Supplementary Material of Outlier-Robust Tensor PCA

Pan Zhou* Jiashi Feng*

* National University of Singapore, Singapore

pzhou@u.edu.cn elefjia@nus.edu.sg

1. Structure of This Document

This document gives details on the deduction of Algorithm 1 and the proofs of Theorems 1 and 2 in the manuscript. Sec. 2 gives some other notations and properties which will be used in the proofs. Sec. 3 presents the details of Algorithm 1. Sec. 4 provides the proofs of Theorem 1. In Sec. 5, we will prove Theorem 2.

2. 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 used later.

2.1. Notations

The tensor spectral (or operator) norm of $\boldsymbol{\mathcal{A}}$ is defined as $\|\boldsymbol{\mathcal{A}}\| = \|\bar{\boldsymbol{A}}\|$. The operator norm of an operator on tensor is defined as $\|\mathfrak{L}\| = \sup_{\|\boldsymbol{\mathcal{A}}\|_F=1} \|\mathfrak{L}(\boldsymbol{\mathcal{A}})\|_F$. The inner product of two tensors $\boldsymbol{\mathcal{A}}$ and $\boldsymbol{\mathcal{B}}$ in $n_1 \times n_2 \times n_3$ is defined as $\langle \boldsymbol{\mathcal{A}}, \boldsymbol{\mathcal{B}} \rangle = \sum_{i=1}^{n_3} \langle \boldsymbol{A}^{(i)}, \boldsymbol{B}^{(i)} \rangle$. The dual norm of tensor $\ell_{2,1}$ norm is the tensor $\ell_{2,\infty}$ norm defined as $\|\boldsymbol{\mathcal{A}}\|_{2,\infty} = \max_i \|\boldsymbol{\mathcal{A}}(:,i,:)\|_F$.

Next, we define the commonly used operators in this document. Let $\boldsymbol{\mathcal{U}} * \boldsymbol{\mathcal{S}} * \boldsymbol{\mathcal{V}}^*$ be the tensor SVD of $\boldsymbol{\mathcal{L}}$. Then, the projection onto the row space $\boldsymbol{\mathcal{V}}$ is given by $\mathcal{P}_{\boldsymbol{\mathcal{V}}}(\boldsymbol{\mathcal{A}}) = \boldsymbol{\mathcal{A}} * \boldsymbol{\mathcal{V}} * \boldsymbol{\mathcal{V}}^*$. The projection to the union of the column space $\boldsymbol{\mathcal{U}}$ and the row space $\boldsymbol{\mathcal{V}}$ is denoted by $\mathcal{P}_{\boldsymbol{\mathcal{T}}}(\boldsymbol{\mathcal{A}}) = \mathcal{P}_{\boldsymbol{\mathcal{U}}}(\boldsymbol{\mathcal{A}}) + \mathcal{P}_{\boldsymbol{\mathcal{V}}}(\boldsymbol{\mathcal{A}}) - \mathcal{P}_{\boldsymbol{\mathcal{U}}}\mathcal{P}_{\boldsymbol{\mathcal{V}}}(\boldsymbol{\mathcal{A}})$, where $\mathcal{P}_{\boldsymbol{\mathcal{U}}}\mathcal{P}_{\boldsymbol{\mathcal{V}}}(\boldsymbol{\mathcal{A}}) = \boldsymbol{\mathcal{U}}*\boldsymbol{\mathcal{U}}^**\boldsymbol{\mathcal{A}}*\boldsymbol{\mathcal{V}}*\boldsymbol{\mathcal{V}}^*$. The orthogonal complement of $\mathcal{P}_{\boldsymbol{\mathcal{U}}}, \mathcal{P}_{\boldsymbol{\mathcal{V}}}$ and $\mathcal{P}_{\boldsymbol{\mathcal{T}}}$ is denoted by $\mathcal{P}_{\boldsymbol{\mathcal{U}}^\perp} = \boldsymbol{\mathcal{I}} - \mathcal{P}_{\boldsymbol{\mathcal{U}}}$, $\mathcal{P}_{\boldsymbol{\mathcal{V}}^\perp} = \boldsymbol{\mathcal{I}} - \mathcal{P}_{\boldsymbol{\mathcal{V}}}$ and $\mathcal{P}_{\boldsymbol{\mathcal{T}}^\perp} = \boldsymbol{\mathcal{I}} - \mathcal{P}_{\boldsymbol{\mathcal{T}}}$, respectively. Note that the variants of above notations also denote the similar meanings.

Finally, we introduce standard tensor basis defined in Definition 1, which is commonly used in the proofs.

Definition 1. ***(Standard tensor basis)*** *[1] For an arbitrary tensor $\boldsymbol{\mathcal{A}} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$, 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 $\mathfrak{e}_{ijk} = \mathring{\mathfrak{e}}_i * \dot{\mathfrak{e}}_k * \mathring{\mathfrak{e}}_j^*$.

2.2. 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 $\boldsymbol{v} \in \mathbb{R}^n$ is given as

$$\bar{\boldsymbol{v}} = \boldsymbol{F}_n \boldsymbol{v} \in \mathbb{C}^n,$$

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

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

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

$$\boldsymbol{F}_n^* \boldsymbol{F}_n = \boldsymbol{F}_n \boldsymbol{F}_n^* = n\boldsymbol{I}_n. \tag{9}$$

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

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

where $\otimes$ denotes the Kronecker product and $(\boldsymbol{F}_{n_3} \otimes \boldsymbol{I}_{n_1})/\sqrt{n_3}$ is orthogonal. By using (9), we have the following properties which will be used frequently:

$$\|\boldsymbol{\mathcal{A}}\|_F^2 = \frac{1}{n_3}\|\bar{\boldsymbol{A}}\|_F^2, \tag{10}$$

$$\langle \boldsymbol{\mathcal{A}}, \boldsymbol{\mathcal{B}} \rangle = \frac{1}{n_3}\langle \bar{\boldsymbol{A}}, \bar{\boldsymbol{B}} \rangle. \tag{11}$$

3. Details of Algorithm 1

In this section, we use ADMM [2] to solve problem problem (2) in manuscript because of its efficiency in solving this problem, which includes linear constraint and involves the nuclear and $\ell_{2,1}$ norms. Besides, since problem (2) is convex with two blocks of variables, the theoretical convergence analysis in [2] can guarantee the convergence of ADMM. We first write down the augmented Lagrangian function:

$$H(\boldsymbol{\mathcal{L}},\boldsymbol{\mathcal{E}},\boldsymbol{\mathcal{J}},\beta) = \|\boldsymbol{\mathcal{L}}\|_* + \lambda\|\boldsymbol{\mathcal{E}}\|_1 + \langle \boldsymbol{\mathcal{J}}, \boldsymbol{\mathcal{X}} - \boldsymbol{\mathcal{L}} - \boldsymbol{\mathcal{E}}\rangle + \frac{\beta}{2}\|\boldsymbol{\mathcal{X}} - \boldsymbol{\mathcal{L}} - \boldsymbol{\mathcal{E}}\|_F^2,$$

where $\boldsymbol{\mathcal{J}}$ is Lagrange multiplier and β is the penalty parameter. Then we can alternately update $\boldsymbol{\mathcal{L}}$ and $\boldsymbol{\mathcal{E}}$ in each iteration by minimizing $H(\boldsymbol{\mathcal{L}},\boldsymbol{\mathcal{E}},\boldsymbol{\mathcal{J}},\beta)$ with other variables fixed.

Updating $\boldsymbol{\mathcal{L}}$: We only need to minimize the following optimization problem:

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

Thus, we can optimize its equivalent problem:

$$\bar{\boldsymbol{L}}_{k+1} = \operatorname*{argmin}_{\bar{\boldsymbol{\mathcal{L}}}} \frac{1}{n_3}\left(\|\bar{\boldsymbol{L}}\|_* + \frac{\beta_k}{2}\|\bar{\boldsymbol{R}} - \bar{\boldsymbol{L}}\|_F^2\right),$$

where $\boldsymbol{\mathcal{R}} = \boldsymbol{\mathcal{X}} - \boldsymbol{\mathcal{E}}_k + \boldsymbol{\mathcal{J}}_k/\beta_k$, $\bar{\boldsymbol{\mathcal{R}}} = \texttt{fft}(\boldsymbol{\mathcal{R}}, [], 3)$ and $\bar{\boldsymbol{R}} = \texttt{bdiag}(\bar{\boldsymbol{\mathcal{R}}})$. Note that $\bar{\boldsymbol{L}}$ is a diagonal block matrix. Accordingly, we only need to update all the diagonal block matrices $\bar{\boldsymbol{L}}^{(i)}$ $(i = 1, \cdots, n_3)$ by

$$\bar{\boldsymbol{L}}_{k+1}^{(i)} = \mathcal{S}_{\frac{1}{\beta_k}}\left(\bar{\boldsymbol{R}}^{(i)}\right), \ (i = 1, \cdots, n_3), \tag{12}$$

where $\mathcal{S}_\nu(\cdot)$ is the singular value thresholding (SVT) operator [3]. Finally, we can compute $\boldsymbol{\mathcal{L}}_{k+1} = \texttt{ifft}(\bar{\boldsymbol{\mathcal{L}}}_{k+1}, [], 3)$.

Updating $\boldsymbol{\mathcal{E}}$: We can update $\boldsymbol{\mathcal{E}}$ by solving

$$\boldsymbol{\mathcal{E}}_{k+1} = \operatorname*{argmin}_{\boldsymbol{\mathcal{E}}} \lambda\|\boldsymbol{\mathcal{E}}\|_{2,1} + \frac{\beta_k}{2}\left\|\boldsymbol{\mathcal{X}} - \boldsymbol{\mathcal{L}}_{k+1} - \boldsymbol{\mathcal{E}} + \frac{\boldsymbol{\mathcal{J}}_k}{\beta_k}\right\|_F^2$$

and obtain its closed form solution:

$$\boldsymbol{\mathcal{E}}(:, i, :)_{k+1} = \begin{cases} \frac{\|\boldsymbol{\mathcal{Q}}^i\|_F - \lambda/\beta_k}{\|\boldsymbol{\mathcal{Q}}^i\|_F}\boldsymbol{\mathcal{Q}}^i, & \text{if } \|\boldsymbol{\mathcal{Q}}^i\|_F \geq \lambda/\beta_k; \\ \mathbf{0}, & \text{otherwise,} \end{cases} \tag{13}$$

where $\boldsymbol{\mathcal{Q}} = \boldsymbol{\mathcal{X}} - \boldsymbol{\mathcal{L}}_{k+1} + \frac{\boldsymbol{\mathcal{J}}_k}{\beta_k}$ and $\boldsymbol{\mathcal{Q}}^i = \boldsymbol{\mathcal{Q}}(:, i, :)$.

Note that our optimization method can be implemented in parallel, since from Eqn. (12), we can observe that at each iteration all lateral slices $\bar{\boldsymbol{L}}_{k+1}^{(i)}$ $(i = 1, \cdots, n_3)$ of $\bar{\boldsymbol{L}}$ can be parallelly updated when updating $\boldsymbol{\mathcal{L}}_{k+1}$. Eqn. (13) also implies that when updating $\boldsymbol{\mathcal{E}}_{k+1}$, its frontal slices $\boldsymbol{\mathcal{E}}_{k+1}(:, i, :)$ $(i = 1, \cdots, n_2)$ can also be parallelly computed.

4. Proofs of Theorem 1

Now we prove Theorem 1 in manuscript. Sec. 4.1 proves the dual conditions of the OR-TPCA problem. Sec. 4.2 provides a way to construct the dual certificates such that dual conditions holds. Sec. 4.3 gives the proofs of some lemmas which are used in Sec. 4.2.

Before we prove Theorem 1, we first prove that any guarantee proved for the Bernoulli distribution equivalently holds for the uniform distribution, which is stated in Lemma 3. According to this conclusion, we will assume that $\Theta \sim \text{Ber}(p)$ in the rest of this document.

Lemma 3. ***(Equivalence of sampling models)*** *If the exact recovery, i.e. exactly recovering the tensor column space and detecting outliers, is guaranteed for the Bernoulli distribution, then it also holds for the uniform distribution. Conversely, if the exact recovery is proved for the uniform distribution, then it also holds for the Bernoulli distribution.*

4.1. Dual Conditions

Before we introduce the dual conditions of OR-TPCA, we first present the subgradient of tensor nuclear norm, which will be used for constructing the dual certificates.

Lemma 4. ***(Subgradient of tensor nuclear norm)*** *[4] Let $\boldsymbol{\mathcal{A}} \in \mathbb{R}^{n_1\times n_2\times n_3}$ with $\text{rank}_t(\boldsymbol{\mathcal{A}}) = r$ and its skinny t-SVD be $\boldsymbol{\mathcal{A}} = \boldsymbol{\mathcal{U}} * \boldsymbol{\mathcal{S}} * \boldsymbol{\mathcal{V}}^*$. The subgradients of $\|\boldsymbol{\mathcal{A}}\|_*$ are $\partial\|\boldsymbol{\mathcal{A}}\|_* = \{\boldsymbol{\mathcal{U}} * \boldsymbol{\mathcal{V}}^* + \boldsymbol{\mathcal{W}} \mid \boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{T}}}(\boldsymbol{\mathcal{W}}) = \mathbf{0}, \|\boldsymbol{\mathcal{W}}\| \leq 1.\}$*

Now we present the dual conditions of OR-TPCA, which is stated in Lemma 5.

Lemma 5. *(Dual conditions of OR-TPCA) Assume that $\text{Range}(\boldsymbol{\mathcal{L}}_0) = \text{Range}(\boldsymbol{\mathcal{P}}_{\Theta_0^\perp}(\boldsymbol{\mathcal{L}}_0))$, $\boldsymbol{\mathcal{E}}_0 \notin \text{Range}(\boldsymbol{\mathcal{L}}_0)$, and $(\widetilde{\boldsymbol{\mathcal{L}}}, \widetilde{\boldsymbol{\mathcal{E}}}) = (\boldsymbol{\mathcal{L}}_0 + \boldsymbol{\mathcal{H}}, \boldsymbol{\mathcal{E}}_0 - \boldsymbol{\mathcal{H}})$ is an arbitrary solution to the OR-TPCA problem. Let $(\boldsymbol{\mathcal{L}}_\star, \boldsymbol{\mathcal{E}}_\star) = (\boldsymbol{\mathcal{L}}_0 + \boldsymbol{\mathcal{P}}_{\Theta_0}\boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{U}}_0}(\boldsymbol{\mathcal{H}}), \boldsymbol{\mathcal{E}}_0 - \boldsymbol{\mathcal{P}}_{\Theta_0}\boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{U}}_0}(\boldsymbol{\mathcal{H}}))$. Suppose that $\|\boldsymbol{\mathcal{P}}_{\Theta_\star}\boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{V}}_\star}\| < 1$, $\lambda > 4\sqrt{\mu_1 r}/\sqrt{n_2 n_3}$, and $\boldsymbol{\mathcal{L}}_\star$ obeys the tensor column-incoherence condition. Then if there are a pair $(\boldsymbol{\mathcal{W}}, \boldsymbol{\mathcal{F}})$ obeying*

$$\boldsymbol{\mathcal{W}} = \lambda\left(\mathcal{B}(\boldsymbol{\mathcal{E}}_\star) + \boldsymbol{\mathcal{F}}\right), \tag{14}$$

with $\boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{V}}_\star}(\boldsymbol{\mathcal{W}}) = 0$, $\|\boldsymbol{\mathcal{W}}\| \leq 1/2$, $\boldsymbol{\mathcal{P}}_{\Theta_\star}(\boldsymbol{\mathcal{F}}) = 0$, and $\|\boldsymbol{\mathcal{F}}\|_{2,\infty} \leq 1/2$, then $\widetilde{\boldsymbol{\mathcal{L}}}$ spans the same tensor column space as that of $\boldsymbol{\mathcal{L}}_0$ and the support set $\widetilde{\Theta}$ of $\widetilde{\boldsymbol{\mathcal{E}}}$ is the same as the support set Θ_0 of $\boldsymbol{\mathcal{E}}_0$.

Proof. The subgradients of tensor nuclear norm and tensor $\ell_{2,1}$ norm can be written as

$$\partial_{\boldsymbol{\mathcal{L}}_\star}\|\boldsymbol{\mathcal{L}}\|_* = \{\boldsymbol{\mathcal{U}}_\star * \boldsymbol{\mathcal{V}}_\star^* + \boldsymbol{\mathcal{W}}_\star, \mid \boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{T}}_\star}(\boldsymbol{\mathcal{W}}_\star) = \mathbf{0}, \ \|\boldsymbol{\mathcal{W}}_\star\| \leq 1\},$$
$$\partial_{\boldsymbol{\mathcal{E}}_\star}\|\boldsymbol{\mathcal{E}}\|_{2,1} = \{\mathcal{B}(\boldsymbol{\mathcal{E}}_\star) + \boldsymbol{\mathcal{H}}_\star, \mid \boldsymbol{\mathcal{P}}_{\Theta_\star}(\boldsymbol{\mathcal{H}}_\star) = \mathbf{0}, \ \|\boldsymbol{\mathcal{H}}_\star\|_{2,\infty} \leq 1\}.$$

We also have

$$\begin{aligned}\|\boldsymbol{\mathcal{U}}_\star * \boldsymbol{\mathcal{V}}_\star^*\|_{2,\infty} &= \max_j \|\boldsymbol{\mathcal{U}}_\star * \boldsymbol{\mathcal{V}}_\star^* * \mathring{\mathfrak{e}}_j\|_F \\ &= \max_j \|\boldsymbol{\mathcal{V}}_\star^* * \mathring{\mathfrak{e}}_j\|_F \\ &\leq \sqrt{\frac{\mu_1 r}{n_2 n_3}}.\end{aligned}$$

Since the tensor nuclear norm and the operator norm are dual, there exists a tensor $\widehat{\mathcal{W}}$ such that $\langle \widehat{\mathcal{W}}, \mathcal{P}_{\mathcal{V}_\star^\perp}\mathcal{P}_{\mathcal{U}_0^\perp}(\mathcal{H})\rangle = \|\mathcal{P}_{\mathcal{V}_\star^\perp}\mathcal{P}_{\mathcal{U}_0^\perp}(\mathcal{H})\|_*$ and $\|\widehat{\mathcal{W}}\| \leq 1$. Then we set $\mathcal{W}_\star = \mathcal{P}_{\mathcal{U}_0^\perp}\mathcal{P}_{\mathcal{V}_\star^\perp}(\widehat{\mathcal{W}})$. In this way, $\mathcal{W}_\star$ obeys $\mathcal{W}_\star \in \mathcal{P}_{\mathcal{T}_\star^\perp}$ since $\mathcal{U}_\star = \mathcal{U}_0$. As $\mathcal{E}_0 \notin \mathrm{Range}(\mathcal{L}_0) = \mathcal{U}_0$, the support set $\Theta_\star$ of $(\mathcal{E}_0 - \mathcal{P}_{\Theta_0}\mathcal{P}_{\mathcal{U}_0}(\mathcal{H}))$ is equal to Θ_0. Then similarly, thanks to the duality between the $\ell_{2,1}$ norm and the $\ell_{2,\infty}$, we can pick a $\mathcal{H}_\star \in \mathcal{P}_{\Theta_\star^\perp}$ such that $\langle \mathcal{H}_\star, \mathcal{P}_{\Theta_\star^\perp}(\mathcal{H})\rangle = \|\mathcal{P}_{\Theta_\star^\perp}(\mathcal{H})\|_{2,1} = \|\mathcal{P}_{\Theta_0^\perp}(\mathcal{H})\|_{2,1}$. On the other hand, we can establish:

$$\begin{aligned}
&\lambda\langle \mathcal{B}(\mathcal{E}_\star), \mathcal{P}_{\mathcal{U}_0^\perp}(\mathcal{H})\rangle \\
=&\langle \mathcal{W} - \lambda\mathcal{F}, \mathcal{P}_{\mathcal{U}_0^\perp}(\mathcal{H})\rangle \\
=&\langle \mathcal{W}, \mathcal{P}_{\mathcal{U}_0^\perp}(\mathcal{H})\rangle - \lambda\langle \mathcal{F}, \mathcal{P}_{\mathcal{U}_0^\perp}(\mathcal{H})\rangle \\
\geq& -\frac{1}{2}\|\mathcal{P}_{\mathcal{V}_\star^\perp}\mathcal{P}_{\mathcal{U}_0^\perp}(\mathcal{H})\|_* - \frac{\lambda}{2}\|\mathcal{P}_{\Theta_\star^\perp}\mathcal{P}_{\mathcal{U}_0^\perp}(\mathcal{H})\|_{2,1},
\end{aligned}$$

where the last inequality holds because we have $\mathcal{W} \in \mathcal{P}_{\mathcal{V}_\star^\perp}$ and $\mathcal{F} \in \mathcal{P}_{\Theta_\star^\perp}$.

Note that both $\mathcal{P}_{\Theta_0}$ and $\mathcal{P}_{\Theta_0^\perp}$ obey $\mathcal{P}_{\Theta_0}\mathcal{P}_{\mathcal{Q}} = \mathcal{P}_{\mathcal{Q}}\mathcal{P}_{\Theta_0}$ and $\mathcal{P}_{\Theta_0^\perp}\mathcal{P}_{\mathcal{Q}} = \mathcal{P}_{\mathcal{Q}}\mathcal{P}_{\Theta_0^\perp}$, where $\mathcal{P}_{\mathcal{Q}}$ can be $\mathcal{P}_{\mathcal{U}_0}, \mathcal{P}_{\mathcal{U}_0^\perp}, \mathcal{P}_{\mathcal{V}_\star}$, *etc.* Define $\widehat{\mathcal{H}} = \mathcal{P}_{\Theta_0^\perp}\mathcal{P}_{\mathcal{U}_0}(\mathcal{H}) + \mathcal{P}_{\Theta_0^\perp}\mathcal{P}_{\mathcal{U}_0^\perp}(\mathcal{H}) + \mathcal{P}_{\Theta_0}\mathcal{P}_{\mathcal{U}_0^\perp}(\mathcal{H})$, we can further obtain

$$\begin{aligned}
&\|\widetilde{\mathcal{L}}\|_* + \lambda\|\widetilde{\mathcal{E}}\|_{2,1} - \|\mathcal{L}_\star\|_* - \lambda\|\mathcal{E}_\star\|_{2,1} \\
\geq&\langle \mathcal{U}_\star * \mathcal{V}_\star^* + \mathcal{W}_\star, \widetilde{\mathcal{L}} - \mathcal{L}_\star\rangle + \lambda\langle \mathcal{B}(\mathcal{E}_\star) + \mathcal{H}_\star, \widetilde{\mathcal{E}} - \mathcal{E}_\star\rangle \\
=&\langle \mathcal{U}_\star * \mathcal{V}_\star^* + \mathcal{W}_\star, \widehat{\mathcal{H}}\rangle - \lambda\langle \mathcal{B}(\mathcal{E}_\star) + \mathcal{H}_\star, \widehat{\mathcal{H}}\rangle \\
=&\langle \mathcal{W}_\star, \mathcal{P}_{\mathcal{U}_0^\perp}(\mathcal{H})\rangle - \lambda\langle \mathcal{H}_\star, \mathcal{P}_{\Theta_0^\perp}(\mathcal{H})\rangle + \langle \mathcal{U}_\star * \mathcal{V}_\star^*, \mathcal{P}_{\Theta_0^\perp}(\mathcal{H})\rangle \\
&- \lambda\langle \mathcal{B}(\mathcal{E}_\star), \mathcal{P}_{\mathcal{U}_0^\perp}(\mathcal{H})\rangle \\
\geq&\|\mathcal{P}_{\mathcal{V}_\star^\perp}\mathcal{P}_{\mathcal{U}_0^\perp}(\mathcal{H})\|_* + \lambda\|\mathcal{P}_{\Theta_\star^\perp}(\mathcal{H})\|_{2,1} - \sqrt{\frac{\mu_1 r}{n_2 n_3}}\|\mathcal{P}_{\Theta_0^\perp}(\mathcal{H})\|_{2,1} \\
&- \frac{1}{2}\|\mathcal{P}_{\mathcal{V}_\star^\perp}\mathcal{P}_{\mathcal{U}_0^\perp}(\mathcal{H})\|_* - \frac{\lambda}{2}\|\mathcal{P}_{\Theta_\star^\perp}\mathcal{P}_{\mathcal{U}_0^\perp}(\mathcal{H})\|_{2,1} \\
=&\frac{1}{2}\|\mathcal{P}_{\mathcal{V}_\star^\perp}\mathcal{P}_{\mathcal{U}_0^\perp}(\mathcal{H})\|_* + \left(\frac{\lambda}{4} - \sqrt{\frac{\mu_1 r}{n_2 n_3}}\right)\|\mathcal{P}_{\Theta_0^\perp}(\mathcal{H})\|_{2,1} \\
&+ \frac{3\lambda}{4}\|\mathcal{P}_{\Theta_\star^\perp}(\mathcal{H})\|_{2,1} - \frac{\lambda}{2}\|\mathcal{P}_{\Theta_\star^\perp}\mathcal{P}_{\mathcal{U}_0^\perp}(\mathcal{H})\|_{2,1} \\
\geq&\frac{1}{2}\|\mathcal{P}_{\mathcal{V}_\star^\perp}\mathcal{P}_{\mathcal{U}_0^\perp}(\mathcal{H})\|_* + \left(\frac{\lambda}{4} - \sqrt{\frac{\mu_1 r}{n_2 n_3}}\right)\|\mathcal{P}_{\Theta_0^\perp}(\mathcal{H})\|_{2,1} \\
&+ \frac{\lambda}{4}\|\mathcal{P}_{\Theta_\star^\perp}\mathcal{P}_{\mathcal{U}_0^\perp}(\mathcal{H})\|_{2,1}.
\end{aligned}$$

As we have $\lambda > 4\sqrt{\mu_1 r}/\sqrt{n_2 n_3}$ and $(\widetilde{\mathcal{L}}, \widetilde{\mathcal{E}})$ is an arbitrary optimal solution, we can obtain $\|\mathcal{P}_{\mathcal{V}_\star^\perp}\mathcal{P}_{\mathcal{U}_0^\perp}(\mathcal{H})\|_* = \|\mathcal{P}_{\Theta_\star^\perp}\mathcal{P}_{\mathcal{U}_0^\perp}(\mathcal{H})\|_{2,1} = \|\mathcal{P}_{\Theta_0^\perp}(\mathcal{H})\|_{2,1} = 0$. Therefore, we have $\mathcal{H} \in \Theta_0$ and $\mathcal{P}_{\mathcal{U}_0^\perp}(\mathcal{H}) \in \mathcal{V}_\star \cap \Theta_\star = \mathbf{0}$, due to $\|\mathcal{P}_{\Theta_\star}\mathcal{P}_{\mathcal{V}_\star}\| < 1$. In this way, we have $\mathcal{H} \in \mathcal{U}_0$ and thus $\mathcal{H} \in \Theta_0 \cap \mathcal{U}_0$, which further demonstrates that $\widetilde{\mathcal{U}} \subseteq \mathcal{U}_0$ and $\widetilde{\Theta} \subseteq \Theta_0$.

On the other hand, since $\mathcal{H} \in \Theta_0 \cap \mathcal{U}_0$ and $\mathrm{Range}(\mathcal{L}_0) = \mathrm{Range}(\mathcal{P}_{\Theta_0^\perp}(\mathcal{L}_0))$, $\widetilde{\mathcal{U}} = \mathrm{Range}(\widetilde{\mathcal{L}}) = \mathrm{Range}(\mathcal{L}_0 + \mathcal{H}) = \mathrm{Range}(\mathcal{L}_0) = \mathcal{U}_0$ holds. Now if $\widetilde{\Theta} \neq \Theta_0$, then there exits an $i \in (\Theta_0 \cap \widetilde{\Theta}^\perp)$ such that $\widetilde{\mathcal{E}}(:, i, :) = \mathbf{0}$. Accordingly, we have $\widetilde{\mathcal{L}}(:, i, :) = \mathcal{L}_0(:\, i, :) + \mathcal{E}_0(:, i, :) \notin \mathcal{U}_0$. But, we have $\widetilde{\mathcal{L}}(:, i, :) \in \widetilde{\mathcal{U}} = \mathcal{U}_0$. Hence, we can obtain $\widetilde{\Theta} = \Theta_0$. $\square$

Lemma 5 implies that if we can find a dual certificate $\mathcal{W}$ obeying

$$\begin{aligned}
&\text{(a)} \quad \mathcal{W} \in \mathcal{V}_\star^\perp, \\
&\text{(b)} \quad \mathcal{P}_{\Theta_\star}(\mathcal{W}) = \lambda\mathcal{B}(\mathcal{E}_\star), \\
&\text{(c)} \quad \|\mathcal{W}\| \leq 1/2, \\
&\text{(d)} \quad \|\mathcal{P}_{\Theta_\star^\perp}(\mathcal{W})\|_{2,\infty} \leq \lambda/2,
\end{aligned} \tag{15}$$

then we can exactly recover the tensor column space $\mathcal{U}_0$ of $\mathcal{L}_0$ and the support set Θ_0 of outliers $\mathcal{E}_0$.

4.2. Dual Certification via Least Squares

Before we construct the dual certificate $\mathcal{W}$, we first give some key lemmas, which will be proved in Sec. 4.3.

Lemma 6. *Assume $\widehat{\Theta} \sim Ber(\kappa)$. Then with high probability,*

$$\left\|\mathcal{P}_{\mathcal{V}_\star} - \frac{1}{\kappa}\mathcal{P}_{\mathcal{V}_\star}\mathcal{P}_{\widehat{\Theta}}\mathcal{P}_{\mathcal{V}_\star}\right\| \leq \epsilon,$$

provided that $\kappa \geq c_3\mu_1 r\log(n_{(1)})/(\epsilon^2 n_2)$ for some numerical constant $c_3 > 0$.

Corollary 7. *Suppose $\Theta_\star \sim Ber(\rho)$. Then with high probability,*

$$\|\mathcal{P}_{\Theta_\star}\mathcal{P}_{\mathcal{V}_\star}\|^2 < (1-\rho)\epsilon + \rho < 1,$$

provided that $1 - \rho \geq c_3\mu_1 r\log(n_{(1)})/(\epsilon^2 n_2)$ for some numerical constant $c_3 > 0$.

Now we construct the dual certificate $\mathcal{W}$ and verify its validity.

Lemma 8. *Suppose that $\Theta_\star \sim Ber(\rho)$ and the assumptions of Theorem 1 (in manuscript) are satisfied. Then with high probability,*

$$\mathcal{W} = \lambda\mathcal{P}_{\mathcal{V}_\star^\perp}\sum_{k=0}^{+\infty}(\mathcal{P}_{\Theta_\star}\mathcal{P}_{\mathcal{V}_\star}\mathcal{P}_{\Theta_\star})^k(\mathcal{B}(\mathcal{E}_\star))$$

obeys the dual conditions (15).

Proof. Note that by Corollary 7, we have $\|\mathcal{P}_{\Theta_\star}\mathcal{P}_{\mathcal{V}_\star}\mathcal{P}_{\Theta_\star}\| = \|\mathcal{P}_{\mathcal{V}_\star}\mathcal{P}_{\Theta_\star}\|^2 < 1$. Thus, $\mathcal{W}$ is well defined. Since we use a smaller space $\mathcal{V}_\star \subset \mathcal{T}_\star$ instead of $\mathcal{T}_\star$ to construct $\mathcal{W}$, we can avoid the case that $\Theta_\star \cap \mathcal{T}_\star \neq \mathbf{0}$, *i.e.*, $\|\mathcal{P}_{\Theta_\star}\mathcal{P}_{\mathcal{T}_\star}\mathcal{P}_{\Theta_\star}\| = \|\mathcal{P}_{\mathcal{T}_\star}\mathcal{P}_{\Theta_\star}\|^2 = 1$, leading to a divergent Neumann series $\sum_{k=0}^{+\infty}(\mathcal{P}_{\Theta_\star}\mathcal{P}_{\mathcal{T}_\star}\mathcal{P}_{\Theta_\star})^k$. Now we verify the conditions in (15) in turn.

Proof of (15) **(a):** It is easy to verify that $\boldsymbol{\mathcal{W}} \in \boldsymbol{\mathcal{V}}_\star^\perp$.

Proof of (15) **(b):** By the construction of $\boldsymbol{\mathcal{W}}$, we have

$$\begin{aligned}
&\boldsymbol{\mathcal{P}}_{\Theta_\star}(\boldsymbol{\mathcal{W}})\\
=&\lambda \boldsymbol{\mathcal{P}}_{\Theta_\star}\boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{V}}_\star^\perp}\sum_{k=0}^{+\infty}(\boldsymbol{\mathcal{P}}_{\Theta_\star}\boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{V}}_\star}\boldsymbol{\mathcal{P}}_{\Theta_\star})^k(\mathcal{B}(\boldsymbol{\mathcal{E}}_\star))\\
=&\lambda \boldsymbol{\mathcal{P}}_{\Theta_\star}(\boldsymbol{\mathcal{I}}-\boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{V}}_\star})\sum_{k=0}^{+\infty}(\boldsymbol{\mathcal{P}}_{\Theta_\star}\boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{V}}_\star}\boldsymbol{\mathcal{P}}_{\Theta_\star})^k(\mathcal{B}(\boldsymbol{\mathcal{E}}_\star))\\
=&\lambda \boldsymbol{\mathcal{P}}_{\Theta_\star}(\boldsymbol{\mathcal{I}}-\boldsymbol{\mathcal{P}}_{\Theta_\star}\boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{V}}_\star}\boldsymbol{\mathcal{P}}_{\Theta_\star})\sum_{k=0}^{+\infty}(\boldsymbol{\mathcal{P}}_{\Theta_\star}\boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{V}}_\star}\boldsymbol{\mathcal{P}}_{\Theta_\star})^k(\mathcal{B}(\boldsymbol{\mathcal{E}}_\star))\\
=&\lambda \boldsymbol{\mathcal{P}}_{\Theta_\star}\left(\sum_{k=0}^{+\infty}(\boldsymbol{\mathcal{P}}_{\Theta_\star}\boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{V}}_\star}\boldsymbol{\mathcal{P}}_{\Theta_\star})^k-\sum_{k=1}^{+\infty}(\boldsymbol{\mathcal{P}}_{\Theta_\star}\boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{V}}_\star}\boldsymbol{\mathcal{P}}_{\Theta_\star})^k\right)(\mathcal{B}(\boldsymbol{\mathcal{E}}_\star))\\
=&\lambda \boldsymbol{\mathcal{P}}_{\Theta_\star}(\mathcal{B}(\boldsymbol{\mathcal{E}}_\star))\\
=&\lambda \mathcal{B}(\boldsymbol{\mathcal{E}}_\star).
\end{aligned}$$

Thus, $\boldsymbol{\mathcal{W}}$ obeys the condition (15) (b).

Proof of (15) **(c):** Let $\boldsymbol{\mathcal{G}} = \sum_{k=0}^{+\infty}(\boldsymbol{\mathcal{P}}_{\Theta_\star}\boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{V}}_\star}\boldsymbol{\mathcal{P}}_{\Theta_\star})^k$. Since $\|\mathcal{B}(\boldsymbol{\mathcal{E}})\| \leq \sqrt{\log{(n_2)}}/4$, then we have

$$\|\boldsymbol{\mathcal{W}}\| \leq \lambda \|\boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{V}}_\star^\perp}\|\|\boldsymbol{\mathcal{G}}\|\|\mathcal{B}(\boldsymbol{\mathcal{E}}_\star)\| = \lambda\frac{1}{1-\sigma^2}\frac{\sqrt{\log{(n_2)}}}{4},$$

where $\sigma = \sqrt{\rho+\epsilon(1-\rho)}$. If $\lambda \leq 2(1-\sigma^2)/\sqrt{\log{(n_2)}}$, then $\|\boldsymbol{\mathcal{W}}\| \leq 1/2$. Note that in Lemma 5, we require $\lambda > 4\sqrt{\frac{\mu_1 r}{n_2 n_3}}$. Thus, $\lambda \in \left(4\sqrt{\frac{\mu_1 r}{n_2 n_3}}, \frac{2(1-\sigma^2)}{\sqrt{\log{(n_2)}}}\right]$.

Proof of (15) **(d):** $\boldsymbol{\mathcal{G}}$ and σ are defined as the same as above. Thus, we can obtain

$$\begin{aligned}
\boldsymbol{\mathcal{P}}_{\Theta_\star^\perp}(\boldsymbol{\mathcal{W}}) =&\lambda \boldsymbol{\mathcal{P}}_{\Theta_\star^\perp}\boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{V}}_\star^\perp}\boldsymbol{\mathcal{G}}(\mathcal{B}(\boldsymbol{\mathcal{E}}_\star))\\
=&\lambda \boldsymbol{\mathcal{P}}_{\Theta_\star^\perp}(\boldsymbol{\mathcal{I}}-\boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{V}}_\star})\boldsymbol{\mathcal{G}}(\mathcal{B}(\boldsymbol{\mathcal{E}}_\star))\\
=&-\lambda \boldsymbol{\mathcal{P}}_{\Theta_\star^\perp}\boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{V}}_\star}\boldsymbol{\mathcal{G}}(\mathcal{B}(\boldsymbol{\mathcal{E}}_\star)).
\end{aligned}$$

We first prove an inequality:

$$\begin{aligned}
\max_j\sum_{i,k}\|\boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{V}}_\star}(\mathfrak{e}_{ijk})\|_F^2 =&\max_j\sum_{i,k}\|\mathring{\mathfrak{e}}_i * \dot{\mathfrak{e}}_k * \mathring{\mathfrak{e}}_j^* * \boldsymbol{\mathcal{V}}_\star * \boldsymbol{\mathcal{V}}_\star^*\|_F^2\\
=&\max_j\sum_{i,k}\|\mathring{\mathfrak{e}}_i * \dot{\mathfrak{e}}_k * \mathring{\mathfrak{e}}_j^* * \boldsymbol{\mathcal{V}}_\star\|_F^2\\
=&\max_j\sum_{i,k}\|(\dot{\mathfrak{e}}_k * \mathring{\mathfrak{e}}_j^* * \boldsymbol{\mathcal{V}}_\star)(i,:,:)\|_F^2\\
=&\max_j\sum_{k}\|\dot{\mathfrak{e}}_k * \mathring{\mathfrak{e}}_j^* * \boldsymbol{\mathcal{V}}_\star\|_F^2\\
=&\max_j n_3\|\mathring{\mathfrak{e}}_j^* * \boldsymbol{\mathcal{V}}_\star\|_F^2\\
\leq& n_3\frac{\mu_1 r}{n_2 n_3}\\
=&\frac{\mu_1 r}{n_2},
\end{aligned}$$

where the third equality holds because $\boldsymbol{\mathcal{Q}}(i,:,:) = \mathring{\mathfrak{e}}_i * \boldsymbol{\mathcal{Q}}$, where $\boldsymbol{\mathcal{Q}} = \dot{\mathfrak{e}}_k * \mathring{\mathfrak{e}}_j^* * \boldsymbol{\mathcal{V}}_\star$, and the fifth equality holds because of $\|\boldsymbol{\mathcal{Q}}\|_F = \|\mathring{\mathfrak{e}}_j^* * \boldsymbol{\mathcal{V}}_\star\|_F$ since $\dot{\mathfrak{e}}_k$ does not change the values of the entries in $\mathring{\mathfrak{e}}_j^* * \boldsymbol{\mathcal{V}}_\star$ but exchanges the positions of entries.

Let $\boldsymbol{\mathcal{Q}} = \boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{V}}_\star}\boldsymbol{\mathcal{G}}(\mathcal{B}(\boldsymbol{\mathcal{E}}_\star))$. Then, we can obtain

$$\begin{aligned}
\|\boldsymbol{\mathcal{Q}}\|_{2,\infty}^2 =&\max_b\sum_{i,k}\langle \boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{V}}_\star}\boldsymbol{\mathcal{G}}(\mathcal{B}(\boldsymbol{\mathcal{E}}_\star)), \mathfrak{e}_{ibk}\rangle^2\\
=&\max_b\sum_{i,k}\langle \mathcal{B}(\boldsymbol{\mathcal{E}}_\star), \boldsymbol{\mathcal{G}}\boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{V}}_\star}(\mathfrak{e}_{ibk})\rangle^2\\
=&\max_b\sum_{i,k}\sum_{j}\langle \mathcal{B}(\boldsymbol{\mathcal{E}}_\star) * \mathring{\mathfrak{e}}_j, \boldsymbol{\mathcal{G}}\boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{V}}_\star}(\mathfrak{e}_{ibk}) * \mathring{\mathfrak{e}}_j\rangle^2\\
\leq&\max_b\sum_{i,j,k}\|\mathcal{B}(\boldsymbol{\mathcal{E}}_\star) * \mathring{\mathfrak{e}}_j\|_F^2\|\boldsymbol{\mathcal{G}}\boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{V}}_\star}(\mathfrak{e}_{ibk}) * \mathring{\mathfrak{e}}_j\|_F^2\\
\leq&\max_b\sum_{i,j,k}\|\boldsymbol{\mathcal{G}}\boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{V}}_\star}(\mathfrak{e}_{ibk}) * \mathring{\mathfrak{e}}_j\|_F^2\\
=&\max_b\sum_{i,k}\|\boldsymbol{\mathcal{G}}\boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{V}}_\star}(\mathfrak{e}_{ibk})\|_F^2\\
\leq&\max_b\sum_{i,k}\|\boldsymbol{\mathcal{G}}\|^2\|\boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{V}}_\star}(\mathfrak{e}_{ibk})\|_F^2\\
\leq&\frac{\mu_1 r}{n_2}(\frac{1}{1-\sigma^2})^2\\
\leq&\frac{1}{4},
\end{aligned}$$

where the second inequality holds since $\mathcal{B}(\boldsymbol{\mathcal{E}}_\star) * \mathring{\mathfrak{e}}_j = (\mathcal{B}(\boldsymbol{\mathcal{E}}_\star))(:,j,:)$ and the last inequality holds because we require $r \leq n_2(1-\sigma^2)^2/(4\mu_1)$. Note that when proving Corollary 7, we demand $r \leq (1-\rho)\epsilon^2 n_2/(c_3\mu_1\log{(n_{(1)})})$. Thus, we can further establish:

$$\begin{aligned}
\|\boldsymbol{\mathcal{P}}_{\Theta_\star^\perp}\boldsymbol{\mathcal{W}}\|_{2,\infty} =&\lambda\|\boldsymbol{\mathcal{P}}_{\Theta_\star^\perp}\boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{V}}_\star^\perp}\boldsymbol{\mathcal{G}}(\mathcal{B}(\boldsymbol{\mathcal{E}}_\star))\|_{2,\infty}\\
\leq&\lambda\|\boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{V}}_\star^\perp}\boldsymbol{\mathcal{G}}(\mathcal{B}(\boldsymbol{\mathcal{E}}_\star))\|_{2,\infty}\\
=&\frac{1}{2}\lambda.
\end{aligned}$$

So $\boldsymbol{\mathcal{W}}$ obeys the condition (15) (d).

Checking the ranges of λ and r: When we prove the above conclusions, we require

$$\lambda \in \left(4\sqrt{\frac{\mu_1 r}{n_2 n_3}}, \frac{2(1-\sigma^2)}{\sqrt{\log{(n_2)}}}\right],$$

and

$$r \leq \min\left(\frac{n_2(1-\sigma^2)^2}{4\mu_1}, \frac{(1-\rho)\epsilon^2 n_2}{c_3\mu_1\log{(n_{(1)})}}\right).$$

Thus, we have

$$r \leq \frac{\rho_r n_2}{\mu_1 \log{(n_{(1)})}},$$

where ρ_r is a constant. On the other hand, let $\rho \leq 0.5-\epsilon$, then we have $2(1-\sigma^2) \geq 1$. Accordingly, we can further obtain

$$\lambda \in \left[\epsilon\sqrt{\frac{8\sqrt{2}}{c_3 n_3}}\frac{1}{\sqrt{\log(n_{(1)})}}, \frac{1}{\sqrt{\log{(n_2)}}}\right],$$

Since ϵ is very small and c_3 is a constant, we have $\epsilon\sqrt{8\sqrt{2}/(c_3 n_3)} \leq 1$. Moreover, we have $n_{(1)} = \max(n_1, n_2) \geq n_2$. So we can just set $\lambda = 1/\sqrt{\log(n_2)}$. The proof is completed. $\square$

4.3. Proofs of Some Lemmas

We first prove Lemma 9 which will be used in the proofs of Lemma 3. Lemma 9 proves that the success of exact recovery is monotone on the outlier number. That is, if there are q outliers and OR-TPCA can exactly recover the tensor column space and detect q outliers, then after removing some outliers from the tensor data, OR-TPCA still can recover the desired tensor column space and detect the remaining outliers.

Lemma 9. *(**Elimination Lemma**) Assume that any optimal solution $(\mathcal{L}, \mathcal{E})$ to the OR-TPCA problem with input tensor $\mathcal{X} = \mathcal{L}_0 + \mathcal{E}_0$ can exactly recover the tensor column space Range$(\mathcal{L}_0)$ and the outlier support Θ_0. Then any optimal solution $(\widetilde{\mathcal{L}}, \widetilde{\mathcal{E}})$ to the OR-TPCA problem with input $\widetilde{\mathcal{X}} = \mathcal{L}_0 + \mathcal{P}_\Theta(\mathcal{E}_0)$ also can exactly recover Range$(\mathcal{L}_0)$ and its outlier support set Θ, where $\Theta \subseteq \Theta_0$.*

Then, we introduce another lemma which is used for proving Lemma 6.

Lemma 10. *(Matrix (Operator) Bernstein Inequality) [5] 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 $\max\left(\|\sum_i \mathbb{E}[\boldsymbol{X}_i \boldsymbol{X}_i^*]\|, \|\sum_i \mathbb{E}[\boldsymbol{X}_i^* \boldsymbol{X}_i]\|\right) \leq \omega$. 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

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

4.3.1 Proof of Lemma 3

Proof. We use $\mathbb{P}_{\text{Unif}(m)}$ and $\mathbb{P}_{\text{Ber}(p)}$ to denote the probabilities calculated under the uniform and Bernoulli models, respectively. Let "$\texttt{Success}$" be the event that the algorithm succeeds, *i.e.* exactly recovering the tensor column space Range$(\mathcal{L}_0)$ and the outlier support Θ_0 of $\mathcal{E}_0$.

By Lemma 9, we know that the success of exactly recovery is monotone on the number of outliers. That is, for any $k \leq m$, $\mathbb{P}_{\text{Unif}(k)}(\texttt{Success}) \geq \mathbb{P}_{\text{Unif}(m)}(\texttt{Success})$. On the other hand, we have

$$\mathbb{P}_{\text{Ber}(p)}(\texttt{Success} \mid |\Theta| = k) = \mathbb{P}_{\text{Unif}(k)}(\texttt{Success}).$$

By using the above two conclusions, we can further obtain

$$\begin{aligned}
&\mathbb{P}_{\text{Ber}(p)}(\texttt{Success})\\
=&\sum_{k=0}^{n_2} \mathbb{P}_{\text{Ber}(p)}(\texttt{Success} \mid |\Theta| = k)\mathbb{P}_{\text{Ber}(p)}(|\Theta| = k)\\
\leq&\sum_{k=0}^{m-1} \mathbb{P}_{\text{Ber}(p)}(|\Theta| = k) + \sum_{k=m}^{n_2} \mathbb{P}_{\text{Unif}(k)}(\texttt{Success})\mathbb{P}_{\text{Ber}(p)}(|\Theta| = k)\\
\leq&\mathbb{P}_{\text{Ber}(p)}(|\Theta| < m) + \mathbb{P}_{\text{Unif}(m)}(\texttt{Success}).
\end{aligned}$$

Let $p = m/n_2 + \epsilon$, where $\epsilon > 0$ is a constant. Then the conclusion follows from $\mathbb{P}_{\text{Ber}(p)}(|\Theta| < m) \leq \exp\left(-\epsilon^2 n_2/(2p)\right)$. In a similar way, we have

$$\begin{aligned}
&\mathbb{P}_{\text{Ber}(p)}(\texttt{Success})\\
=&\sum_{k=0}^{n_2} \mathbb{P}_{\text{Ber}(p)}(\texttt{Success} \mid |\Theta| = k)\mathbb{P}_{\text{Ber}(p)}(|\Theta| = k)\\
=&\sum_{k=0}^{n_2} \mathbb{P}_{\text{Unif}(k)}(\texttt{Success})\mathbb{P}_{\text{Ber}(p)}(|\Theta| = k)\\
\geq&\sum_{k=0}^{m} \mathbb{P}_{\text{Unif}(k)}(\texttt{Success})\mathbb{P}_{\text{Ber}(p)}(|\Theta| = k)\\
\geq&\mathbb{P}_{\text{Unif}(m)}(\texttt{Success})\sum_{k=0}^{m} \mathbb{P}_{\text{Ber}(p)}(|\Theta| = k)\\
\geq&\mathbb{P}_{\text{Unif}(m)}(\texttt{Success})\mathbb{P}_{\text{Ber}(p)}(|\Theta| \leq m).
\end{aligned}$$

By choosing an appropriate m such that $\mathbb{P}_{\text{Ber}(p)}(|\Theta| > m)$ is exponentially small. The proof is completed. $\square$

4.3.2 Proof of Lemma 6

Proof. Define a set $\phi = \{(\mathcal{Z}_1, \mathcal{Z}_2) \mid \|\mathcal{Z}_1\|_F \leq 1, \ \mathcal{Z}_2 = \pm \mathcal{Z}_1\}$. Since $\left(\kappa^{-1}\mathcal{P}_{\mathcal{V}_\star}\mathcal{P}_{\widehat{\Theta}}\mathcal{P}_{\mathcal{V}_\star} - \mathcal{P}_{\mathcal{V}_\star}\right)$ is a self-adjoint operator, we can use the variational form of the operator norm to compute its operator norm:

$$\begin{aligned}
&\left\|\frac{1}{\kappa}\mathcal{P}_{\mathcal{V}_\star}\mathcal{P}_{\widehat{\Theta}}\mathcal{P}_{\mathcal{V}_\star} - \mathcal{P}_{\mathcal{V}_\star}\right\|\\
=&\sup_{\phi} \sum_{i,j,k}\left(\frac{\delta_j}{\kappa} - 1\right)\left\langle \mathcal{P}_{\mathcal{V}_\star}(\mathcal{Z}_1), \mathfrak{e}_{ijk}\right\rangle \left\langle \mathcal{P}_{\mathcal{V}_\star}(\mathcal{Z}_2), \mathfrak{e}_{ijk}\right\rangle\\
=&\sup_{\phi} \sum_{i,j,k}\frac{\delta_j - \kappa}{\kappa n_3^2}\left\langle \texttt{bdiag}\left(\overline{\mathcal{P}_{\mathcal{V}_\star}(\mathfrak{e}_{ijk})}\right), \bar{\mathcal{Z}}_1\right\rangle \left\langle \texttt{bdiag}\left(\overline{\mathcal{P}_{\mathcal{V}_\star}(\mathfrak{e}_{ijk})}\right), \bar{\mathcal{Z}}_2\right\rangle\\
=&\left\|\sum_{i,j,k}\frac{\delta_j - \kappa}{\kappa n_3}\texttt{bdiag}\left(\overline{\mathcal{P}_{\mathcal{V}_\star}(\mathfrak{e}_{ijk})}\right) \otimes \texttt{bdiag}\left(\overline{\mathcal{P}_{\mathcal{V}_\star}(\mathfrak{e}_{ijk})}\right)\right\|\\
=&\left\|\sum_{j}\frac{\delta_j - \kappa}{\kappa n_3}\sum_{i,k}\texttt{bdiag}\left(\overline{\mathcal{P}_{\mathcal{V}_\star}(\mathfrak{e}_{ijk})}\right) \otimes \texttt{bdiag}\left(\overline{\mathcal{P}_{\mathcal{V}_\star}(\mathfrak{e}_{ijk})}\right)\right\|,
\end{aligned}$$

where δ_j obeys *i.i.d.* Bernoulli distribution Ber(κ) and $\otimes$ denotes the tensor product. Note that the fourth equality holds because $\sum_{i,j,k} \frac{\delta_j-\kappa}{\kappa n_3}\texttt{bdiag}\left(\overline{\mathcal{P}_{\mathcal{V}_\star}(\mathfrak{e}_{ijk})}\right) \otimes \texttt{bdiag}\left(\overline{\mathcal{P}_{\mathcal{V}_\star}(\mathfrak{e}_{ijk})}\right)$ is a self-adjoint matrix operator and its operator norm is defined on a set $\varphi = \{(\bar{\boldsymbol{D}}_1, \bar{\boldsymbol{D}}_2) \mid \|\bar{\boldsymbol{D}}_1\|_F \leq 1, \bar{\boldsymbol{D}}_2 = \pm\bar{\boldsymbol{D}}_1\}$ instead of ϕ which leads to a factor n_3. Note that $\texttt{bdiag}\left(\overline{\mathcal{P}_{\mathcal{V}_\star}(\mathfrak{e}_{ijk})}\right)$ is a diagonal matrix and thus $\bar{\boldsymbol{D}}_1$ and $\bar{\boldsymbol{D}}_2$ are also diagonal matrices and there exist two tensors $\boldsymbol{\mathcal{D}}_1$ and $\boldsymbol{\mathcal{D}}_2$ such that $\bar{\boldsymbol{D}}_1 = \texttt{bdiag}(\texttt{fft}(\boldsymbol{\mathcal{D}}_1, [], 3))$ and $\bar{\boldsymbol{D}}_2 = \texttt{bdiag}(\texttt{fft}(\boldsymbol{\mathcal{D}}_2, [], 3))$. For brevity, we further define

$$\bar{\boldsymbol{H}}_j = \frac{\delta_j-\kappa}{\kappa n_3}\sum_{i,k}\texttt{bdiag}\left(\overline{\mathcal{P}_{\mathcal{V}_\star}(\mathfrak{e}_{ijk})}\right) \otimes \texttt{bdiag}\left(\overline{\mathcal{P}_{\mathcal{V}_\star}(\mathfrak{e}_{ijk})}\right),$$

and have

$$\left\|\frac{1}{\kappa}\mathcal{P}_{\mathcal{V}_\star}\mathcal{P}_{\widehat{\Theta}}\mathcal{P}_{\mathcal{V}_\star} - \mathcal{P}_{\mathcal{V}_\star}\right\| = \left\|\sum_j \bar{\boldsymbol{H}}_j\right\|.$$

By the above definitions, we have $\bar{\boldsymbol{H}}_j$ is self-adjoint and $\mathbb{E}[\bar{\boldsymbol{H}}_j] = \boldsymbol{0}$. To prove the conclusions by the matrix Bernstein inequality, we need to bound $\|\bar{\boldsymbol{H}}_j\|$ and $\left\|\sum_j \mathbb{E}[\bar{\boldsymbol{H}}_j\bar{\boldsymbol{H}}_j^*]\right\|$. We first prove a vital inequality:

$$\begin{aligned}
&\|\mathcal{P}_{\mathcal{V}_\star}(\boldsymbol{\mathcal{Z}})\|_{2,\infty}^2 \\
=&\max_j \sum_{i,k}\langle\boldsymbol{\mathcal{Z}}, \mathcal{P}_{\mathcal{V}_\star}(\mathfrak{e}_{ijk})\rangle^2 \\
=&\max_j \sum_{i,k}\langle\boldsymbol{\mathcal{Z}}, \mathring{\mathfrak{e}}_i * \dot{\mathfrak{e}}_k * \mathring{\mathfrak{e}}_j^* * \boldsymbol{\mathcal{V}}_\star * \boldsymbol{\mathcal{V}}_\star^*\rangle^2 \\
=&\max_j \sum_{i,k}\langle\dot{\mathfrak{e}}_k^* * \mathring{\mathfrak{e}}_i^* * \boldsymbol{\mathcal{Z}}, \mathring{\mathfrak{e}}_j^* * \boldsymbol{\mathcal{V}}_\star * \boldsymbol{\mathcal{V}}_\star^*\rangle^2 \\
\leq&\max_j \sum_{i,k}\|\dot{\mathfrak{e}}_k^* * \mathring{\mathfrak{e}}_i^* * \boldsymbol{\mathcal{Z}}\|_F^2\|\mathring{\mathfrak{e}}_j^* * \boldsymbol{\mathcal{V}}_\star * \boldsymbol{\mathcal{V}}_\star^*\|_F^2 \\
=&\max_j \sum_{i,k}\|\boldsymbol{\mathcal{Z}}(i,:,:)\|_F^2\|\mathring{\mathfrak{e}}_j^* * \boldsymbol{\mathcal{V}}_\star\|_F^2 \\
=&\max_j \sum_{i} n_3\|\boldsymbol{\mathcal{Z}}(i,:,:)\|_F^2\|\mathring{\mathfrak{e}}_j^* * \boldsymbol{\mathcal{V}}_\star\|_F^2 \\
=&n_3\sum_{i}\|\boldsymbol{\mathcal{Z}}(i,:,:)\|_F^2 \max_j\|\mathring{\mathfrak{e}}_j^* * \boldsymbol{\mathcal{V}}_\star\|_F^2 \\
\leq&n_3\|\boldsymbol{\mathcal{Z}}\|_F^2\frac{\mu_1 r}{n_2 n_3} \\
=&\frac{\mu_1 r}{n_2}\|\boldsymbol{\mathcal{Z}}\|_F^2,
\end{aligned}$$

where the fourth equality holds because $\boldsymbol{\mathcal{Z}}(i,:,:) = \mathring{\mathfrak{e}}_i^* * \boldsymbol{\mathcal{Z}}$ and if let $\boldsymbol{\mathcal{Q}} = \dot{\mathfrak{e}}_k^* * \boldsymbol{\mathcal{Z}}(i,:,:)$, then $\|\boldsymbol{\mathcal{Q}}\|_F = \|\boldsymbol{\mathcal{Z}}(i,:,:)\|_F$ since $\dot{\mathfrak{e}}_k$ does not change the values of the entries in $\boldsymbol{\mathcal{Z}}(i,:,:)$ but exchanges the positions of the entries. Then, by utilizing the above inequality, we can bound $\|\bar{\boldsymbol{H}}_j\|$:

$$\begin{aligned}
&\|\bar{\boldsymbol{H}}_j\| \\
=&\sup_\varphi\frac{\delta_j-\kappa}{\kappa n_3}\sum_{i,k}\left\langle\texttt{bdiag}\left(\overline{\mathcal{P}_{\mathcal{V}_\star}(\mathfrak{e}_{ijk})}\right), \bar{\boldsymbol{D}}_1\right\rangle\left\langle\texttt{bdiag}\left(\overline{\mathcal{P}_{\mathcal{V}_\star}(\mathfrak{e}_{ijk})}\right), \bar{\boldsymbol{D}}_2\right\rangle \\
\leq&\sup_\varphi\frac{n_3}{\kappa}|\delta_j-\kappa|\sum_{i,k}|\langle\boldsymbol{\mathcal{D}}_1, \mathcal{P}_{\mathcal{V}_\star}(\mathfrak{e}_{ijk})\rangle||\langle\boldsymbol{\mathcal{D}}_2, \mathcal{P}_{\mathcal{V}_\star}(\mathfrak{e}_{ijk})\rangle| \\
=&\sup_\varphi\frac{n_3}{\kappa}|\delta_j-\kappa|\sum_{i,k}|\langle\mathcal{P}_{\mathcal{V}_\star}(\boldsymbol{\mathcal{D}}_1), \mathfrak{e}_{ijk}\rangle|^2 \\
\leq&\sup_\varphi\frac{n_3}{\kappa}\sum_{i,k}\langle\mathcal{P}_{\mathcal{V}_\star}(\boldsymbol{\mathcal{D}}_1), \mathfrak{e}_{ijk}\rangle^2 \\
\leq&\sup_\varphi\frac{n_3}{\kappa}\|\mathcal{P}_{\mathcal{V}_\star}(\boldsymbol{\mathcal{D}}_1)\|_{2,\infty}^2 \\
\leq&\sup_\varphi\frac{n_3}{\kappa}\frac{\mu_1 r}{n_2}\|\boldsymbol{\mathcal{D}}_1\|_F^2 \\
\leq&\frac{\mu_1 r}{\kappa n_2} := \nu.
\end{aligned}$$

We can also establish:

$$\begin{aligned}
&\|\sum_j\mathbb{E}[\bar{\boldsymbol{H}}_j\bar{\boldsymbol{H}}_j^*]\| \\
=&\sup_\varphi\sum_j\frac{\mathbb{E}\left[(\delta_j-\kappa)^2\right]}{\kappa n_3}\left(\sum_{i,k}\langle\boldsymbol{\mathcal{D}}_1, \mathcal{P}_{\mathcal{V}_\star}(\mathfrak{e}_{ijk})\rangle\langle\boldsymbol{\mathcal{D}}_2, \mathcal{P}_{\mathcal{V}_\star}(\mathfrak{e}_{ijk})\rangle\right)^2 \\
=&\frac{(1-\kappa)n_3^2}{\kappa}\sup_\varphi\sum_j\left(\sum_{i,k}\langle\boldsymbol{\mathcal{D}}_1, \mathcal{P}_{\mathcal{V}_\star}(\mathfrak{e}_{ijk})\rangle\langle\boldsymbol{\mathcal{D}}_2, \mathcal{P}_{\mathcal{V}_\star}(\mathfrak{e}_{ijk})\rangle\right)^2 \\
\leq&\frac{(1-\kappa)n_3^2}{\kappa}\sup_\varphi\sum_j\left(\sum_{i,k}|\langle\boldsymbol{\mathcal{D}}_1, \mathcal{P}_{\mathcal{V}_\star}(\mathfrak{e}_{ijk})\rangle||\langle\boldsymbol{\mathcal{D}}_2, \mathcal{P}_{\mathcal{V}_\star}(\mathfrak{e}_{ijk})\rangle|\right)^2 \\
=&\frac{(1-\kappa)n_3^2}{\kappa}\sup_\varphi\sum_j\left(\sum_{i,k}\langle\boldsymbol{\mathcal{D}}_1, \mathcal{P}_{\mathcal{V}_\star}(\mathfrak{e}_{ijk})\rangle^2\right)^2 \\
\leq&\frac{(1-\kappa)n_3^2}{\kappa}\sup_\varphi\sum_j\|\mathcal{P}_{\mathcal{V}_\star}(\boldsymbol{\mathcal{D}}_1)\|_{2,\infty}^2\left(\sum_{i,k}\langle\boldsymbol{\mathcal{D}}_1, \mathcal{P}_{\mathcal{V}_\star}(\mathfrak{e}_{ijk})\rangle^2\right) \\
\leq&\frac{(1-\kappa)n_3^2}{\kappa}\frac{\mu_1 r}{n_2 n_3}\sup_\varphi\sum_j\left(\sum_{i,k}\langle\boldsymbol{\mathcal{D}}_1, \mathcal{P}_{\mathcal{V}_\star}(\mathfrak{e}_{ijk})\rangle^2\right) \\
\leq&\frac{\mu_1 r n_3}{\kappa n_2}\sup_\varphi\sum_j\left(\sum_{i,k}\langle\boldsymbol{\mathcal{D}}_1, \mathcal{P}_{\mathcal{V}_\star}(\mathfrak{e}_{ijk})\rangle^2\right) \\
\leq&\frac{\mu_1 r n_3}{\kappa n_2}\sup_\varphi\|\boldsymbol{\mathcal{D}}_1\|_F^2 \\
=&\frac{\mu_1 r}{\kappa n_2} := \omega
\end{aligned}$$

As $\omega/\nu = 1 > \epsilon$, then by Lemma 10 we can establish

$$\begin{aligned}
&\mathbb{P}\left(\left\|\mathcal{P}_{\mathcal{V}_\star} - \frac{1}{\kappa}\mathcal{P}_{\mathcal{V}_\star}\mathcal{P}_{\widehat{\Theta}}\mathcal{P}_{\mathcal{V}_\star}\right\| > \epsilon\right)\\
=&\mathbb{P}\left(\left\|\sum_j \mathbb{E}[\bar{\boldsymbol{H}}_j]\right\| > \epsilon\right)\\
\leq&(n_1+n_2)n_3\exp\left(-\frac{3\epsilon^2}{8\omega}\right)\\
\leq&(n_1+n_2)n_3\exp\left(-\frac{3\epsilon^2\kappa n_2}{8\mu_1 r}\right).
\end{aligned}$$

Let $\kappa \geq c_3\mu_1 r\log(n_{(1)})/(\epsilon^2 n_2)$. Then, the following inequality holds.

$$\begin{aligned}
&\mathbb{P}\left(\left\|\mathcal{P}_{\mathcal{V}_\star} - \frac{1}{\kappa}\mathcal{P}_{\mathcal{V}_\star}\mathcal{P}_{\widehat{\Theta}}\mathcal{P}_{\mathcal{V}_\star}\right\| \leq \epsilon\right)\\
=&1-\mathbb{P}\left(\left\|\mathcal{P}_{\mathcal{V}_\star} - \frac{1}{\kappa}\mathcal{P}_{\mathcal{V}_\star}\mathcal{P}_{\widehat{\Theta}}\mathcal{P}_{\mathcal{V}_\star}\right\| > \epsilon\right)\\
\geq&1-(n_1+n_2)n_3\exp\left(-\frac{3\epsilon^2\kappa n_2}{8\mu_1 r}\right)\\
\geq&1-2n_3(n_{(1)})^{-\frac{3c_3}{8}+1}.
\end{aligned}$$

By choosing an appropriate c_3, we have $\mathbb{P}(\|\mathcal{P}_{\mathcal{V}_\star} - \kappa^{-1}\mathcal{P}_{\mathcal{V}_\star}\mathcal{P}_{\widehat{\Theta}}\mathcal{P}_{\mathcal{V}_\star}\| \leq \epsilon) \geq 1 - n_{(1)}^{-10}$. The proof is completed. □

4.3.3 Proof of Corollary 7

Proof. Since $\Theta_\star^\perp \sim \text{Ber}(1-\rho)$, by Lemma 6 we have

$$\left\|\mathcal{P}_{\mathcal{V}_\star} - \frac{1}{1-\rho}\mathcal{P}_{\mathcal{V}_\star}\mathcal{P}_{\Theta_\star^\perp}\mathcal{P}_{\mathcal{V}_\star}\right\| \leq \epsilon,$$

provided that $1-\rho \geq c_3\mu_1 r\log(n_{(1)})/(\epsilon^2 n_2)$. Note that $\mathcal{I} = \mathcal{P}_{\Theta_\star} + \mathcal{P}_{\Theta_\star^\perp}$. We can further obtain

$$\begin{aligned}
&\left\|\mathcal{P}_{\mathcal{V}_\star} - \frac{1}{1-\rho}\mathcal{P}_{\mathcal{V}_\star}\mathcal{P}_{\Theta_\star^\perp}\mathcal{P}_{\mathcal{V}_\star}\right\|\\
=&\frac{1}{1-\rho}\left\|\rho\mathcal{P}_{\mathcal{V}_\star} - \mathcal{P}_{\mathcal{V}_\star}\mathcal{P}_{\Theta_\star}\mathcal{P}_{\mathcal{V}_\star}\right\|.
\end{aligned}$$

Thus, we have

$$\begin{aligned}
\|\mathcal{P}_{\Theta_\star}\mathcal{P}_{\mathcal{V}_\star}\|^2 =&\|\mathcal{P}_{\mathcal{V}_\star}\mathcal{P}_{\Theta_\star}\mathcal{P}_{\mathcal{V}_\star}\|\\
\leq&\|\mathcal{P}_{\mathcal{V}_\star}\mathcal{P}_{\Theta_\star}\mathcal{P}_{\mathcal{V}_\star} - \rho\mathcal{P}_{\mathcal{V}_\star}\| + \|\rho\mathcal{P}_{\mathcal{V}_\star}\|\\
\leq&(1-\rho)\epsilon + \rho.
\end{aligned}$$

Thus, the conclusion is established. □

4.3.4 Proof of Lemma 9

Proof. First, it is easy to obtain the following inequality:

$$\begin{aligned}
&\|\widetilde{\mathcal{L}}\|_* + \lambda\|\widetilde{\mathcal{E}} + \mathcal{P}_{\Theta^\perp\cap\Theta_0}(\mathcal{E})\|_{2,1}\\
\leq&\|\widetilde{\mathcal{L}}\|_* + \lambda\|\widetilde{\mathcal{E}}\|_{2,1} + \lambda\|\mathcal{P}_{\Theta^\perp\cap\Theta_0}(\mathcal{E})\|_{2,1}\\
\leq&\|\mathcal{L}\|_* + \lambda\|\mathcal{P}_{\Theta}(\mathcal{E})\|_{2,1} + \lambda\|\mathcal{P}_{\Theta^\perp\cap\Theta_0}(\mathcal{E})\|_{2,1}\\
\leq&\|\mathcal{L}\|_* + \lambda\|\mathcal{P}_{\Theta_0}(\mathcal{E})\|_{2,1}\\
\leq&\|\mathcal{L}\|_* + \lambda\|\mathcal{E}\|_{2,1},
\end{aligned}$$

where the second inequality holds because $(\widetilde{\mathcal{L}}, \widetilde{\mathcal{E}})$ is the optimal solution to the OR-TPCA problem with input $\widetilde{\mathcal{X}} = \mathcal{L}_0 + \mathcal{P}_\Theta(\mathcal{E}_0)$ and thus

$$\|\widetilde{\mathcal{L}}\|_* + \lambda\|\widetilde{\mathcal{E}}\|_{2,1} \leq \|\mathcal{L}\|_* + \lambda\|\mathcal{P}_\Theta(\mathcal{E})\|_{2,1}.$$

On the other hand, we have

$$\widetilde{\mathcal{L}} + \widetilde{\mathcal{E}} + \mathcal{P}_{\Theta^\perp\cap\Theta_0}(\mathcal{E}) = \widetilde{\mathcal{X}} + \mathcal{P}_{\Theta^\perp\cap\Theta_0}(\mathcal{E}) = \mathcal{X}.$$

This means that $\left(\widetilde{\mathcal{L}}, \widetilde{\mathcal{E}} + \mathcal{P}_{\Theta^\perp\cap\Theta_0}(\mathcal{E})\right)$ is also an optimal solution to the OR-TPCA problem with input $\mathcal{X}$. Therefore, it can also exactly recover the tensor column space Range$(\mathcal{L}_0)$ and the outlier support Θ_0. That is, the tensor column space of $\widetilde{\mathcal{L}}$ is Range$(\mathcal{L}_0)$ and $\widetilde{\mathcal{E}} \notin$ Range$(\widetilde{\mathcal{L}})$ = Range$(\mathcal{L}_0)$. So the support set $\widetilde{\Theta}$ of $\widetilde{\mathcal{E}}$ obeys $\Theta \subseteq \widetilde{\Theta}$. If $\Theta \neq \widetilde{\Theta}$, then there exists an index $i \in \Theta^\perp \cap \widetilde{\Theta}$. Accordingly, we have $\widetilde{\mathcal{E}}(:,i,:) \notin$ Range$(\mathcal{L}_0)$. However, since $\widetilde{\mathcal{X}} = \mathcal{L}_0 + \mathcal{P}_\Theta(\mathcal{E}_0)$, $\widetilde{\mathcal{E}}(:,i,:) = \mathcal{L}_0(:,i,:) \in$ Range$(\mathcal{L}_0)$. Therefore, $\Theta = \widetilde{\Theta}$ holds. The proof is completed. □

5. Proofs of Theorem 2

5.1. Main Proof

We first give two theorems that respectively corresponding to Steps 1 and 2 in Algorithm 2 in the manuscript, which will be used later.

Theorem 11. *Suppose that each lateral slice of $\mathcal{X} \in \mathbb{R}^{n_1\times n_2\times n_3}$ is sampled by i.i.d Bernoulli distribution Ber(s/n_2). Let $\mathcal{X}_l$ be the selected lateral slices from $\mathcal{X}$. Then with probability at least $1-\delta$, the clean data in $\mathcal{X}_l$ exactly spans the desired tensor column space Range$(\mathcal{P}_{\Theta_0^\perp}(\mathcal{X}))$ = Range$(\mathcal{L}_0)$, provided that*

$$s \geq 2\mu_1 r\log\left(\frac{r}{\delta}\right),$$

where $r = $ rank$(\mathcal{L}_0)$ and μ_1 is the tensor column-incoherence parameter in Eqn. (3).

Theorem 11 implies that Step 1 in Algorithm 2 can guarantee that the clean data in the sampled data spans the desired tensor column space which is the basis of the subsequent steps in Algorithm 2.

Theorem 12. *Assume that all the assumptions in Theorem 1 are fulfilled for the pair $(\mathcal{L}_l, \mathcal{E}_l)$. Then Step 2 in Algorithm 2 exactly recovers the subspace of $\mathcal{L}_l$ and the support set Θ_l of $\mathcal{E}_l$ with a high probability at least $1 - c_1 n_{(1)}^{-10}$, where c_1 is a positive constant, provided that*

$$s \geq c_2 \mu_1 r \log(n_{(1)}),$$

where c_2 is a constant and μ_1 is the tensor column incoherence parameter in Eqn. (3).

Now we utilize the above Theorems 11 and 12 to prove Theorem 2.

Proof. To prove Theorem 2 in manuscript, we have to prove: (1) Step 1 in Algorithm 2 can guarantees that the clean data in the randomly sampled data $\mathcal{X}_l$ can exactly span the desired column space Range$(\mathcal{L}_0)$ with a probability at least $1-\delta$; (2) Step 2 in Algorithm 2 can exactly recover the tensor column space of the sampled data which is shared by the entire data and detect the outliers in the sampled data; (3) Step 3 can successfully detect the outliers in the remaining data. As for Step 3, it is easy to prove it if Steps 1 and 2 succeed. Since outliers are not in the subspace of the clean data which is guaranteed by the unambiguity condition in the manuscript, accordingly, the outliers in the remaining data are not located in the recovered subspace and thus can be detected.

As for Step 1, by Theorem 11, we know that if

$$s \geq 2\mu_1 r \log\left(\frac{r}{\delta}\right),$$

then with a probability at least $1-\delta$, the clean data of $\mathcal{X}_l$ spans the same tensor column space Range$(\mathcal{L}_0)$.

About Step 2, by Theorem 12, we have that if

$$s \geq c_2 \mu_1 r \log(n_{(1)}),$$

then with probability at least $1-c_1 n_{(1)}^{-10}$, Step 2 exactly recovers tensor column space Range$(\mathcal{L}_l)$ of $\mathcal{X}_l$, which also obeys Range$(\mathcal{L}_l) =$ Range$(\mathcal{L}_0)$, and the support set Θ_l of $\mathcal{E}_l$.

Thus, combine Theorems 11 and 12, if

$$s \geq \max\left(c_2 \mu_1 r \log(n_{(1)}), 2\mu_1 r \log\left(\frac{r}{\delta}\right)\right),$$

the first two steps exactly recover the tensor column space Range$(\mathcal{L}_0)$ and detect outliers $\mathcal{E}_l$ in $\mathcal{X}_l$ with a probability at least $1-\delta$. At the same time, because of Range$(\mathcal{L}_l) =$ Range$(\mathcal{L}_0)$, the data in $\mathcal{X}_r$ except outliers $\mathcal{E}_r$ can be linearly represented by $\mathcal{L}_l$ and by this way, we also can distinguish normal samples and outliers. The proof is completed. □

5.2. Proofs of Some Theorems

Before we prove Theorem 11, we first give one lemma and one theorem, which will be used later.

Theorem 13. *(Matrix Chernoff Bound) [6] Consider a finite sequence $\{\boldsymbol{X}_k\}$ of independent, random, Hermitian matrices. Suppose that*

$$0 \leq \lambda_{min}(\boldsymbol{X}_k) \leq \lambda_{max}(\boldsymbol{X}_k) \leq l.$$

Define $\boldsymbol{Y} = \sum_k \boldsymbol{X}_k$, and μ_r as the r-th largest eigenvalue of the expectation $\mathbb{E}(\boldsymbol{Y})$. That is, $\mu_r = \lambda_r(\mathbb{E}(\boldsymbol{Y}))$. Then for $\epsilon \in [0, 1)$,

$$\begin{aligned}\mathbb{P}\left(\lambda_r(\boldsymbol{Y}) > (1-\epsilon)\mu_r\right) \geq & 1 - r\left(\frac{e^{-\epsilon}}{(1-\epsilon)^{1-\epsilon}}\right)^{\frac{\mu_r}{l}} \\ \geq & 1 - re^{-\frac{\mu_r \epsilon^2}{2l}}\end{aligned}$$

holds.

Lemma 14. *[7] For an arbitrary matrix $\boldsymbol{X} \in \mathbb{C}^{m\times n}$, assume that the skinny SVD of $\boldsymbol{X}$ is $\boldsymbol{X} = \boldsymbol{U\Sigma V}^*$. Then for any set of coordinates Θ, we have rank$(\boldsymbol{X}_{(\Theta,:)}) =$ rank$(\boldsymbol{U}_{(\Theta,:)})$ and rank$(\boldsymbol{X}_{(:,\Theta)}) =$ rank$(\boldsymbol{V}_{(:,\Theta)})$.*

5.2.1 Proof of Theorem 11

Proof. For brevity, we use $\mathcal{Z} \in \mathbb{R}^{n_1\times n_2\times n_3}$ denote the clean data $\mathcal{P}_{\Theta_0^\perp}(\mathcal{X})$ in $\mathcal{X}$. Note that $\mathcal{P}_{\Theta_0}(\mathcal{Z}) = \boldsymbol{0}$. Assume that $\mathcal{Q} \in \mathbb{R}^{n_1\times n_2\times n_3}$ is the sampled lateral slices of $\mathcal{Z}$ with $i.i.d$ Bernoulli distribution Ber(s/n_2). Note that if the i-sampling-probability is larger or equal than s/n_2, then the i-th lateral slice $\mathcal{Q}(:, i, :) = \mathcal{Z}(:, i, :)$; otherwise, $\mathcal{Q}(:, i, :) = \boldsymbol{0}$. In this way, $\mathcal{Q}$ is the clean data of the sampled data in $\mathcal{X}_l$. So we only need to prove Range$(\mathcal{Q}) =$ Range$(\mathcal{L}_0)$.

As defined in Definition 2.5 in the manuscript, for an arbitrary tensor, its tensor column space is the union of the column spaces of its all frontal slices. So we only need to prove that the i-th frontal slice $\bar{\boldsymbol{Q}}^{(i)}$ of $\bar{\mathcal{Q}}$ spans the same column space as the i-th frontal slice $\bar{\boldsymbol{Z}}^{(i)}$ of $\bar{\mathcal{Z}}$, *i.e.* Range$(\bar{\boldsymbol{Q}}^{(i)}) =$ Range$(\bar{\boldsymbol{Z}}^{(i)})$ $(i = 1, \cdots, n_3)$. Assume $\boldsymbol{r} = (\text{rank}(\bar{\boldsymbol{Z}}^{(1)}); \cdots; \text{rank}(\bar{\boldsymbol{Z}}^{(n_3)})) \in \mathbb{R}^{n_3}$ and the tensor tubal rank $r = \text{rank}_t(\mathcal{Z}) = \max(\boldsymbol{r}_1, \cdots, \boldsymbol{r}_{n_3})$.

On the other hand, when conducting DFT on $\mathcal{Z}$, the t-th column $\bar{\boldsymbol{Z}}^{(i)}_{(:,t)}$ of the i-th frontal slice $\bar{\boldsymbol{Z}}^{(i)}$ is computed from the t-th sample $\mathcal{Z}(:, t, :)$. Indeed, we have

$$\bar{\boldsymbol{Z}}^{(i)} = \left[(\boldsymbol{f}_i\boldsymbol{M}^1)^*, (\boldsymbol{f}_i\boldsymbol{M}^2)^*, \cdots, (\boldsymbol{f}_i\boldsymbol{M}^j)^*, \cdots, (\boldsymbol{f}_i\boldsymbol{M}^{n_2})^*\right],$$

where $\boldsymbol{f}_i$ is the i-th row of the DFT matrix $\boldsymbol{F}_{n_3}$ and $\boldsymbol{M}^j \in \mathbb{R}^{n_3\times n_1}$ is the transpose matrix of the j-th lateral slice of $\mathcal{Z}$, *i.e.* $\boldsymbol{M}^j = (\mathcal{Z}(:, j, :))^*$. Therefore, we have $\bar{\boldsymbol{Q}}^{(i)}_{(:,t)} = \delta_t(\boldsymbol{f}_i\boldsymbol{M}^t) = \delta_t\bar{\boldsymbol{Z}}^{(i)}\mathfrak{e}_t$ $(t = 1, \cdots, n_2;\ i = 1, \cdots, n_3)$, where $\mathfrak{e}_t \in \mathbb{R}^{n_2}$ is standard matrix basis with 1 at the t-th entry and $\delta_t = \text{Ber}(s/n_2)$. We can further obtain $\bar{\boldsymbol{Q}}^{(i)} = \sum_{t=1}^{n_2} \bar{\boldsymbol{Z}}^{(i)}_{(:,t)}\mathfrak{e}_t^*$.

Assume that the skinny SVD of matrix $\bar{\boldsymbol{Z}}^{(i)}$ is $\bar{\boldsymbol{U}}_i\bar{\boldsymbol{S}}_i\bar{\boldsymbol{V}}_i^*$. Let $\boldsymbol{T}_i = \sum_{t=1}^{n_2} \delta_t(\bar{\boldsymbol{V}}_i^*)_{(:,t)}\mathfrak{e}_t^*$. Then, we have rank$(\bar{\boldsymbol{Q}}^i) =$

rank$(\boldsymbol{T}_i)$. Define a positive semi-definite matrix,

$$\boldsymbol{D}_i = \boldsymbol{T}_i\boldsymbol{T}_i^* = \sum_{t=1}^{n_2} \delta_t(\bar{\boldsymbol{V}}_i^*)_{(:,t)} \left((\bar{\boldsymbol{V}}_i^*)_{(:,t)}\right)^*.$$

Accordingly, we have $\sigma_{\boldsymbol{r}_i}(\boldsymbol{T}_i)^2 = \lambda_{\boldsymbol{r}_i}(\boldsymbol{D}_i)$, where $\sigma_{\boldsymbol{r}_i}(\boldsymbol{T}_i)$ is the $\boldsymbol{r}_i$-th singular value and $\lambda_{\boldsymbol{r}_i}(\boldsymbol{D}_i)$ is the $\boldsymbol{r}_i$-th eigenvalue. Then we have

$$\begin{aligned}\mathbb{E}(\boldsymbol{D}_i) =& \mathbb{E}\left(\sum_{t=1}^{n_2} \delta_t(\bar{\boldsymbol{V}}_i^*)_{(:,t)} \left((\bar{\boldsymbol{V}}_i^*)_{(:,t)}\right)^*\right)\\ =& \frac{s}{n_2}\sum_{t=1}^{n_2}(\bar{\boldsymbol{V}}_i^*)_{(:,t)} \left((\bar{\boldsymbol{V}}_i^*)_{(:,t)}\right)^*\\ =& \frac{s}{n_2}\bar{\boldsymbol{V}}_i^*\bar{\boldsymbol{V}}_i\\ =& \frac{s}{n_2}\boldsymbol{I}.\end{aligned}$$

Thus, we have $\lambda_{\boldsymbol{r}_i}(\mathbb{E}(\boldsymbol{D}_i)) = s/n_2$. Besides, we can establish

$$\begin{aligned}\lambda_{\max}\left(\delta_t(\bar{\boldsymbol{V}}_i^*)_{(:,t)} \left((\bar{\boldsymbol{V}}_i^*)_{(:,t)}\right)^*\right) =& \|\delta_t(\bar{\boldsymbol{V}}_i^*)_{(:,t)}\|_2^2\\ \leq& \|(\bar{\boldsymbol{V}}_i^*)_{(:,t)}\|_2^2\\ \leq& \|\bar{\boldsymbol{\mathcal{V}}}(:,t,:)\|_F^2\\ \leq& n_3\|\boldsymbol{\mathcal{V}}(:,t,:)\|_F^2\\ \leq& \frac{\mu_1 r}{n_2}.\end{aligned}$$

By utilizing Theorem 13, we can set $\mu_{\boldsymbol{r}_i} = \lambda_{\boldsymbol{r}_i}(\mathbb{E}(\boldsymbol{D}_i)) = s/n_2$ and $l = \lambda_{\max}\left(\delta_t(\bar{\boldsymbol{V}}_i^*)_{(:,t)} \left((\bar{\boldsymbol{V}}_i^*)_{(:,t)}\right)^*\right) \leq \mu_1 r/n_2$. Therefore, we have

$$\mathbb{P}\left(\sigma_{\boldsymbol{r}_i}(\boldsymbol{T}_i) = \lambda_{\boldsymbol{r}_i}(\boldsymbol{D}_i) > 0\right) \geq 1 - \boldsymbol{r}_i e^{-\frac{\mu_{\boldsymbol{r}_i}}{2l}} = 1 - \boldsymbol{r}_i e^{-\frac{s}{2\mu_1 r}}.$$

Note that $\sigma_{\boldsymbol{r}_i}(\boldsymbol{T}_i) > 0$ implies that rank$\left((\bar{\boldsymbol{V}}_i)_{(:,\Theta)}\right) =$ rank$(\boldsymbol{T}_i) \geq \boldsymbol{r}_i$, where Θ is the support of selected samples by $i.i.d.$ Ber(s/n_2). By Theorem 14, we have rank$(\bar{\boldsymbol{Q}}^{(i)}) =$ rank$(\bar{\boldsymbol{Q}}^{(i)}_{(:,\Theta)}) =$ rank$\left((\bar{\boldsymbol{V}}_i)_{(:,\Theta)}\right) \geq \boldsymbol{r}_i$. On the other hand, Range$(\bar{\boldsymbol{Q}}^{(i)}) \subseteq$ Range$(\bar{\boldsymbol{Z}}^{(i)})$. Therefore, we can further obtain Range$(\bar{\boldsymbol{Q}}^{(i)}) =$ Range$(\bar{\boldsymbol{Z}}^{(i)})$.

Since the tensor tubal rank of $\boldsymbol{\mathcal{Z}}$ is $r = \max(\boldsymbol{r}_1, \cdots, \boldsymbol{r}_{n_3})$, there exists a frontal slice $\bar{\boldsymbol{Z}}^{(j)}$ such that rank$(\bar{\boldsymbol{Z}}^{(j)}) = r$. If we set

$$\mathbb{P}(\sigma_{\boldsymbol{r}_j}(\boldsymbol{T}_j) > 0) \geq 1 - re^{-\frac{s}{2\mu_1 r}} \geq 1 - \delta,$$

where $s \geq 2\mu_1 r\log(r/\delta)$, it can guarantee that $\mathbb{P}(\sigma_{\boldsymbol{r}_i}(\boldsymbol{T}_i) > 0) \geq 1-\delta$ $(i \neq j)$. Thus, if $s \geq 2\mu_1 r\log(r/\delta)$, with a probability at least $1-\delta$, Range$(\boldsymbol{\mathcal{Q}}) =$ Range$(\boldsymbol{\mathcal{Z}}) =$ Range$(\boldsymbol{\mathcal{L}}_0)$ holds. Thus, the conclusion in Theorem 11 holds. □

5.2.2 Proof of Theorem 12

Proof. By Theorem 1 in the manuscript, if the sampled number n_2' obeys

$$n_2' \geq c_4\mu_1 r\log(n_{(1)}), \tag{16}$$

where c_4 is a constant, Step 2 in Algorithm 2 can exactly recover the tensor column space of $\boldsymbol{\mathcal{L}}_l$ and the support set Θ_l of $\boldsymbol{\mathcal{E}}_l$ with a high probability $1 - c_1 n_{(1)}^{-10}$.

Now we prove that if we sample each lateral slice of $\boldsymbol{\mathcal{X}}$ with $i.i.d.$ Bernoulli distribution Ber(s'/n_2), where $s' = 2c_4\mu_1 r\log(n_{(1)})$, then the sampled number n_2' obeys Eqn. (16). According to Bernoulli trial property in [7], which states that if $\Theta \sim$ Ber(d/n_2), then with a probability at least $1 - n_2^{-10}$,

$$\frac{1}{2}d \leq |\Theta| = n_2' \leq 2d,$$

provide that $d \geq c_5\log(n_2)$, where c_5 is a constant. Therefore, if $s' = c_2\mu_1 r\log(n_{(1)})$, where $c_2 = 2c_4$, then n_2' obeys the condition (16) with a probability at least $1 - n_2^{-10}$. The proof is completed. □

References

[1] Z. Zhang and S. Aeron, "Exact tensor completion using t-SVD," *arXiv preprint arXiv:1502.04689*, 2015. 1

[2] Z. Lin, R. Liu, and Z. Su, "Linearized alternating direction method with adaptive penalty for low-rank representation," in *Proc. Conf. Neutral Information Processing Systems*, 2011. 1, 2

[3] 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

[4] 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. 2

[5] J. A. Tropp, "An introduction to matrix concentration inequalities," *arXiv preprint:1501.01571*, 2015. 5

[6] A. Gittens and J. A. Tropp, "Tail bounds for all eigenvalues of a sum of random matrices," *arXiv preprint arXiv:1104.4513*, 2011. 8

[7] H. Zhang, Z. Lin, and C. Zhang, "Completing low-rank matrices with corrupted samples from few coefficients in general basis," *IEEE Trans. on Information Theory*, vol. 62, no. 8, pp. 4748–4768, 2016. 8, 9