcal{P}}_\Omega\boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{T}}_0^{\boldsymbol{\mathcal{A}}}}\|\leq\sigma=\sqrt{\rho+\epsilon(1-\rho)}$. By using Lemma 7, we have

$$\begin{aligned}\|\boldsymbol{\mathcal{P}}_\Omega\boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{T}}_0^{\boldsymbol{\mathcal{A}}}}(\mathfrak{e}_{ijk})\|_F\leq&\|\boldsymbol{\mathcal{P}}_\Omega\boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{T}}_0^{\boldsymbol{\mathcal{A}}}}\|\|\boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{T}}_0^{\boldsymbol{\mathcal{A}}}}(\mathfrak{e}_{ijk})\|_F\\ \leq&\sigma\sqrt{\frac{2\mu^{\boldsymbol{\mathcal{A}}}r}{n_{(2)}n_3}},\end{aligned}$$

on the event $\{\|\boldsymbol{\mathcal{P}}_\Omega\boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{T}}_0^{\boldsymbol{\mathcal{A}}}}\|\leq\sigma\}$. On the same event, we have $\|\boldsymbol{\mathcal{G}}\|\leq\frac{1}{1-\sigma^2}$ and thus $\|\boldsymbol{\mathcal{Q}}(i,j,k)\|_F^2\leq\frac{2\sigma^2}{(1-\sigma^2)^2}\frac{\mu^{\boldsymbol{\mathcal{A}}}r}{n_{(2)}n_3}$. Then, unconditionally,

$$\begin{aligned}&\mathbb{P}\left(\sup_{i,j,k}|\boldsymbol{\mathcal{F}}^1_{ijk}|\geq\epsilon\right)\\ \leq&2n_1n_2n_3\exp\left(-\frac{(1-\sigma^2)^2n_{(2)}n_3\epsilon^2}{4\sigma^2\mu^{\boldsymbol{\mathcal{A}}}r}\right)+\mathbb{P}(\|\boldsymbol{\mathcal{P}}_\Omega\boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{T}}_0^{\boldsymbol{\mathcal{A}}}}\|\geq\sigma).\end{aligned}$$

So when $r\leq c_4(1-\sigma^2)^2\epsilon^2n_{(2)}n_3/(4\sigma^2\mu^{\boldsymbol{\mathcal{A}}}\log(n_{(1)}n_3))$ where c_4 is a constant, then we have

$$\begin{aligned}\mathbb{P}\left(\|\boldsymbol{\mathcal{F}}^1\|_\infty<\epsilon\right)\leq&1-2n_{(1)}^{2-c_4}n_3^{1-c_4}-\mathbb{P}(\|\boldsymbol{\mathcal{P}}_\Omega\boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{T}}_0^{\boldsymbol{\mathcal{A}}}}\|\geq\sigma)\\ \leq&1-2n_{(1)}^{2-c_4}n_3^{1-c_4}-2(n_{(1)}n_3)^{-8}.\end{aligned}$$

Thus, by choosing an appropriate constant $c_4=10$, $\|\boldsymbol{\mathcal{F}}^1\|_\infty<\epsilon$ holds with a probability at least $1-4(n_{(1)}n_3)^{-8}$.

(2) Bound $\|\boldsymbol{\mathcal{F}}^2\|_\infty$.

Let $\boldsymbol{\mathcal{Q}}=\boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{T}}_0^{\boldsymbol{\mathcal{A}}}}\boldsymbol{\mathcal{G}}\left((\boldsymbol{\mathcal{A}}^*)^\dagger*\boldsymbol{\mathcal{U}}*\boldsymbol{\mathcal{V}}^*\right)$. Then $\boldsymbol{\mathcal{F}}^2=\frac{1}{\lambda}\boldsymbol{\mathcal{P}}_{\Omega^\perp}(\boldsymbol{\mathcal{Q}})$. First, we give an inequality:

$$\begin{aligned}&\|\boldsymbol{\mathcal{Q}}\|_\infty\\ =&\max_{ijk}\left\langle\boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{T}}_0^{\boldsymbol{\mathcal{A}}}}\boldsymbol{\mathcal{G}}\left((\boldsymbol{\mathcal{A}}^*)^\dagger*\boldsymbol{\mathcal{U}}*\boldsymbol{\mathcal{V}}^*\right),\mathfrak{e}_{ijk}\right\rangle\\ =&\max_{ijk}\left\langle(\boldsymbol{\mathcal{A}}^*)^\dagger*\boldsymbol{\mathcal{U}}*\boldsymbol{\mathcal{V}}^*,\boldsymbol{\mathcal{G}}\boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{T}}_0^{\boldsymbol{\mathcal{A}}}}(\mathfrak{e}_{ijk})\right\rangle\\ =&\max_{ijk}\sum_b\left\langle(\boldsymbol{\mathcal{A}}^*)^\dagger*\boldsymbol{\mathcal{U}}*\boldsymbol{\mathcal{V}}^**\mathring{\mathfrak{e}}_b,\boldsymbol{\mathcal{G}}\boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{T}}_0^{\boldsymbol{\mathcal{A}}}}(\mathfrak{e}_{ijk})*\mathring{\mathfrak{e}}_b\right\rangle\\ \leq&\max_{ijk}\sum_b\left\|(\boldsymbol{\mathcal{A}}^*)^\dagger*\boldsymbol{\mathcal{U}}*\boldsymbol{\mathcal{V}}^**\mathring{\mathfrak{e}}_b\right\|_F\left\|\boldsymbol{\mathcal{G}}\boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{T}}_0^{\boldsymbol{\mathcal{A}}}}(\mathfrak{e}_{ijk})*\mathring{\mathfrak{e}}_b\right\|_F\\ \leq&\max_{ijk}\sum_b\left\|(\boldsymbol{\mathcal{A}}^*)^\dagger*\boldsymbol{\mathcal{U}}*\boldsymbol{\mathcal{V}}^*\right\|_{2,\infty}\left\|\boldsymbol{\mathcal{G}}\boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{T}}_0^{\boldsymbol{\mathcal{A}}}}(\mathfrak{e}_{ijk})*\mathring{\mathfrak{e}}_b\right\|_F\\ =&\max_{ijk}\left\|(\boldsymbol{\mathcal{A}}^*)^\dagger*\boldsymbol{\mathcal{U}}*\boldsymbol{\mathcal{V}}^*\right\|_{2,\infty}\left\|\boldsymbol{\mathcal{G}}\boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{T}}_0^{\boldsymbol{\mathcal{A}}}}(\mathfrak{e}_{ijk})\right\|_F\\ \leq&\max_{ijk}\left\|(\boldsymbol{\mathcal{A}}^*)^\dagger*\boldsymbol{\mathcal{U}}*\boldsymbol{\mathcal{V}}^*\right\|_{2,\infty}\|\boldsymbol{\mathcal{G}}\|\left\|\boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{T}}_0^{\boldsymbol{\mathcal{A}}}}(\mathfrak{e}_{ijk})\right\|_F\\ \leq&\frac{1}{1-\sigma^2}\sqrt{\frac{2\mu^{\boldsymbol{\mathcal{A}}}r}{n_{(2)}n_3}}\left\|(\boldsymbol{\mathcal{A}}^*)^\dagger*\boldsymbol{\mathcal{U}}*\boldsymbol{\mathcal{V}}^*\right\|_{2,\infty}.\end{aligned}$$

By substituting $\mu_3^{\boldsymbol{\mathcal{A}}}(\boldsymbol{\mathcal{L}}_0)$ into the above inequality, we further obtain

$$\begin{aligned}\|\boldsymbol{\mathcal{F}}^2\|_\infty=&\|\frac{1}{\lambda}\boldsymbol{\mathcal{P}}_{\Omega^\perp}(\boldsymbol{\mathcal{Q}})\|_\infty\\ \leq&\frac{1}{\lambda}\|\boldsymbol{\mathcal{Q}}\|_\infty\\ \leq&\frac{\kappa^{\boldsymbol{\mathcal{A}}}}{\lambda(1-\sigma^2)}\sqrt{\frac{2\mu^{\boldsymbol{\mathcal{A}}}\mu_3^{\boldsymbol{\mathcal{A}}}r\log(n_{(1)}n_3)}{n_1n_2n_3^2}}\\ \leq&1-\epsilon,\end{aligned}$$

where the last inequality holds because we require

$$\lambda\geq\frac{\kappa^{\boldsymbol{\mathcal{A}}}}{(1-\epsilon)(1-\sigma^2)}\sqrt{\frac{2\mu^{\boldsymbol{\mathcal{A}}}\mu_3^{\boldsymbol{\mathcal{A}}}r\log(n_{(1)}n_3)}{n_1n_2n_3^2}},$$

which is consistent with $\lambda=1/\sqrt{n_{(1)}n_3}$. So the condition (34) (c) is also satisfied.

Proof of (34) (d): We first give some useful equalities and inequalities.

$$\begin{aligned}&\boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{T}}^\perp}\Big(\boldsymbol{\mathcal{A}}^**\boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{T}}_0^{\boldsymbol{\mathcal{A}}}}\boldsymbol{\mathcal{P}}_{\Omega^\perp}\boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{T}}_0^{\boldsymbol{\mathcal{A}}}}\boldsymbol{\mathcal{G}}\left((\boldsymbol{\mathcal{A}}^*)^\dagger*\boldsymbol{\mathcal{U}}*\boldsymbol{\mathcal{V}}^*\right)\Big)\\ =&\boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{T}}^\perp}\Big(\boldsymbol{\mathcal{A}}^**\boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{T}}_0^{\boldsymbol{\mathcal{A}}}}(\boldsymbol{\mathcal{I}}-\boldsymbol{\mathcal{P}}_\Omega)\boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{T}}_0^{\boldsymbol{\mathcal{A}}}}\boldsymbol{\mathcal{G}}\left((\boldsymbol{\mathcal{A}}^*)^\dagger*\boldsymbol{\mathcal{U}}*\boldsymbol{\mathcal{V}}^*\right)\Big)\\ =&\boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{T}}^\perp}\Big(\boldsymbol{\mathcal{A}}^**\boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{T}}_0^{\boldsymbol{\mathcal{A}}}}(\boldsymbol{\mathcal{I}}-\boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{T}}_0^{\boldsymbol{\mathcal{A}}}}\boldsymbol{\mathcal{P}}_\Omega\boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{T}}_0^{\boldsymbol{\mathcal{A}}}})\boldsymbol{\mathcal{G}}\left((\boldsymbol{\mathcal{A}}^*)^\dagger*\boldsymbol{\mathcal{U}}*\boldsymbol{\mathcal{V}}^*\right)\Big)\\ =&\boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{T}}^\perp}\Big(\boldsymbol{\mathcal{A}}^**\boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{T}}_0^{\boldsymbol{\mathcal{A}}}}\left((\boldsymbol{\mathcal{A}}^*)^\dagger*\boldsymbol{\mathcal{U}}*\boldsymbol{\mathcal{V}}^*\right)\Big)\\ =&\boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{T}}^\perp}(\boldsymbol{\mathcal{U}}*\boldsymbol{\mathcal{V}}^*)\\ =&\mathbf{0},\end{aligned}$$

where the last second equality holds because $(\boldsymbol{\mathcal{A}}^*)^\dagger*\boldsymbol{\mathcal{U}}*\boldsymbol{\mathcal{V}}^*\in\boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{T}}_0^{\boldsymbol{\mathcal{A}}}}$ and $\boldsymbol{\mathcal{V}}_{\boldsymbol{\mathcal{A}}}*\boldsymbol{\mathcal{V}}_{\boldsymbol{\mathcal{A}}}{}^**\boldsymbol{\mathcal{U}}=\boldsymbol{\mathcal{U}}$. Hence, we can further establish

$$\begin{aligned}&\lambda\boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{T}}^\perp}\left(\boldsymbol{\mathcal{A}}^**(\mathrm{sgn}\left(\boldsymbol{\mathcal{E}}_0\right)+\boldsymbol{\mathcal{F}})\right)\\ =&\boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{T}}^\perp}\Big(\boldsymbol{\mathcal{A}}^**\boldsymbol{\mathcal{P}}_{\Omega^\perp}\boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{T}}_0^{\boldsymbol{\mathcal{A}}}}\boldsymbol{\mathcal{G}}\left((\boldsymbol{\mathcal{A}}^*)^\dagger*\boldsymbol{\mathcal{U}}*\boldsymbol{\mathcal{V}}^*\right)\Big)\\ &+\lambda\boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{T}}^\perp}\Big(\boldsymbol{\mathcal{A}}^**(\boldsymbol{\mathcal{I}}-\boldsymbol{\mathcal{P}}_{\Omega^\perp}\boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{T}}_0^{\boldsymbol{\mathcal{A}}}}\boldsymbol{\mathcal{G}}\boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{T}}_0^{\boldsymbol{\mathcal{A}}}})\left(\mathrm{sgn}\left(\boldsymbol{\mathcal{E}}_0\right)\right)\Big)\\ =&\boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{T}}^\perp}\Big(\boldsymbol{\mathcal{A}}^**(\boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{T}}_0^{\boldsymbol{\mathcal{A}}}}+\boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{T}}_0^{\boldsymbol{\mathcal{A}}\perp}})\boldsymbol{\mathcal{P}}_{\Omega^\perp}\boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{T}}_0^{\boldsymbol{\mathcal{A}}}}\boldsymbol{\mathcal{G}}\left((\boldsymbol{\mathcal{A}}^*)^\dagger*\boldsymbol{\mathcal{U}}*\boldsymbol{\mathcal{V}}^*\right)\Big)\\ &+\lambda\boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{T}}^\perp}\Big(\boldsymbol{\mathcal{A}}^**(\boldsymbol{\mathcal{I}}-\boldsymbol{\mathcal{P}}_{\Omega^\perp}\boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{T}}_0^{\boldsymbol{\mathcal{A}}}}\boldsymbol{\mathcal{G}}\boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{T}}_0^{\boldsymbol{\mathcal{A}}}})\left(\mathrm{sgn}\left(\boldsymbol{\mathcal{E}}_0\right)\right)\Big)\\ =&\boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{T}}^\perp}\Big(\boldsymbol{\mathcal{A}}^**\boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{T}}_0^{\boldsymbol{\mathcal{A}}\perp}}\boldsymbol{\mathcal{P}}_{\Omega^\perp}\boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{T}}_0^{\boldsymbol{\mathcal{A}}}}\boldsymbol{\mathcal{G}}\left((\boldsymbol{\mathcal{A}}^*)^\dagger*\boldsymbol{\mathcal{U}}*\boldsymbol{\mathcal{V}}^*\right)\Big)\\ &+\lambda\boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{T}}^\perp}\Big(\boldsymbol{\mathcal{A}}^**(\boldsymbol{\mathcal{I}}-\boldsymbol{\mathcal{P}}_{\Omega^\perp}\boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{T}}_0^{\boldsymbol{\mathcal{A}}}}\boldsymbol{\mathcal{G}}\boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{T}}_0^{\boldsymbol{\mathcal{A}}}})\left(\mathrm{sgn}\left(\boldsymbol{\mathcal{E}}_0\right)\right)\Big)\\ =&\boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{T}}^\perp}\Big(\boldsymbol{\mathcal{A}}^**\boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{T}}_0^{\boldsymbol{\mathcal{A}}\perp}}(\boldsymbol{\mathcal{I}}-\boldsymbol{\mathcal{P}}_\Omega)\boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{T}}_0^{\boldsymbol{\mathcal{A}}}}\boldsymbol{\mathcal{G}}\left((\boldsymbol{\mathcal{A}}^*)^\dagger*\boldsymbol{\mathcal{U}}*\boldsymbol{\mathcal{V}}^*\right)\Big)\\ &+\lambda\boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{T}}^\perp}\Big(\boldsymbol{\mathcal{A}}^**(\boldsymbol{\mathcal{I}}-\boldsymbol{\mathcal{P}}_{\Omega^\perp}\boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{T}}_0^{\boldsymbol{\mathcal{A}}}}\boldsymbol{\mathcal{G}}\boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{T}}_0^{\boldsymbol{\mathcal{A}}}})\left(\mathrm{sgn}\left(\boldsymbol{\mathcal{E}}_0\right)\right)\Big)\\ =&\boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{T}}^\perp}\Big(\boldsymbol{\mathcal{A}}^**\boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{T}}_0^{\boldsymbol{\mathcal{A}}\perp}}\boldsymbol{\mathcal{P}}_\Omega\boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{T}}_0^{\boldsymbol{\mathcal{A}}}}\boldsymbol{\mathcal{G}}\left((\boldsymbol{\mathcal{A}}^*)^\dagger*\boldsymbol{\mathcal{U}}*\boldsymbol{\mathcal{V}}^*\right)\Big)\\ &+\lambda\boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{T}}^\perp}\Big(\boldsymbol{\mathcal{A}}^**(\boldsymbol{\mathcal{I}}-\boldsymbol{\mathcal{P}}_{\Omega^\perp}\boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{T}}_0^{\boldsymbol{\mathcal{A}}}}\boldsymbol{\mathcal{G}}\boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{T}}_0^{\boldsymbol{\mathcal{A}}}})\left(\mathrm{sgn}\left(\boldsymbol{\mathcal{E}}_0\right)\right)\Big).\end{aligned}$$

Thus, we have

$$\begin{aligned}&\|\lambda\boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{T}}^\perp}(\boldsymbol{\mathcal{A}}^**(\mathrm{sgn}\left(\boldsymbol{\mathcal{E}}_0\right)+\boldsymbol{\mathcal{F}}))\|\\ =&\left\|\boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{T}}^\perp}\Big(\boldsymbol{\mathcal{A}}^**\boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{T}}_0^{\boldsymbol{\mathcal{A}}\perp}}\boldsymbol{\mathcal{P}}_\Omega\boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{T}}_0^{\boldsymbol{\mathcal{A}}}}\boldsymbol{\mathcal{G}}\left((\boldsymbol{\mathcal{A}}^*)^\dagger*\boldsymbol{\mathcal{U}}*\boldsymbol{\mathcal{V}}^*\right)\Big)\right\|\\ &+\lambda\left\|\boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{T}}^\perp}\Big(\boldsymbol{\mathcal{A}}^**(\boldsymbol{\mathcal{I}}-\boldsymbol{\mathcal{P}}_{\Omega^\perp}\boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{T}}_0^{\boldsymbol{\mathcal{A}}}}\boldsymbol{\mathcal{G}}\boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{T}}_0^{\boldsymbol{\mathcal{A}}}})\left(\mathrm{sgn}(\boldsymbol{\mathcal{E}}_0)\right)\Big)\right\|\\ \leq&\|\boldsymbol{\mathcal{A}}\|\|\boldsymbol{\mathcal{P}}_\Omega\boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{T}}_0^{\boldsymbol{\mathcal{A}}}}\|\|\boldsymbol{\mathcal{G}}\|\|\boldsymbol{\mathcal{A}}^\dagger\|+\lambda\|\boldsymbol{\mathcal{A}}\|\left(\|\boldsymbol{\mathcal{I}}\|+\|\boldsymbol{\mathcal{G}}\|\right)\|\mathrm{sgn}(\boldsymbol{\mathcal{E}}_0)\|\\ \leq&\frac{\sigma}{1-\sigma^2}\kappa^{\boldsymbol{\mathcal{A}}}+\frac{2-\sigma^2}{1-\sigma^2}\lambda\varphi(\rho)\sqrt{n_{(1)}n_3}\|\boldsymbol{\mathcal{A}}\|\\ \overset{①}{<}&1,\end{aligned}$$

where ① holds because we require

$$\lambda<\frac{1-\sigma^2}{\varphi(\rho)(2-\sigma^2)\sqrt{n_{(1)}n_3}\|\boldsymbol{\mathcal{A}}\|}\left(1-\frac{\sigma\kappa^{\boldsymbol{\mathcal{A}}}}{1-\sigma^2}\right).$$

Let $\rho \leq (\sqrt{4\kappa^{\mathcal{A}}+1}-1)^2/(4(\kappa^{\mathcal{A}})^2)-\epsilon$, then we have $1-\sigma\kappa^{\mathcal{A}}/(1-\sigma^2)>0$. Since $\varphi(\rho)$ is monotone increasing on $\rho \in [0,1]$ and $\lim_{\rho\to 0^+}\varphi(\rho)=0$, there must exist a constant ρ' such that

$$\frac{(1-\sigma^2)\left(1-\frac{\sigma\kappa^{\mathcal{A}}}{1-\sigma^2}\right)}{\varphi(\rho')(2-\sigma^2)\|\mathcal{A}\|}\geq 1.$$

Accordingly, we have

$$\rho \leq \rho^* = \min\left(\rho', (\sqrt{4\kappa^{\mathcal{A}}+1}-1)^2/(4(\kappa^{\mathcal{A}})2)-\epsilon\right).$$

So we can set $\lambda = 1/\sqrt{n_{(1)}n_3}$. Thus, the condition (34) (d) is also satisfied.

Checking the range of r: When we prove Corollary 1 and bound $\|\boldsymbol{\mathcal{F}}^1\|_\infty$ in Proof of (10) (c), we require

$$r \leq \min\left(\frac{\epsilon^2 n_{(2)}}{48(1-\rho)\mu^{\mathcal{A}}\log(n_{(1)}n_3)}, \frac{5(1-\sigma^2)^2\epsilon^2 n_{(2)}}{2\sigma^2\mu^{\mathcal{A}}\log(n_{(1)}n_3)}\right).$$

Thus, we have

$$r \leq \frac{\rho_r n_{(2)}}{\mu^{\mathcal{A}}\log(n_{(1)}n_3)},$$

where $\rho_r = \epsilon^2/(48(1-\rho))$ is a constant. The proof is completed. □

7.3 Proof of Some Lemmas

We first introduce Lemma 12 which will be used in the proofs of Lemma 9.

Lemma 12. *(Matrix (Operator) Bernstein Inequality) [8] Let $\boldsymbol{X}_i \in \mathbb{R}^{d_1\times d_2}$ $(i=1,\cdots,s)$ be independent zero-mean, matrix valued random variables. Suppose $\|\boldsymbol{X}_i\| \leq \nu$ and $\max\left(\|\sum_i \mathbb{E}\left[\boldsymbol{X}_i\boldsymbol{X}_i^*\right]\|, \|\sum_i \mathbb{E}\left[\boldsymbol{X}_i^*\boldsymbol{X}_i\right]\|\right) \leq \omega$. Then, for any $t\geq 0$, we have*

$$\mathbb{P}\left[\left\|\sum_{i=1}^{s}\boldsymbol{X}_i\right\| > t\right] \leq (d_1+d_2)exp\left(-\frac{t^2}{2\omega+\frac{2}{3}\nu t}\right).$$

If $t \leq \omega/\nu$, then

$$\mathbb{P}\left[\left\|\sum_{i=1}^{s}\boldsymbol{X}_i\right\| > t\right] \leq (d_1+d_2)exp\left(-\frac{3t^2}{8\omega}\right).$$

7.3.1 Proof of Lemma 6

Proof. We prove (a)-(f) in turn. **(a)** Since $\boldsymbol{\mathcal{U}}*\boldsymbol{\mathcal{S}}*\boldsymbol{\mathcal{V}}^* = \boldsymbol{\mathcal{A}}^\dagger*\boldsymbol{\mathcal{L}}_0 = \boldsymbol{\mathcal{V}}_{\boldsymbol{\mathcal{A}}}*\boldsymbol{\mathcal{S}}_{\boldsymbol{\mathcal{A}}}^\dagger*\boldsymbol{\mathcal{U}}_{\boldsymbol{\mathcal{A}}}*\boldsymbol{\mathcal{L}}_0$, $\boldsymbol{\mathcal{U}}$ and $\boldsymbol{\mathcal{V}}_{\boldsymbol{\mathcal{A}}}$ span the same space. Hence, we have $\boldsymbol{\mathcal{V}}_{\boldsymbol{\mathcal{A}}}*\boldsymbol{\mathcal{V}}_{\boldsymbol{\mathcal{A}}}^**\boldsymbol{\mathcal{U}} = \boldsymbol{\mathcal{U}}$.

(b) Since $\boldsymbol{\mathcal{U}}*\boldsymbol{\mathcal{S}}*\boldsymbol{\mathcal{V}}^* = \boldsymbol{\mathcal{A}}^\dagger*\boldsymbol{\mathcal{L}}_0$, we can obtain $\widehat{\boldsymbol{U}}^{(i)}\widehat{\boldsymbol{S}}^{(i)}(\widehat{\boldsymbol{V}}^{(i)})^* = (\bar{\boldsymbol{A}}^{(i)})^\dagger\bar{\boldsymbol{L}}_0^{(i)}$ $(i=1,\cdots,n_3)$. So we can further establish

$$\begin{aligned}
&\widehat{\boldsymbol{V}}^{(i)}(\widehat{\boldsymbol{V}}^{(i)})^*\\
=&\widehat{\boldsymbol{V}}^{(i)}(\widehat{\boldsymbol{S}}^{(i)})^{-1}(\widehat{\boldsymbol{U}}^{(i)})^*(\bar{\boldsymbol{A}}^{(i)})^\dagger\bar{\boldsymbol{L}}_0^{(i)}\\
=&((\bar{\boldsymbol{A}}^{(i)})^\dagger\bar{\boldsymbol{L}}_0^{(i)})^\dagger(\bar{\boldsymbol{A}}^{(i)})^\dagger\bar{\boldsymbol{L}}_0^{(i)}\\
=&\widehat{\boldsymbol{V}}_{\boldsymbol{\mathcal{L}}_0}^{(i)}(\widehat{\boldsymbol{S}}_{\boldsymbol{\mathcal{L}}_0}^{(i)})^{-1}(\widehat{\boldsymbol{U}}_{\boldsymbol{\mathcal{L}}_0}^{(i)})^*\widehat{\boldsymbol{U}}_{\boldsymbol{\mathcal{A}}}^{(i)}\widehat{\boldsymbol{S}}_{\boldsymbol{\mathcal{A}}}^{(i)}(\widehat{\boldsymbol{V}}_{\boldsymbol{\mathcal{A}}}^{(i)})^*\widehat{\boldsymbol{V}}_{\boldsymbol{\mathcal{A}}}^{(i)}(\widehat{\boldsymbol{S}}_{\boldsymbol{\mathcal{A}}}^{(i)})^{-1}(\widehat{\boldsymbol{U}}_{\boldsymbol{\mathcal{A}}}^{(i)})^*\\
&\widehat{\boldsymbol{U}}_{\boldsymbol{\mathcal{L}}_0}^{(i)}\widehat{\boldsymbol{S}}_{\boldsymbol{\mathcal{L}}_0}^{(i)}(\widehat{\boldsymbol{V}}_{\boldsymbol{\mathcal{L}}_0}^{(i)})^*\\
=&\widehat{\boldsymbol{V}}_{\boldsymbol{\mathcal{L}}_0}^{(i)}(\widehat{\boldsymbol{V}}_{\boldsymbol{\mathcal{L}}_0}^{(i)})^*.
\end{aligned}$$

Therefore we can obtain $[\widehat{\boldsymbol{V}}^{(i)},\widetilde{\boldsymbol{V}}^{(i)}][\widehat{\boldsymbol{V}}^{(i)},\widetilde{\boldsymbol{V}}^{(i)}]^* = [\widehat{\boldsymbol{V}}_{\boldsymbol{\mathcal{L}}_0}^{(i)},\widetilde{\boldsymbol{V}}^{(i)}][\widehat{\boldsymbol{V}}_{\boldsymbol{\mathcal{L}}_0}^{(i)},\widetilde{\boldsymbol{V}}^{(i)}]^*$, where $\widetilde{\boldsymbol{V}}^{(i)} \in \mathbb{R}^{n_2\times r-r_i}$ obeys $(\widehat{\boldsymbol{V}}^{(i)})^*\widetilde{\boldsymbol{V}}^{(i)} = \boldsymbol{0}$ and $(\widetilde{\boldsymbol{V}}^{(i)})^*\widetilde{\boldsymbol{V}}^{(i)} = \boldsymbol{I}$. Thus, when we compute the skinny t-SVDs of $\boldsymbol{\mathcal{L}}_0$ and $\boldsymbol{\mathcal{A}}^\dagger*\boldsymbol{\mathcal{L}}_0$, we can let $\bar{\boldsymbol{V}}_{\boldsymbol{\mathcal{L}}_0}^{(i)} = [\widehat{\boldsymbol{V}}_{\boldsymbol{\mathcal{L}}_0}^{(i)},\widetilde{\boldsymbol{V}}^{(i)}] \in \mathbb{R}^{n_2\times r}$ and $\bar{\boldsymbol{V}}^{(i)} = [\widehat{\boldsymbol{V}}^{(i)},\widetilde{\boldsymbol{V}}^{(i)}] \in \mathbb{R}^{n_2\times r}$. This is because the last $\boldsymbol{r}_i$ columns of $\bar{\boldsymbol{V}}_{\boldsymbol{\mathcal{L}}_0}^{(i)}$ corresponding to $\boldsymbol{r}_i$ zero singular values and thus they can be constructed by the above method. Thus, we can construct $\bar{\boldsymbol{V}}^{(i)}$ by similar way. Therefore, we can obtain $\boldsymbol{\mathcal{V}}*\boldsymbol{\mathcal{V}}^* = \boldsymbol{\mathcal{V}}_0*\boldsymbol{\mathcal{V}}_0^*$.

(c) Similarly to (b), we have $\widehat{\boldsymbol{V}}_{\boldsymbol{\mathcal{A}}}^{(i)}(\widehat{\boldsymbol{S}}_{\boldsymbol{\mathcal{A}}}^{(i)})^{-1}(\widehat{\boldsymbol{U}}_{\boldsymbol{\mathcal{A}}}^{(i)})^*\bar{\boldsymbol{L}}_0^{(i)} = \widehat{\boldsymbol{U}}^{(i)}\widehat{\boldsymbol{S}}^{(i)}(\widehat{\boldsymbol{V}}^{(i)})^*$ $(i=1,\cdots,n_3)$. Therefore, we can establish $\bar{\boldsymbol{L}}_0^{(i)}\bar{\boldsymbol{V}}^{(i)}(\widehat{\boldsymbol{S}}^{(i)})^{-1}(\widehat{\boldsymbol{U}}^{(i)})^* = \widehat{\boldsymbol{U}}_{\boldsymbol{\mathcal{A}}}^{(i)}\widehat{\boldsymbol{S}}_{\boldsymbol{\mathcal{A}}}^{(i)}(\widehat{\boldsymbol{V}}_{\boldsymbol{\mathcal{A}}}^{(i)})^*$. So we have $\boldsymbol{\mathcal{A}} = \boldsymbol{\mathcal{L}}_0*\boldsymbol{\mathcal{V}}*\boldsymbol{\mathcal{S}}^\dagger*\boldsymbol{\mathcal{U}}^*$, *i.e.* $\boldsymbol{\mathcal{A}}*\boldsymbol{\mathcal{U}} = \boldsymbol{\mathcal{L}}_0*\boldsymbol{\mathcal{V}}*\boldsymbol{\mathcal{S}}^\dagger$. Thus, $\boldsymbol{\mathcal{A}}*\boldsymbol{\mathcal{U}} = \boldsymbol{\mathcal{L}}_0*\boldsymbol{\mathcal{V}}*\boldsymbol{\mathcal{S}}^\dagger \in \mathcal{P}_{\boldsymbol{\mathcal{U}}_0}$ holds.

(d) Since $(\boldsymbol{\mathcal{A}}^*)^\dagger*\boldsymbol{\mathcal{U}}*\boldsymbol{\mathcal{V}}^* = \boldsymbol{\mathcal{U}}_{\boldsymbol{\mathcal{A}}}*\boldsymbol{\mathcal{S}}_A^\dagger*\boldsymbol{\mathcal{V}}_{\boldsymbol{\mathcal{A}}}^**\boldsymbol{\mathcal{U}}*\boldsymbol{\mathcal{V}}^*$, and $\boldsymbol{\mathcal{V}}*\boldsymbol{\mathcal{V}}^* = \boldsymbol{\mathcal{V}}_0*\boldsymbol{\mathcal{V}}_0^*$, we have

$$\begin{aligned}
&\mathcal{P}_{\mathcal{T}_0^{\mathcal{A}}}\left((\boldsymbol{\mathcal{A}}^*)^\dagger*\boldsymbol{\mathcal{U}}*\boldsymbol{\mathcal{V}}^*\right)\\
=&\boldsymbol{\mathcal{U}}_0*\boldsymbol{\mathcal{U}}_0^**\left(\boldsymbol{\mathcal{U}}_{\boldsymbol{\mathcal{A}}}*\boldsymbol{\mathcal{S}}_{\boldsymbol{\mathcal{A}}}^\dagger*\boldsymbol{\mathcal{V}}_{\boldsymbol{\mathcal{A}}}^**\boldsymbol{\mathcal{U}}*\boldsymbol{\mathcal{V}}^*\right)\\
&+\boldsymbol{\mathcal{U}}_{\boldsymbol{\mathcal{A}}}*\boldsymbol{\mathcal{U}}_{\boldsymbol{\mathcal{A}}}^**\left(\boldsymbol{\mathcal{U}}_{\boldsymbol{\mathcal{A}}}*\boldsymbol{\mathcal{S}}_{\boldsymbol{\mathcal{A}}}^\dagger*\boldsymbol{\mathcal{V}}_{\boldsymbol{\mathcal{A}}}^**\boldsymbol{\mathcal{U}}*\boldsymbol{\mathcal{V}}^*\right)*\boldsymbol{\mathcal{V}}_0*\boldsymbol{\mathcal{V}}_0^*\\
&-\boldsymbol{\mathcal{U}}_0*\boldsymbol{\mathcal{U}}_0^**\left(\boldsymbol{\mathcal{U}}_{\boldsymbol{\mathcal{A}}}*\boldsymbol{\mathcal{S}}_{\boldsymbol{\mathcal{A}}}^\dagger*\boldsymbol{\mathcal{V}}_{\boldsymbol{\mathcal{A}}}^**\boldsymbol{\mathcal{U}}*\boldsymbol{\mathcal{V}}^*\right)*\boldsymbol{\mathcal{V}}_0*\boldsymbol{\mathcal{V}}_0^*\\
=&\boldsymbol{\mathcal{U}}_{\boldsymbol{\mathcal{A}}}*\boldsymbol{\mathcal{S}}_{\boldsymbol{\mathcal{A}}}^\dagger*\boldsymbol{\mathcal{V}}_{\boldsymbol{\mathcal{A}}}^**\boldsymbol{\mathcal{U}}*\boldsymbol{\mathcal{V}}^*.
\end{aligned}$$

Hence, we can obtain $(\boldsymbol{\mathcal{A}}^*)^\dagger*\boldsymbol{\mathcal{U}}*\boldsymbol{\mathcal{V}}^* \in \mathcal{P}_{\mathcal{T}_0^{\mathcal{A}}}$.

(e) Assume that $\boldsymbol{\mathcal{Q}}$ is an arbitrary tensor of proper size. Then we have

$$\begin{aligned}
&\boldsymbol{\mathcal{A}}*\mathcal{P}_{\mathcal{T}}(\boldsymbol{\mathcal{Q}})\\
=&\boldsymbol{\mathcal{A}}*\mathcal{P}_{\boldsymbol{\mathcal{U}}}(\boldsymbol{\mathcal{Q}})+\boldsymbol{\mathcal{A}}*\mathcal{P}_{\boldsymbol{\mathcal{V}}}(\boldsymbol{\mathcal{Q}})-\boldsymbol{\mathcal{A}}*\mathcal{P}_{\boldsymbol{\mathcal{U}}}\mathcal{P}_{\boldsymbol{\mathcal{V}}}(\boldsymbol{\mathcal{Q}})\\
=&\boldsymbol{\mathcal{A}}*\mathcal{P}_{\boldsymbol{\mathcal{U}}}(\boldsymbol{\mathcal{Q}})+\boldsymbol{\mathcal{A}}*\mathcal{P}_{\boldsymbol{\mathcal{V}}_0}(\boldsymbol{\mathcal{Q}})-\boldsymbol{\mathcal{A}}*\mathcal{P}_{\boldsymbol{\mathcal{U}}}\mathcal{P}_{\boldsymbol{\mathcal{V}}_0}(\boldsymbol{\mathcal{Q}})\\
\overset{①}{=}&\mathcal{P}_{\boldsymbol{\mathcal{U}}_0}(\boldsymbol{\mathcal{A}}*\mathcal{P}_{\boldsymbol{\mathcal{U}}}(\boldsymbol{\mathcal{Q}}))+\mathcal{P}_{\boldsymbol{\mathcal{U}}_{\boldsymbol{\mathcal{A}}}}(\boldsymbol{\mathcal{A}}*\mathcal{P}_{\boldsymbol{\mathcal{V}}}(\boldsymbol{\mathcal{Q}}))\\
&-\mathcal{P}_{\boldsymbol{\mathcal{U}}_0}\mathcal{P}_{\boldsymbol{\mathcal{V}}_0}(\boldsymbol{\mathcal{A}}*\mathcal{P}_{\boldsymbol{\mathcal{U}}}\mathcal{P}_{\boldsymbol{\mathcal{V}}}(\boldsymbol{\mathcal{Q}}))\\
=&\mathcal{P}_{\mathcal{T}_0^{\mathcal{A}}}\left(\boldsymbol{\mathcal{A}}*\mathcal{P}_{\mathcal{T}}(\boldsymbol{\mathcal{Q}})\right),
\end{aligned}$$

where ① holds because $\boldsymbol{\mathcal{A}}*\boldsymbol{\mathcal{U}} \in \mathcal{P}_{\boldsymbol{\mathcal{U}}_0}$, $\mathcal{P}_{\boldsymbol{\mathcal{U}}_{\boldsymbol{\mathcal{A}}}}(\boldsymbol{\mathcal{U}}_0) = \boldsymbol{\mathcal{U}}_0$ and $\boldsymbol{\mathcal{V}}*\boldsymbol{\mathcal{V}}^* = \boldsymbol{\mathcal{V}}_0*\boldsymbol{\mathcal{V}}_0^*$. Hence, we have $\boldsymbol{\mathcal{A}}*\mathcal{P}_{\mathcal{T}}(\boldsymbol{\mathcal{Q}}) \in \mathcal{P}_{\mathcal{T}_0^{\mathcal{A}}}(\cdot)$.

(f) Since by (c), we have $\mathcal{P}_{\boldsymbol{\mathcal{U}}_0}(\boldsymbol{\mathcal{A}}*\boldsymbol{\mathcal{U}}) = \boldsymbol{\mathcal{A}}*\boldsymbol{\mathcal{U}}$. Thus, $\boldsymbol{\mathcal{U}}^**\boldsymbol{\mathcal{A}}^* = \boldsymbol{\mathcal{U}}^**\boldsymbol{\mathcal{A}}^**\boldsymbol{\mathcal{U}}_0*\boldsymbol{\mathcal{U}}_0^*$ holds. Suppose that $\boldsymbol{\mathcal{Q}}$ is an arbitrary tensor of proper size. Then we can establish:

$$\begin{aligned}
&\mathcal{P}_{\mathcal{T}}\left(\boldsymbol{\mathcal{A}}^*(\boldsymbol{\mathcal{Q}})\right)\\
=&\boldsymbol{\mathcal{U}}*\boldsymbol{\mathcal{U}}^**\boldsymbol{\mathcal{A}}^**\boldsymbol{\mathcal{Q}}+\boldsymbol{\mathcal{A}}^**\boldsymbol{\mathcal{Q}}*\boldsymbol{\mathcal{V}}*\boldsymbol{\mathcal{V}}^*\\
&-\boldsymbol{\mathcal{U}}*\boldsymbol{\mathcal{U}}^**\boldsymbol{\mathcal{A}}^**\boldsymbol{\mathcal{Q}}*\boldsymbol{\mathcal{V}}*\boldsymbol{\mathcal{V}}^*\\
=&\boldsymbol{\mathcal{U}}*\boldsymbol{\mathcal{U}}^**\boldsymbol{\mathcal{A}}^**\mathcal{P}_{\boldsymbol{\mathcal{U}}_0}(\boldsymbol{\mathcal{Q}})+\boldsymbol{\mathcal{A}}^**\mathcal{P}_{\boldsymbol{\mathcal{U}}_{\boldsymbol{\mathcal{A}}}}\mathcal{P}_{\boldsymbol{\mathcal{V}}_0}(\boldsymbol{\mathcal{Q}})*\boldsymbol{\mathcal{V}}*\boldsymbol{\mathcal{V}}^*\\
&-\boldsymbol{\mathcal{U}}*\boldsymbol{\mathcal{U}}^**\boldsymbol{\mathcal{A}}^**\mathcal{P}_{\boldsymbol{\mathcal{U}}_0}\mathcal{P}_{\boldsymbol{\mathcal{V}}_0}(\boldsymbol{\mathcal{Q}})*\boldsymbol{\mathcal{V}}*\boldsymbol{\mathcal{V}}^*\\
=&\mathcal{P}_{\mathcal{T}}\left(\boldsymbol{\mathcal{A}}^**\mathcal{P}_{\mathcal{T}_0^{\mathcal{A}}}(\boldsymbol{\mathcal{Q}})\right).
\end{aligned}$$

Hence, we have $\mathcal{P}_{\mathcal{T}}\left(\boldsymbol{\mathcal{A}}(\cdot)\right) = \mathcal{P}_{\mathcal{T}}\left(\boldsymbol{\mathcal{A}}^**\mathcal{P}_{\mathcal{T}_0^{\mathcal{A}}}\right)$.

(g) Since $\boldsymbol{\mathcal{A}}$ further obeys that the rank of $\bar{\boldsymbol{A}}^{(i)}$ $(i=1,\cdots,n_3)$ are equal to each other, $\bar{\boldsymbol{A}}^{(i)}(\bar{\boldsymbol{A}}^{(i)})^\dagger = \left(\bar{\boldsymbol{U}}_{\boldsymbol{\mathcal{A}}}^{(i)}\bar{\boldsymbol{S}}_{\boldsymbol{\mathcal{A}}}^{(i)}(\bar{\boldsymbol{V}}_{\boldsymbol{\mathcal{A}}}^{(i)})^*\right)\left(\bar{\boldsymbol{V}}_{\boldsymbol{\mathcal{A}}}^{(i)}(\bar{\boldsymbol{S}}_{\boldsymbol{\mathcal{A}}}^{(i)})^\dagger(\bar{\boldsymbol{U}}_{\boldsymbol{\mathcal{A}}}^{(i)})^*\right) = \left(\bar{\boldsymbol{U}}_{\boldsymbol{\mathcal{A}}}^{(i)}\bar{\boldsymbol{S}}_{\boldsymbol{\mathcal{A}}}^{(i)}(\bar{\boldsymbol{V}}_{\boldsymbol{\mathcal{A}}}^{(i)})^*\right)\left(\bar{\boldsymbol{V}}_{\boldsymbol{\mathcal{A}}}^{(i)}(\bar{\boldsymbol{S}}_{\boldsymbol{\mathcal{A}}}^{(i)})^{-1}(\bar{\boldsymbol{U}}_{\boldsymbol{\mathcal{A}}}^{(i)})^*\right) = \bar{\boldsymbol{U}}^{(i)}(\bar{\boldsymbol{U}}^{(i)})^*$ $(i=1,\cdots,n_3)$. Thus we have $\boldsymbol{\mathcal{A}}*\boldsymbol{\mathcal{A}}^\dagger = \boldsymbol{\mathcal{U}}_{\boldsymbol{\mathcal{A}}}*(\boldsymbol{\mathcal{U}}_{\boldsymbol{\mathcal{A}}})^*$. □

7.3.2 Proof of Lemma 7

Proof. Firstly, we give an useful inequality:

$$
\begin{aligned}
&\|\mathcal{P}_{\mathcal{T}_0^{\mathcal{A}}}(\mathfrak{e}_{ijk})\|_F^2\\
=&\|\mathcal{P}_{\mathcal{U}_0}(\mathfrak{e}_{ijk})+\mathcal{P}_{\mathcal{U}_{\mathcal{A}}}\mathcal{P}_{\mathcal{V}_0}(\mathfrak{e}_{ijk})-\mathcal{P}_{\mathcal{U}_0}\mathcal{P}_{\mathcal{V}_0}(\mathfrak{e}_{ijk})\|_F^2\\
=&\|\mathcal{P}_{\mathcal{U}_0}\mathcal{P}_{\mathcal{V}_0^\perp}(\mathfrak{e}_{ijk})+\mathcal{P}_{\mathcal{U}_{\mathcal{A}}}\mathcal{P}_{\mathcal{V}_0}(\mathfrak{e}_{ijk})\|_F^2\\
\leq&\|\mathcal{P}_{\mathcal{U}_0}\mathcal{P}_{\mathcal{V}_0^\perp}(\mathfrak{e}_{ijk})\|_F^2+\|\mathcal{P}_{\mathcal{U}_{\mathcal{A}}}\mathcal{P}_{\mathcal{V}_0}(\mathfrak{e}_{ijk})\|_F^2\\
\leq&\|\mathcal{P}_{\mathcal{U}_0}(\mathfrak{e}_{ijk})\|_F^2+\|\mathcal{P}_{\mathcal{U}_{\mathcal{A}}}\mathcal{P}_{\mathcal{V}_0}(\mathfrak{e}_{ijk})\|_F^2\\
\leq&\frac{\mu_2(\mathcal{L}_0)r}{n_1n_3}+\|\mathcal{P}_{\mathcal{U}_{\mathcal{A}}}\mathcal{P}_{\mathcal{V}_0}(\mathfrak{e}_{ijk})\|_F^2\\
=&\frac{\mu_2(\mathcal{L}_0)r}{n_1n_3}+\langle\mathcal{U}_{\mathcal{A}}*\mathcal{U}_{\mathcal{A}}^**\mathfrak{e}_{ijk}*\mathcal{V}_0*\mathcal{V}_0^*,\mathfrak{e}_{ijk}\rangle\\
=&\frac{\mu_2(\mathcal{L}_0)r}{n_1n_3}+\langle\mathcal{U}_{\mathcal{A}}*\mathcal{U}_{\mathcal{A}}^**\mathfrak{e}_{ijk},\mathfrak{e}_{ijk}*\mathcal{V}_0*\mathcal{V}_0^*\rangle\\
=&\frac{\mu_2(\mathcal{L}_0)r}{n_1n_3}+\frac{1}{n_3}\langle\overline{\mathcal{U}_{\mathcal{A}}*\mathcal{U}_{\mathcal{A}}^**\mathfrak{e}_{ijk}},\overline{\mathfrak{e}_{ijk}*\mathcal{V}_0*\mathcal{V}_0^*}\rangle\\
\leq&\frac{\mu_2(\mathcal{L}_0)r}{n_1n_3}+\frac{1}{n_3}\|\overline{\mathcal{U}_{\mathcal{A}}*\mathcal{U}_{\mathcal{A}}^**\mathfrak{e}_{ijk}}\|_F\|\overline{\mathfrak{e}_{ijk}*\mathcal{V}_0*\mathcal{V}_0^*}\|_F\\
=&\frac{\mu_2(\mathcal{L}_0)r}{n_1n_3}+\|\mathcal{U}_{\mathcal{A}}*\mathcal{U}_{\mathcal{A}}^**\mathfrak{e}_{ijk}\|_F\|\mathfrak{e}_{ijk}*\mathcal{V}_0*\mathcal{V}_0^*\|_F\\
\leq&\frac{\mu_2(\mathcal{L}_0)r}{n_1n_3}+\frac{\mu_2(\mathcal{A})r^{\mathcal{A}}}{n_1n_3}\frac{\mu_1(\mathcal{L}_0)r}{n_2n_3}\\
=&\frac{\mu_2(\mathcal{L}_0)r}{n_1n_3}+\frac{\mu_1^{\mathcal{A}}(\mathcal{L}_0)r}{n_2n_3}\\
\leq&\frac{2\mu^{\mathcal{A}}r}{n_{(2)}n_3},
\end{aligned}
$$

where $\mu^{\mathcal{A}}=\max(\mu_2(\mathcal{L}_0),\mu_1^{\mathcal{A}}(\mathcal{L}_0))$. □

7.3.3 Proof of Lemma 9

Proof. For any tensor $\mathcal{Z}$, we can write

$$
\begin{aligned}
&(\rho^{-1}\mathcal{P}_{\mathcal{T}_0^{\mathcal{A}}}\mathcal{P}_\Omega\mathcal{P}_{\mathcal{T}_0^{\mathcal{A}}}-\mathcal{P}_{\mathcal{T}_0^{\mathcal{A}}})\mathcal{Z}\\
=&\sum_{ijk}\left(\frac{\delta_{ijk}}{\rho}-1\right)\langle\mathfrak{e}_{ijk},\mathcal{P}_{\mathcal{T}_0^{\mathcal{A}}}(\mathcal{Z})\rangle\mathcal{P}_{\mathcal{T}_0^{\mathcal{A}}}(\mathfrak{e}_{ijk})\\
:=&\sum_{ijk}\mathcal{H}_{ijk}(\mathcal{Z})
\end{aligned}
$$

where $\mathcal{H}_{ijk}:\mathbb{R}^{n_1\times n_2\times n_3}\to\mathbb{R}^{n_1\times n_2\times n_3}$ is a self-adjoint random operator with $\mathbb{E}[\mathcal{H}_{ijk}]=\mathbf{0}$. Define the matrix operator $\bar{\boldsymbol{H}}_{ijk}:\mathbb{B}\to\mathbb{B}$, where $\mathbb{B}=\{\bar{\boldsymbol{B}}:\mathcal{B}\in\mathbb{R}^{n_1\times n_2\times n_3}\}$ denotes the set consisting of block diagonal matrices with the blocks as the frontal slices of $\bar{\boldsymbol{B}}$, as

$$
\bar{\boldsymbol{H}}_{ijk}(\bar{\boldsymbol{Z}})=\left(\frac{\delta_{ijk}}{\rho}-1\right)\left\langle\mathfrak{e}_{ijk},\mathcal{P}_{\mathcal{T}_0^{\mathcal{A}}}(\mathcal{Z})\right\rangle\texttt{bdiag}\left(\overline{\mathcal{P}_{\mathcal{T}_0^{\mathcal{A}}}(\mathfrak{e}_{ijk})}\right).
$$

By the above definitions, we have $\|\mathcal{H}_{ijk}\|=\|\bar{\boldsymbol{H}}_{ijk}\|$ and $\|\sum_{ijk}\mathcal{H}_{ijk}\|=\|\sum_{ijk}\bar{\boldsymbol{H}}_{ijk}\|$. Also $\bar{\boldsymbol{H}}_{ijk}$ is self-adjoint and $\mathbb{E}[\bar{\boldsymbol{H}}_{ijk}]=0$. To prove the result by the non-commutative Bernstein inequality, we need to bound $\|\bar{\boldsymbol{H}}_{ijk}\|$ and $\|\sum_{ijk}\mathbb{E}[\bar{\boldsymbol{H}}_{ijk}\bar{\boldsymbol{H}}_{ijk}^*]\|$. For brevity, we define a set $\phi=\{\bar{\boldsymbol{Z}}\mid\|\bar{\boldsymbol{Z}}\|_F\leq1\}$. Then, we have

$$
\begin{aligned}
&\|\bar{\boldsymbol{H}}_{ijk}\|\\
=&\sup_\phi\|\bar{\boldsymbol{H}}_{ijk}(\bar{\boldsymbol{Z}})\|_F\\
=&\sup_\phi\left\|\left(\frac{\delta_{ijk}}{\rho}-1\right)\left\langle\mathfrak{e}_{ijk},\mathcal{P}_{\mathcal{T}_0^{\mathcal{A}}}(\mathcal{Z})\right\rangle\texttt{bdiag}\left(\overline{\mathcal{P}_{\mathcal{T}_0^{\mathcal{A}}}(\mathfrak{e}_{ijk})}\right)\right\|_F\\
=&\sup_\phi\left\|\left(\frac{\delta_{ijk}}{\rho}-1\right)\left\langle\mathcal{P}_{\mathcal{T}_0^{\mathcal{A}}}(\mathfrak{e}_{ijk}),\mathcal{Z}\right\rangle\texttt{bdiag}\left(\overline{\mathcal{P}_{\mathcal{T}_0^{\mathcal{A}}}(\mathfrak{e}_{ijk})}\right)\right\|_F\\
\leq&\sup_\phi\frac{1}{\rho}\left\|\mathcal{P}_{\mathcal{T}_0^{\mathcal{A}}}(\mathfrak{e}_{ijk})\right\|_F\left\|\texttt{bdiag}\left(\overline{\mathcal{P}_{\mathcal{T}_0^{\mathcal{A}}}(\mathfrak{e}_{ijk})}\right)\right\|_F\|\mathcal{Z}\|_F\\
=&\sup_\phi\frac{1}{\rho}\left\|\mathcal{P}_{\mathcal{T}_0^{\mathcal{A}}}(\mathfrak{e}_{ijk})\right\|_F^2\|\bar{\boldsymbol{Z}}\|_F\\
\leq&\frac{2\mu^{\mathcal{A}}r}{\rho n_{(2)}n_3}:=\nu,
\end{aligned}
$$

Note that $\mathbb{E}[(\rho^{-1}\delta_{ijk}-1)^2]=\rho^{-1}(1-\rho)\leq\rho^{-1}$. Then we can further obtain

$$
\begin{aligned}
&\left\|\sum_{ijk}\mathbb{E}[\bar{\boldsymbol{H}}_{ijk}\bar{\boldsymbol{H}}_{ijk}^*]\right\|\\
\leq&\frac{1}{\rho}\sup_\phi\left\|\sum_{ijk}\left\langle\mathfrak{e}_{ijk},\mathcal{P}_{\mathcal{T}_0^{\mathcal{A}}}(\mathcal{Z})\right\rangle\left\langle\mathfrak{e}_{ijk},\mathcal{P}_{\mathcal{T}_0^{\mathcal{A}}}(\mathfrak{e}_{ijk})\right\rangle\texttt{bdiag}\left(\overline{\mathcal{P}_{\mathcal{T}_0^{\mathcal{A}}}(\mathfrak{e}_{ijk})}\right)\right\|_F\\
\leq&\frac{1}{\rho}\sqrt{n_3}\max_{ijk}\|\mathcal{P}_{\mathcal{T}_0^{\mathcal{A}}}(\mathfrak{e}_{ijk})\|_F^2\left\|\sum_{ijk}\left\langle\mathfrak{e}_{ijk},\mathcal{P}_{\mathcal{T}_0^{\mathcal{A}}}(\mathcal{Z})\right\rangle\right\|_F\\
=&\frac{1}{\rho}\sqrt{n_3}\max_{ijk}\left\|\mathcal{P}_{\mathcal{T}_0^{\mathcal{A}}}(\mathfrak{e}_{ijk})\right\|_F^2\left\|\mathcal{P}_{\mathcal{T}_0^{\mathcal{A}}}(\mathcal{Z})\right\|_F\\
=&\frac{1}{\rho}\sqrt{n_3}\max_{ijk}\left\|\mathcal{P}_{\mathcal{T}_0^{\mathcal{A}}}(\mathfrak{e}_{ijk})\right\|_F^2\|\mathcal{P}_{\mathcal{U}_{\mathcal{A}}}\mathcal{P}_{\mathcal{T}_0}(\mathcal{Z})\|_F\\
\leq&\frac{1}{\rho}\sqrt{n_3}\max_{ijk}\left\|\mathcal{P}_{\mathcal{T}_0^{\mathcal{A}}}(\mathfrak{e}_{ijk})\right\|_F^2\|\mathcal{Z}\|_F\\
=&\frac{1}{\rho}\max_{ijk}\left\|\mathcal{P}_{\mathcal{T}_0^{\mathcal{A}}}(\mathfrak{e}_{ijk})\right\|_F^2\|\bar{\boldsymbol{Z}}\|_F\\
\leq&\frac{2\mu^{\mathcal{A}}r}{\rho n_{(2)}n_3}:=\omega.
\end{aligned}
$$

Since ϵ is small, we have $\omega/\nu=1>\epsilon$. By Lemma 12, we can establish:

$$
\begin{aligned}
&\mathbb{P}\left[\left\|\mathcal{P}_{\mathcal{T}_0^{\mathcal{A}}}-\frac{1}{\rho}\mathcal{P}_{\mathcal{T}_0^{\mathcal{A}}}\mathcal{P}_\Omega\mathcal{P}_{\mathcal{T}_0^{\mathcal{A}}}\right\|>\epsilon\right]\\
=&\mathbb{P}\left[\left\|\sum_{ijk}\mathcal{H}_{ijk}\right\|>\epsilon\right]\\
=&\mathbb{P}\left[\left\|\sum_{ijk}\bar{\boldsymbol{H}}_{ijk}\right\|>\epsilon\right]\\
\leq&(n_1+n_2)n_3\exp\left(-\frac{3}{8}\cdot\frac{\epsilon^2\rho n_{(2)}n_3}{2\mu^{\mathcal{A}}r}\right).
\end{aligned}
$$

Let $\rho\geq c_3\mu^{\mathcal{A}}r\log(n_{(1)}n_3)/(\epsilon^2n_{(2)}n_3)$. Then, the following inequality holds.

$$\begin{aligned}&\mathbb{P}\left(\left\|\boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{T}}_0^{\boldsymbol{\mathcal{A}}}}-\frac{1}{\rho}\boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{T}}_0^{\boldsymbol{\mathcal{A}}}}\boldsymbol{\mathcal{P}}_{\Omega}\boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{T}}_0^{\boldsymbol{\mathcal{A}}}}\right\|\leq\epsilon\right)\\ =&1-\mathbb{P}\left(\left\|\boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{T}}_0^{\boldsymbol{\mathcal{A}}}}-\frac{1}{\rho}\boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{T}}_0^{\boldsymbol{\mathcal{A}}}}\boldsymbol{\mathcal{P}}_{\Omega}\boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{T}}_0^{\boldsymbol{\mathcal{A}}}}\right\|>\epsilon\right)\\ \geq&1-(n_1+n_2)n_3\exp\left(-\frac{3c_3\log(n_{(1)})}{16}\right)\\ \geq&1-2(n_{(1)}n_3)^{-\frac{3}{16}c_3+1}.\end{aligned}$$

By choosing $c_3=48$, we have $\mathbb{P}(\|\boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{T}}_0^{\boldsymbol{\mathcal{A}}}}-\rho^{-1}\boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{T}}_0^{\boldsymbol{\mathcal{A}}}}\boldsymbol{\mathcal{P}}_{\Omega}\boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{T}}_0^{\boldsymbol{\mathcal{A}}}}\|\leq\epsilon)\geq 1-2(n_{(1)}n_3)^{-8}$. The proof is completed. □

7.3.4 Proof of Corollary 1

Proof. Since $\Omega^{\perp}\sim\mathrm{Ber}(1-\rho)$, by Lemma 9, with a probability at least $1-(n_{(1)}n_3)^{-8}$,

$$\left\|\boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{T}}_0^{\boldsymbol{\mathcal{A}}}}-\frac{1}{1-\rho}\boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{T}}_0^{\boldsymbol{\mathcal{A}}}}\boldsymbol{\mathcal{P}}_{\Omega^{\perp}}\boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{T}}_0^{\boldsymbol{\mathcal{A}}}}\right\|\leq\epsilon,$$

provided that $1-\rho\geq 48\mu^{\boldsymbol{\mathcal{A}}}r\log(n_{(1)}n_3)/(\epsilon^2 n_{(2)}n_3)$. Note that $\boldsymbol{\mathcal{I}}=\boldsymbol{\mathcal{P}}_{\Omega}+\boldsymbol{\mathcal{P}}_{\Omega^{\perp}}$, we have

$$\begin{aligned}&\left\|\boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{T}}_0^{\boldsymbol{\mathcal{A}}}}-\frac{1}{1-\rho}\boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{T}}_0^{\boldsymbol{\mathcal{A}}}}\boldsymbol{\mathcal{P}}_{\Omega^{\perp}}\boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{T}}_0^{\boldsymbol{\mathcal{A}}}}\right\|\\ =&\frac{1}{1-\rho}\left\|\boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{T}}_0^{\boldsymbol{\mathcal{A}}}}\boldsymbol{\mathcal{P}}_{\Omega}\boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{T}}_0^{\boldsymbol{\mathcal{A}}}}-\rho\boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{T}}_0^{\boldsymbol{\mathcal{A}}}}\right\|.\end{aligned}$$

Then, by the triangular inequality

$$\begin{aligned}\|\boldsymbol{\mathcal{P}}_{\Omega}\boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{T}}_0^{\boldsymbol{\mathcal{A}}}}\|^2=&\|\boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{T}}_0^{\boldsymbol{\mathcal{A}}}}\boldsymbol{\mathcal{P}}_{\Omega}\boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{T}}_0^{\boldsymbol{\mathcal{A}}}}\|\\ \leq&\|\boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{T}}_0^{\boldsymbol{\mathcal{A}}}}\boldsymbol{\mathcal{P}}_{\Omega}\boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{T}}_0^{\boldsymbol{\mathcal{A}}}}-\rho\boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{T}}_0^{\boldsymbol{\mathcal{A}}}}\|+\|\rho\boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{T}}_0^{\boldsymbol{\mathcal{A}}}}\|\\ \leq&(1-\rho)\epsilon+\rho.\end{aligned}$$

Thus, the conclusion is established. □

8 Proofs of Theorem 5

Proof. Assume that the pair $(\boldsymbol{\mathcal{Z}}_{\star},\boldsymbol{\mathcal{E}}_{\star})$ is the minimizer to the TLRR problem (problem (2) in the manuscript). Then we have $\boldsymbol{\mathcal{Z}}_{\star}=\boldsymbol{\mathcal{A}}^{\dagger}*(\boldsymbol{\mathcal{X}}-\boldsymbol{\mathcal{E}}_{\star})$. Thus, there exists $\boldsymbol{\mathcal{Z}}_{\star}'$ such that $\boldsymbol{\mathcal{Z}}_{\star}=\boldsymbol{\mathcal{V}}_{\boldsymbol{\mathcal{A}}}*\boldsymbol{\mathcal{Z}}_{\star}'$, where $\boldsymbol{\mathcal{U}}_{\boldsymbol{\mathcal{A}}}*\boldsymbol{\mathcal{S}}_{\boldsymbol{\mathcal{A}}}*\boldsymbol{\mathcal{V}}_{\boldsymbol{\mathcal{A}}}^*$ is the skinny t-SVD of $\boldsymbol{\mathcal{A}}$. On the other hand, we have

$$\begin{aligned}\|\boldsymbol{\mathcal{Z}}_{\star}\|_*&=\|\boldsymbol{\mathcal{V}}_{\boldsymbol{\mathcal{A}}}*\boldsymbol{\mathcal{Z}}_{\star}'\|_*\\ &=\frac{1}{n_3}\sum_{i=1}^{n_3}\|\bar{\boldsymbol{V}}_{\boldsymbol{\mathcal{A}}}^{(i)}(\bar{\boldsymbol{Z}}_{\star}')^{(i)}\|_*\\ &=\frac{1}{n_3}\sum_{i=1}^{n_3}\mathrm{tr}\left(\sqrt{((\bar{\boldsymbol{Z}}_{\star}')^{(i)})^*(\bar{\boldsymbol{V}}_{\boldsymbol{\mathcal{A}}}^{(i)})^*\bar{\boldsymbol{V}}_{\boldsymbol{\mathcal{A}}}^{(i)}(\bar{\boldsymbol{Z}}_{\star}')^{(i)}}\right)\\ &\overset{①}{=}\frac{1}{n_3}\sum_{i=1}^{n_3}\mathrm{tr}\left(\sqrt{((\bar{\boldsymbol{Z}}_{\star}')^{(i)})^*(\bar{\boldsymbol{Z}}_{\star}')^{(i)}}\right)\\ &=\frac{1}{n_3}\sum_{i=1}^{n_3}\|(\bar{\boldsymbol{Z}}_{\star}')^{(i)})\|_*\\ &=\|\boldsymbol{\mathcal{Z}}_{\star}'\|_*,\end{aligned}$$

where ① holds since $(\bar{\boldsymbol{V}}_{\boldsymbol{\mathcal{A}}}^{(i)})^**\bar{\boldsymbol{V}}_{\boldsymbol{\mathcal{A}}}^{(i)}=\boldsymbol{I}$. Thus, we can substitute $\boldsymbol{\mathcal{Z}}_{\star}=\boldsymbol{\mathcal{V}}_{\boldsymbol{\mathcal{A}}}*\boldsymbol{\mathcal{Z}}_{\star}'$ into problem (2) in the manuscript and then obtain its equivalent problem (13). Thus, if $(\boldsymbol{\mathcal{Z}}_{\star}',\boldsymbol{\mathcal{E}}_{\star})$ is an arbitrary optimal solution to problem (13), then we can obtain the minimizer $(\boldsymbol{\mathcal{V}}_{\boldsymbol{\mathcal{A}}}*\boldsymbol{\mathcal{Z}}_{\star}',\boldsymbol{\mathcal{E}}_{\star})$ of problem (2). □

References

[1] J. Cai, E. Candès, and Z. Shen, "A singular value thresholding algorithm for matrix completion," *SIAM J. on Optimization*, vol. 20, no. 4, pp. 1956–1982, 2008.

[2] D. Donoho and I. Johnstone, "Adapting to unknown smoothness via wavelet shrinkage," *Journal of the American Statistical Association*, vol. 90, no. 432, pp. 1200–1224, 1999.

[3] C. Lu, J. Feng, Y. Chen, W. Liu, Z. Lin, and S. Yan, "Tensor robust principal component analysis: Exact recovery of corrupted low-rank tensors via convex optimization," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, 2016, pp. 5249–5257.

[4] G. Liu, H. Xu, J. Tang, Q. Liu, and S. Yan, "A deterministic analysis for LRR," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 38, no. 3, pp. 417–430, 2016.

[5] M. Kilmer and C. Martin, "Factorization strategies for third-order tensors," *Linear Algebra and its Applications*, vol. 435, no. 3, pp. 641–658, 2011.

[6] G. Liu, Z. Lin, and Y. Yu, "Robust subspace segmentation by low-rank representation," in *Proc. Int'l Conf. Machine Learning*, 2010, pp. 663–670.

[7] R. Horn and C. Johnson, "Topics in matrix analysis," *Cambridge University Presss*, 1991.

[8] J. A. Tropp, "An introduction to matrix concentration inequalities," *arXiv preprint:1501.01571*, 2015.