Supplementary Material for "Fast Guaranteed Tensor Recovery with Adaptive Tensor Nuclear Norm"

Jiangjun Peng[1,2], **Hailin Wang**[3], **Xiangyong Cao**[4], **Shuang Xu**[1,2*]
[1] School of Mathematics and Statistics, Northwestern Polytechnical University, Xi'an 710129, China
[2] Shenzhen Research Institute of Northwestern Polytechnical University, Shenzhen 518057, China
[3] School of Mathematics and Statistics, Xi'an Jiaotong University, Xi'an 710049, China
[4] School of Computer Science and Technology, Xi'an Jiaotong University, Xi'an 710049, China
{pengjj, xs}@nwpu.edu.cn,wanghailin97@163.com, caoxiangyong45@gmail.com

Abstract

In this document, we first introduce the notations, preliminaries, and models in Section 1. The proof about the optimal of ATNN regularizer in the manuscript is placed in Section 2. Next, we provide the proofs of the exact recoverability theories for Tensor Robust Principal Component Analysis (TRPCA) (i.e., Theorem 2) and Tensor Completion (TC) (i.e., Theorem 3) in Sections 3 and 4, respectively. Section 5 presents detailed information about Algorithm 1 and Algorithm 2 mentioned in the manuscript. Finally, in Section 6, we provide additional experimental evidence to further validate the effectiveness of our proposed models. Our code and Supplementary Material are available at https://github.com/andrew-pengjj/adaptive_tensor_nuclear_norm.

Contents

1 Notations and Preliminaries

1.1 Notations

Before completing the proofs, it is necessary to introduce some symbols that will be used throughout the document. In this paper, we denote tensors by boldface Euler script letters, e.g., $\mathcal{A}$. Matrices are denoted by boldface capital letters, e.g., A; vectors are denoted by boldface lowercase letters, e.g., $\boldsymbol{a}$, and scalars are denoted by lowercase letters, e.g., a. We denote I_n as the $n \times n$ identity matrix. For a 3-order tensor $\mathcal{A} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$, we denote its (i, j, k)-th entry as $\mathcal{A}_{ijk}$ or a_{ijk} and use $\mathcal{A}(i, :, :)$, $\mathcal{A}(:, i, :)$ and $\mathcal{A}(:, :, i)$ to denote respectively the i-th horizontal, lateral and frontal slice (see definition in [1]). More often, the frontal slice $\mathcal{A}(:, :, i)$ is denoted compactly as $\mathrm{A}^{(i)}$. The tube is denoted as $\mathcal{A}(i, j, :)$. The mode-n unfolding matrix of $\mathcal{A}$ is denoted as $\mathrm{A}_{(n)} = \text{unfold}_n(\mathcal{A})$, and $\text{fold}_n(\mathrm{A}_{(n)}) = \mathcal{A}$, where fold_n is the inverse of unfolding operator. The mode-n product of a tensor $\mathcal{X} \in \mathbb{R}^{I_1 \times I_2 \times I_3}$ and a matrix $\mathrm{A} \in \mathbb{R}^{J_n \times I_n}$ is denoted as $\mathcal{Y} := \mathcal{X} \times_n \mathrm{A}$ (see definition in [2]). The inner product of between A and B is denoted as $\langle \mathrm{A}, \mathrm{B} \rangle = \mathrm{Tr}(\mathrm{A}^T \mathrm{B})$. The inner product between $\mathcal{A}$ and $\mathcal{B}$ is denoted as $\langle \mathcal{A}, \mathcal{B} \rangle = \sum_{i=1}^{n_3} \langle \mathrm{A}^{(i)}, \mathrm{B}^{(i)} \rangle$.

Some norms of vector, matrix and tensor are used. We denote the $\|\mathcal{A}\|_1 = \sum_{ijk} |a_{ijk}|$, the infinity norm as $\|\mathcal{A}\|_\infty = \max_{ijk} |a_{ijk}|$ and the Frobenius norm as $\|\mathcal{A}\|_F = \sqrt{\sum_{ijk} |a_{ijk}|^2}$, respectively. The spectral norm of a matrix

*Corresponding author

$\mathbf{A}$ is denoted as $\|\mathrm{A}\| = \max_i \sigma_i(\mathrm{A})$, where $\sigma_i(\mathrm{A})$ is the i-th largest singular values of A. The matrix nuclear norm is $\|\mathrm{A}\|_* = \sum_i \sigma_i(\mathrm{A})$.

For a given scalar x, we denote by $\mathrm{sgn}(x)$ the sign of x, which we take to be zero if $x = 0$. By extension, $\mathrm{sgn}(\mathcal{E})$ is the matrix whose entries are the signs of those of $\mathcal{E}$. We recall that any subgradient of the ℓ_1 norm at $\mathcal{E}$ supported on $\mathbf{\Omega}$, is of the form

$$\mathrm{sgn}(\mathcal{E}_0) + \mathcal{F}, \tag{1}$$

where $\mathcal{F}$ vanishes on $\mathbf{\Omega}$, i.e. $\mathcal{P}_\Omega \mathcal{F} = 0$, and obeys $\|\mathcal{F}\|_\infty \le 1$.

Let $\mathrm{A} = \mathrm{USV}^T$ be the skinny SVD of A. It is known that any subgradient of the nuclear norm at A is of the form $\mathrm{UV}^T + \mathrm{W}$, where $\mathrm{U}^T\mathrm{W} = \mathbf{0}$, $\mathrm{WV} = \mathbf{0}$ and $\|\mathrm{W}\| \le 1$ [3].

Similarly, for $\mathcal{A} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ with tubal rank R, we also have the skinny t-SVD, i.e., $\mathcal{A} = \mathcal{U} *_L \mathcal{S} *_L \mathcal{V}^T$, where $\mathcal{U} \in \mathbb{R}^{n_1 \times R \times r_3}$, $\mathcal{S} \in \mathbb{R}^{R \times R \times r_3}$ and $\mathcal{V} \in \mathbb{R}^{n_2 \times R \times r_3}$, in which $\mathcal{U}^T *_L \mathcal{U} = \mathcal{I}$ and $\mathcal{V}^T *_L \mathcal{V} = \mathcal{I}$, where $\mathrm{L} \in \mathbb{R}^{n_3 \times r_3}$. The skinny t-SVD will be used throughout this paper. With skinny t-SVD, we introduce the subgradient of the tensor nuclear norm, which plays an important role in the proofs.

1.2 Subgradient of Tensor Nuclear Norm

Theorem 1 (Subgradient of tensor nuclear norm)**.** *Let $\mathcal{A} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ with $\mathrm{rank}_t(\mathcal{A}) = R$ and its skinny t-SVD be $\mathcal{A} = \mathcal{U} *_L \mathcal{S} *_L \mathcal{V}^T$ under the COM $\mathrm{L} \in \mathbb{R}^{n_3 \times r_3}$. The subdifferential (the set of subgradients) of $\|\mathcal{A}\|_*$ is:*

$$\partial\|\mathcal{A}\|_* = \{\mathcal{U} *_L \mathcal{V}^T + \mathcal{W} | \mathcal{U}^T *_L \mathcal{W} = \mathbf{0}, \mathcal{W} *_L \mathcal{V} = \mathbf{0}, \|\mathcal{W}\| \le 1\}. \tag{2}$$

Proof. The proof is by construction. According to t-product definition in the manuscript, we have

$$\mathcal{A} = \mathcal{U} *_L \mathcal{S} *_L \mathcal{V}^T \iff \overline{\mathrm{A}} = \overline{\mathrm{U}}\,\overline{\mathrm{S}}\,\overline{\mathrm{V}}^T, \tag{3}$$

where $\overline{\mathrm{U}} = \mathrm{bdiag}(\overline{\mathcal{U}})$, $\overline{\mathrm{V}}^T = \mathrm{bdiag}(\overline{\mathcal{V}}^T)$ and $\overline{\mathrm{S}} = \mathrm{bdiag}(\overline{\mathcal{S}})$. According to the following equation:

$$\|\mathcal{A}\|_* = \|\mathcal{S}\|_1 = \|\overline{\mathcal{S}}\|_* = \|\overline{\mathrm{A}}\|_* = \|\overline{\mathcal{A}}\|_*, \tag{4}$$

we have $\partial\|\mathcal{A}\|_* = \partial\|\overline{\mathrm{A}}\|_*$. Since $\|\overline{\mathrm{A}}\|_*$ is diagonal block matrix, we have $\|\overline{\mathrm{A}}\|_* = \sum_{i=1}^{r_3} \|\overline{\mathrm{A}}^{(i)}\|_*$. Performing matrix singular vector decomposition (SVD) operation on each frontal slice $\overline{\mathrm{A}}^{(i)}$, we have $\overline{\mathrm{A}}^{(i)} = \mathrm{U}_{(i)}\mathrm{S}_{(i)}\mathrm{V}_{(i)}^T$, where $\mathrm{U}_{(i)}, \mathrm{V}_{(i)}^T$ are orthogonal matrix and $\mathrm{S}_{(i)}$ is a diagonal matrix. Merge the SVD of each frontal slice together, we can set

$$\overline{\mathrm{U}} = \begin{bmatrix} \mathrm{U}_{(i)} & & \\ & \ddots & \\ & & \mathrm{U}_{(r_3)} \end{bmatrix}, \overline{\mathrm{S}} = \begin{bmatrix} \mathrm{S}_{(1)} & & \\ & \ddots & \\ & & \mathrm{S}_{(r_3)} \end{bmatrix}, \quad \overline{\mathrm{V}} = \begin{bmatrix} \mathrm{V}_{(i)} & & \\ & \ddots & \\ & & \mathrm{V}_{(r_3)} \end{bmatrix}, \tag{5}$$

Next, we prove the form of subgradient of $\partial\|\mathcal{A}\|_*$.

For each frontal slice $\|\overline{\mathrm{A}}^{(i)}\|_*$, its subgradient is: $\partial\|\overline{\mathrm{A}}^{(i)}\|_* = \mathrm{U}_{(i)}\mathrm{V}_{(i)}^T + \mathrm{W}_{(i)}$, where $\mathrm{W}_{(i)}$ satisfies $\mathrm{U}_{(i)}^T\mathrm{W}_{(i)} = \mathbf{0}$, $\mathrm{W}_{(i)}\mathrm{V}_{(i)} = \mathbf{0}$ and $\|\mathrm{W}_{(i)}\| \le 1$. Defining

$$\overline{\mathrm{W}} = \begin{bmatrix} \mathrm{W}_{(i)} & & \\ & \ddots & \\ & & \mathrm{W}_{(r_3)} \end{bmatrix}, \tag{6}$$

we can easily obtain that $\overline{\mathrm{W}}$ satisfies $\overline{\mathrm{U}}^T\,\overline{\mathrm{W}} = 0$, $\overline{\mathrm{W}}\,\overline{\mathrm{V}} = 0$ and $\|\overline{\mathrm{W}}\| \le 1$. Then, we have

$$\begin{aligned} \partial\|\mathrm{A}\|_* = \partial\|\overline{\mathrm{A}}\|_* &= \sum_{i=1}^{r_3}\{\mathrm{U}_{(i)}\mathrm{V}_{(i)}^T + \mathrm{W}_{(i)}\} \\ &= \overline{\mathrm{U}}\,\overline{\mathrm{V}}^T + \overline{\mathrm{W}} = \mathrm{U} *_L \mathrm{V}^T + \mathrm{W}, \end{aligned} \tag{7}$$

with $\mathrm{U}^T *_L \mathrm{W} = 0$, $\mathrm{W} *_L \mathrm{V} = 0$, $\|\mathrm{W}\| \le 1$, where $\mathrm{U} = \mathrm{bfold}(\overline{\mathrm{U}}) \in \mathbb{R}^{n_1 \times R \times n_3}$, $\mathrm{V} = \mathrm{bfold}(\overline{\mathrm{V}}) \in \mathbb{R}^{n_2 \times R \times n_3}$, and $\mathrm{W} = \mathrm{bfold}(\overline{\mathrm{W}}) \in \mathbb{R}^{n_1 \times n_2 \times n_3}$. This completes the proof. □

Furthermore, we define the $\ell_{\infty,2}$-norm of the tensor $\mathcal{A}$ as

$$\|\mathcal{A}\|_{\infty,2} = \max\{\max_i \|\mathcal{A}(i,:,:)\|_F, \max_j \|\mathcal{A}(:,j,:)\|_F\}. \tag{8}$$

Define the projection $\mathcal{P}_\Omega(\mathcal{Z}) = \sum_{i,j,k} \delta_{ijk} z_{ijk} \mathfrak{e}_{ijk}$, where $\delta_{ijk} = 1_{(i,j,k)\in\Omega}$, where $1_{(.)}$ is the indicator function. Also Ω^c denotes the complement of Ω and $\mathcal{P}_{\Omega^\perp}$ is the projection onto Ω^c. Denote T by the set

$$T = \{\mathcal{U} *_L \mathcal{Y}^T + \mathcal{W} *_L \mathcal{V}^T, \mathcal{Y}, \mathcal{W} \in \mathbb{R}^{n \times r \times n_3}\}, \tag{9}$$

and by $T^\perp$ its orthogonal complement. Then the projections onto T and $T^\perp$ are respectively

$$\begin{aligned} \mathcal{P}_T(\mathcal{Z}) &= \mathcal{U} *_L \mathcal{U}^T *_L \mathcal{Z} + \mathcal{Z} *_L \mathcal{U} *_L \mathcal{U}^T \\ &\quad - \mathcal{U} *_L \mathcal{U}^T *_L \mathcal{Z} *_L \mathcal{U} *_L \mathcal{U}^T, \\ \mathcal{P}_{T^\perp}(\mathcal{Z}) &= \mathcal{Z} - \mathcal{P}_T(\mathcal{Z}) = (\mathcal{I}_{n_1} - \mathcal{U} *_L \mathcal{U}^T) *_L \\ &\quad \mathcal{Z} *_L (\mathcal{I}_{n_2} - \mathcal{V} *_L \mathcal{V}^T). \end{aligned} \tag{10}$$

We denote $\mathring{\mathfrak{e}}_i$ as the tensor column basis, which is a tensor of size $n_1 \times 1 \times n_3$ with its $(i,1,1)$-th entry equaling 1 and the rest equaling 0 [1]. We also define the tensor tube basis $\dot{\mathfrak{e}}_j$, which is a tensor of size $1 \times 1 \times n_3$ with its $(1,1,k)$-th entry equaling 1 and the rest equaling 0.

For $i = 1, \cdots, n_1$, $j = 1, \cdots, n_2$ and $k = 1, \cdots, n_3$, we define the random variable $\delta_{ijk} = 1_{(i,j,k)\in\Omega}$. Then the projection $\mathcal{R}_\Omega$ is given by

$$\mathcal{R}_\Omega := \frac{1}{p}\mathcal{P}_\Omega(\mathcal{Z}) = \sum_{i,j,k} \frac{1}{p}\delta_{ijk} z_{ijk} \mathfrak{e}_{ijk}, \tag{11}$$

where $\mathfrak{e}_{ijk} = \mathring{\mathfrak{e}}_i \dot{\mathfrak{e}}_j \mathring{\mathfrak{e}}_k$ is an $n \times n \times n_3$ sized tensor with (i,j,k)-th entry equaling 1 and the rest equaling 0. Also Ω^c denotes the complement of Ω and $\mathcal{P}_{\Omega^\perp}$ is the projection onto Ω^c. Then we can get

$$\|\mathcal{P}_T(\mathfrak{e}_{ijk})\|_F^2 \le \frac{\mu R(n_1 + n_2)}{n_1 n_2} = \frac{2\mu R}{n}, \text{if } \mathrm{n}_1 = \mathrm{n}_2 = \mathrm{n}, \tag{12}$$

by using the Definition 1, i.e., the following tensor incoherence condition (13).

Definition 1 (Tensor Incoherence Conditions). *For $\mathcal{X}_0 \in \mathbb{R}^{n_1\times n_2\times n_3}$ with t-SVD rank R, it has the skinny t-SVD $\mathcal{X}_0 = \mathcal{U} *_L \mathcal{S} *_L \mathcal{V}^T$. Then $\mathcal{X}_0$ is said to satisfy the tensor incoherence conditions with parameter μ if*

$$\begin{aligned}&\max_{i\in[1,n_1]} \|\mathcal{U}^T *_L \mathring{\mathfrak{e}}_i\|_F \le \sqrt{\frac{\mu R}{n_1}},\\ &\max_{j\in[1,n_2]} \|\mathcal{V}^T *_L \mathring{\mathfrak{e}}_j\|_F \\ &\le \sqrt{\frac{\mu R}{n_2}}, \|\mathcal{U} *_L \mathcal{V}^T\|_F \le \sqrt{\frac{\mu R}{n_1 n_2}}.\end{aligned} \tag{13}$$

1.3 Models

Two types of models are given in this paper, i.e.,

$$\min_{\mathcal{T}=\mathcal{B}\times_3 \mathrm{D}} \|\mathcal{B}\|_*, s.t., \mathcal{P}_\Omega(\mathcal{Y}) = \mathcal{P}_\Omega(\mathcal{T}), \tag{14}$$

and

$$\min_{\mathcal{T}=\mathcal{B}\times_3 \mathrm{D},\mathcal{E}} \|\mathcal{B}\|_* + \lambda\|\mathcal{E}\|_1, s.t., \mathcal{Y} = \mathcal{T} + \mathcal{E}. \tag{15}$$

Ideally, we know the transformation matrix and can fix it, but more often the transformation matrix is unknown. The following theorem demonstrates that the representation of the ground-truth tensor $\mathcal{T}_0$ under the learned COM D^* preserves information.

Theorem 2. *Suppose $\mathcal{T}_0 \in \mathbb{R}^{n_1\times n_2\times n_3}$ is the ground-truth tensor, it can be decomposed as $\mathcal{T}_0 = \mathcal{B}_0 \times_3 \mathrm{D}_0$, where $\mathrm{D}_0 \in \mathbb{R}^{n_3\times r_3}(r_3 \le n_3)$ is the column-orthogonal matrix. Then, for any column-orthogonal matrix D of the same size as D_0, $\mathcal{T}_0$ can be represented exactly.*

Proof. Since both $\mathrm{D}_0 \in \mathbb{R}^{n_3\times r_3}$ and $\mathrm{D} \in \mathbb{R}^{n_3\times r_3}(r_3 \le n_3)$ are column-orthogonal matrices, there exists an orthogonal matrix $\mathrm{Q} \in \mathbb{R}^{r_3\times r_3}$ that satisfies $\mathrm{D}_0 = \mathrm{DQ}$. Then we have

$$\mathcal{T}_0 = \mathcal{B}_0 \times_3 \mathrm{D}_0 = \mathcal{B}_0 \times_3 (\mathrm{DQ}) = \underbrace{\mathcal{B}_0 \times_3 \mathrm{Q}}_{\mathcal{B}} \times_3 \mathrm{D}. \tag{16}$$

This completes the proof. □

Once we get the optimal COM D_0 learned from the data, we can fixed D_0 in ATNN framework. Therefore, we we prove the exact recoverability theory of models (14) and (15) with fixed COM D_0.

2 The Proof of the Adaptive Transformation

If D is the full-rank invertible orthogonal transform, and regard $\mathcal{B}$ as a transformed tensor $\mathcal{T}_{\mathfrak{L}} = \mathfrak{L}(\mathcal{T}) = \mathcal{T} \times_3 \mathrm{D}^T$, then we can get obtain that

$$\begin{aligned}\mathfrak{L}^{-1}(\mathfrak{L}(\mathcal{T})) &= \mathfrak{L}^{-1}(\mathcal{T}_{\mathfrak{L}}) = \mathcal{T}_{\mathfrak{L}} \times_3 \mathrm{D}\\ &= \mathcal{T} \times_3 \mathrm{D}^T \times_3 \mathrm{D} = \mathcal{B} \times_3 \mathrm{D} = \mathcal{T}.\end{aligned} \tag{17}$$

Next, we prove that even if L is not a full-rank transformation, as long as L is decomposed from real data, the above equation still holds, that is, the following theorem holds.

Theorem 3. *For a three-order tensor $\mathcal{T} = \mathcal{B} \times_3 \mathrm{D} \in \mathbb{R}^{n_1\times n_2\times n_3}$ composed of $\mathcal{B} \in \mathbb{R}^{n_1\times n_2\times r_3}$ and $\mathrm{D} \in \mathbb{R}^{n_3\times r_3}$, if the transformation matrix is not chosen as D^T, Eq. (17) does not hold.*

Proof. For any column-orthogonal matrix $\mathrm{D}_1 \in \mathbb{R}^{n_3\times r_3}$, we have

$$\begin{aligned}\mathcal{T} - \mathfrak{L}^{-1}(\mathfrak{L}(\mathcal{T})) &= \mathcal{T} \times_3 \mathrm{D}^T \times_3 \mathrm{D} - \mathcal{T} \times_3 \mathrm{D}_1^T \times_3 \mathrm{D}_1\\ &= \mathcal{T} \times_3 (\mathrm{DD}^T - \mathrm{D}_1\mathrm{D}_1^T) \ne 0.\end{aligned}$$

Therefore, Eq. (17) is valid only when the transform matrix is set as the transforma matrix D^T embedded in the data. □

Next, we will focus on why the column orthogonal matrix D obtained from the original tensor is used as the transformation matrix.

The original ATNN models is:

$$\begin{aligned}\|\mathcal{T}\|_{\circledast,A} := \sum_{k=1}^{r_3} \|\mathcal{B}(:,:,k)\|_* = \|\mathcal{B}\|_*,\\ s.t.\ \mathcal{B} = \mathcal{T} \times_3 \mathrm{D}^T, \mathrm{D}^T\mathrm{D} = \mathrm{I}.\end{aligned} \tag{18}$$

ATNN model (18) is a definition with constraints. A natural question is what kind of transformation operator is appropriate? A suitable transformation operator must first ensure that information is not lost. In fact, it can minimize the objective function of ATNN, that is, the transformation operator can induce relatively low-rank transform domain data, so that the nuclear norm can play a better role in the optimization process. To achieve better approximation, setting $\overline{\mathcal{T}} = \mathcal{T} \times_3 \mathrm{D}^T$, then finding the appropriate transformation matrix becomes the following optimization problem:

$$\min_{\mathrm{D},\mathrm{D}^T\mathrm{D}=\mathrm{I}} \sum_{i=1}^{r_3} \frac{1}{r_3}\|\overline{\mathcal{T}}(:,:,i)\|_*, s.t. \mathcal{T} = \mathcal{T} \times_3 \mathrm{D}^T \times_3 \mathrm{D}. \tag{19}$$

Theorem 4. *For optimization problem (19), we have the following two assertions.*

a) The upper bound objective of the optimization problem (19) is $\|\mathcal{T}_{(3)}\mathrm{D}\|_{2,1}$.

b) If the transform matrix is set as the component matrix D^T of tensor $\mathcal{T}$, then the upper bound objective can obtain the smallest value.

Proof. Proof of a): According to the norm triangle inequality, i.e., $\frac{1}{n}\|\mathrm{X}\|_* \le \|\mathrm{X}\| \le \|\mathrm{X}\|_F$ for any $\mathrm{X} \in \mathbb{R}^{n_1\times n_2}$, we can obtain that

$$\begin{aligned}\sum_{i=1}^{r_3} \frac{1}{r_3}\|\overline{\mathcal{T}}(:,:,i)\|_* \le \sum_{i=1}^{r_3} \|\overline{\mathcal{T}}(:,:,i)\|_F = \|\overline{\mathcal{T}}_{(3)}\|_{2,1}\\ = \|\mathcal{T}_{(3)}\mathrm{D}\|_{2,1}.\end{aligned} \tag{20}$$

This completes the assertion (a).

Proof of b): Without loss of generality, assume $n_1 \le n_2$. Given a fixed matrix $\mathcal{T}_{(3)} \in \mathbb{R}^{n_1\times n_3}$ with rank r, and its skinny SVD as $\mathcal{T}_{(3)} = \mathrm{U\Sigma V}^T$ with $\mathrm{U} \in \mathbb{R}^{n_1\times r}$, $\Sigma \in \mathbb{R}^{r\times r}$ and $\mathrm{V} \in \mathbb{R}^{n_2\times r}$. For any matrix $\mathrm{X} \in \mathbb{R}^{r\times n_2}$ and column-orthogonal matrix $\mathrm{U} \in \mathbb{R}^{n_1\times r}$ with $\mathrm{U}^T\mathrm{U} = \mathrm{I}$, then for any $i = 1\cdots, n_2$, we have $\|\mathrm{UX}\|_{2,1} = \|\mathrm{X}\|_{2,1}$ since

$$\begin{aligned}\|\mathrm{UX}(:,j)\|_2 &= \sqrt{(\mathrm{UX}(:,j))^T(\mathrm{UX}(:,j))}\\ &= \sqrt{\mathrm{X}(:,j)^T\mathrm{U}^T\mathrm{UX}(:,j)} = \|\mathrm{X}(:,j)\|_2.\end{aligned} \tag{21}$$

Then we can get

$$\|\mathcal{T}_{(3)}\mathrm{D}\|_{2,1} = \|\mathrm{U}\Sigma\mathrm{V}^T\mathrm{D}\|_{2,1} = \|\Sigma\mathrm{V}^T\mathrm{D}\|_{2,1}. \quad (22)$$

Set $\mathrm{Z} := \mathrm{V}^T\mathrm{D}$, it is easily to validate that $\mathrm{Z} \in \mathbb{R}^{r\times r}$ with $\mathrm{Z}^T\mathrm{Z} = \mathrm{I}$. We then prove that $\mathrm{Z} = \mathrm{I}$ optimize the following problem:

$$\min_{\mathrm{Z},\mathrm{Z}^T\mathrm{Z}=\mathrm{I}} \|\Sigma\mathrm{Z}\|_{2,1}. \quad (23)$$

We use Z_{ij} to represent the (i,j)-th element of matrix Z, and use $\mathbf{z}_i$ to represent the i-th column of matrix Z. We give the following deduction:

$$\|\Sigma\mathrm{Z}\|_{2,1} = \sum_{i=1}^{r}\|\Sigma\mathbf{z}_i\|_2 = \sum_{i=1}^{r}\sqrt{\sum_{j=1}^{r}(\sigma_j\mathrm{Z}_{ij})^2} \quad (24)$$

$$\overset{(1)}{=} \sum_{i=1}^{r}\sqrt{\sum_{j=1}^{r}(\sigma_j\mathrm{Z}_{ij})^2 \times \sum_{j=1}^{r}\mathrm{Z}_{ij}^2} \quad (25)$$

$$\overset{(2)}{\geq} \sum_{i=1}^{r}\sum_{j=1}^{r}\sigma_j\mathrm{Z}_{ij}^2 \overset{(3)}{=} \sum_{j=1}^{r}\sigma_j\sum_{i=1}^{r}\mathrm{Z}_{ij}^2 \quad (26)$$

$$\overset{(4)}{=} \sum_{j=1}^{r}\sigma_j = \|\Sigma\|_{2,1}, \quad (27)$$

where (1) holds due to that Z is an orthogonal matrix with normalized columns, (2) holds because of Cauchy inequality, (3) holds with exchanging the order of two summations, (4) holds due to that Z is an orthogonal matrix with normalized rows. Notice that the equality in (2) holds if and only if the two vectors $(\sigma_1\mathrm{Z}_{i1}, \sigma_2\mathrm{Z}_{i2}\cdots,\sigma_r\mathrm{Z}_{ir})$ and $(\mathrm{Z}_{i1},\mathrm{Z}_{i2}\cdots,\mathrm{Z}_{ir})$ are parallel. It can be seen that when $\mathrm{Z} = \mathrm{I}$ (i.e., $\mathrm{D} = \mathrm{V}$), the condition are satisfied.

This completes the assertion (b).

□

3 The Proof of Exact Recovery Theorem about TRPCA Model

In this section, we first introduce conditions for $(\mathcal{T}_0, \mathcal{E}_0)$ to be the unique solution to TRPCA model (14). Then we construct a dual certificate in subsection 3.2 which satisfies the conditions in subsection 3.1, and thus our main results in Theorem 2 in our paper are proved.

3.1 Dual Certificates

Lemma 1. *Assume that* $\|\mathcal{P}_\Omega\mathcal{P}_T\| \leq \frac{1}{2}$ *and* $\lambda < \frac{1}{\sqrt{n_3}}$. *Then* $(\mathcal{T}_0, \mathcal{E}_0)$ *is the unique solution to the TRPCA problem if there is a pair* $(\mathcal{W}, \mathcal{F})$ *obeying*

$$\mathcal{U} *_{\mathrm{D}} \mathcal{V}^T + \mathcal{W} = \lambda(\mathrm{sgn}(\mathcal{S}_0) + \mathcal{F} + \mathrm{mathcalP}_\Omega\mathcal{D}), \quad (28)$$

with $\mathcal{P}_T\mathcal{W} = \mathbf{0}$, $\|\mathcal{W}\| \leq \frac{1}{2}$, $\mathcal{P}_\Omega\mathcal{F} = \mathbf{0}$ *and* $\|\mathcal{F}\|_\infty \leq \frac{1}{2}$ *and* $\|\mathcal{P}_\Omega\mathcal{D}\|_F \leq \frac{1}{4}$.

Proof. For any $\mathcal{H} \neq 0$, $(\mathcal{T}_0 + \mathcal{H}, \mathcal{E}_0 - \mathcal{H})$ is also a feasible solution. We show that its objective is larger than that at $(\mathcal{T}_0, \mathcal{E}_0)$, hence proving that $(\mathcal{T}_0, \mathcal{E}_0)$ is the unique solution. To do this, let $\mathcal{U} *_{\mathrm{D}} \mathcal{V}^T + \mathcal{W}_0$ be an arbitrary subgradient of the tensor nuclear norm at $\mathcal{X}_0$ under the COM D, and $\mathrm{sgn}(\mathcal{E}_0)+\mathcal{F}_0$ be an arbitrary subgradient of the ℓ_1-norm at $\mathcal{E}_0$. Then we have

$$\|\mathcal{T}_0 + \mathcal{H}\|_* + \lambda\|\mathcal{E}_0 - \mathcal{H}\|_1 \geq \|\mathcal{T}_0\|_* + \lambda\|\mathcal{E}_0\|_1 + \left\langle \mathcal{U} *_{\mathrm{D}} \mathcal{V}^T + \mathcal{W}_0, \mathcal{H}\right\rangle - \lambda\left\langle \mathrm{sgn}(\mathcal{E}_0) + \mathcal{F}_0, \mathcal{H}\right\rangle$$

Now pick $\mathcal{W}_0$ such that $\langle \mathcal{W}_0, \mathcal{H}\rangle = \|\mathcal{P}_{T^\perp}\mathcal{H}\|_*$ and $\langle \mathcal{F}_0, \mathcal{H}\rangle = \|\mathcal{P}_{\Omega^\perp}\mathcal{H}\|$. We have

$$\|\mathcal{T}_0 + \mathcal{H}\|_* + \lambda\|\mathcal{E}_0 - \mathcal{H}\|_1 \geq \|\mathcal{T}_0\|_* + \lambda\|\mathcal{E}_0\|_1 + \|\mathcal{P}_{T^\perp}\mathcal{H}\|_* + \lambda\|\mathcal{P}_{\Omega^\perp}\mathcal{H}\|_1 + \left\langle \mathcal{U} *_{\mathrm{D}} \mathcal{V}^T - \mathrm{sgn}(\mathcal{E}_0), \mathcal{H}\right\rangle.$$

By assumption, we have

$$\begin{aligned} &\left|\left\langle \mathcal{U} *_D \mathcal{V}^T - \mathrm{sgn}(\mathcal{E}_0), \mathcal{H}\right\rangle\right| \leq |\langle \mathcal{W}, \mathcal{H}\rangle| \\ &\quad + \lambda\,|\langle \mathcal{F}, \mathcal{H}\rangle| + \lambda\,|\langle \mathcal{P}_\Omega\mathcal{D}, \mathcal{H}\rangle| \\ &\leq \beta\left(\|\mathcal{P}_{T^\perp}\mathcal{H}\|_* + \lambda\|\mathcal{P}_{\Omega^\perp}\mathcal{H}\|_1\right) + \frac{\lambda}{4}\|\mathcal{P}_\Omega\mathcal{H}\|_F, \end{aligned} \quad (29)$$

where $\beta = \max(\|\mathcal{W}\|, \|\mathcal{F}\|_\infty) < \frac{1}{2}$. Thus we have

$$\|\mathcal{T}_0 + \mathcal{H}\|_* + \lambda\|\mathcal{E}_0 - \mathcal{H}\|_1 \geq \|\mathcal{T}_0\|_* + \lambda\|\mathcal{E}_0\|_1 + \frac{1}{2}\left(\|\mathcal{P}_{T^\perp}\mathcal{H}\|_* + \lambda\|\mathcal{P}_{\Omega^\perp}\mathcal{H}\|_1\right) - \frac{\lambda}{4}\|\mathcal{P}_\Omega\mathcal{H}\|_F.$$

On the other hand,

$$\begin{aligned} \|\mathcal{P}_\Omega\mathcal{H}\|_F &\leq \|\mathcal{P}_\Omega\mathcal{P}_T\mathcal{H}\|_F + \|\mathcal{P}_\Omega\mathcal{P}_{T^\perp}\mathcal{H}\|_F \\ &\leq \frac{1}{2}\|\mathcal{H}\|_F + \|\mathcal{P}_{T^\perp}\mathcal{H}\|_F \\ &\leq \frac{1}{2}\|\mathcal{P}_\Omega\mathcal{H}\|_F + \frac{1}{2}\|\mathcal{P}_{\Omega^\perp}\mathcal{H}\|_F + \|\mathcal{P}_{T^\perp}\mathcal{H}\|_F. \end{aligned}$$

Thus we can obtain

$$\begin{aligned} \|\mathcal{P}_\Omega\mathcal{H}\|_F &\leq \|\mathcal{P}_{\Omega^\perp}\mathcal{H}\|_F + 2\|\mathcal{P}_{T^\perp}\mathcal{H}\|_F \\ &\leq \|\mathcal{P}_{\Omega^\perp}\mathcal{H}\|_1 + 2\sqrt{n_3}\|\mathcal{P}_{T^\perp}\mathcal{H}\|_*. \end{aligned}$$

In conclusion,

$$\|\mathcal{T}_0 + \mathcal{H}\|_* + \lambda\|\mathcal{E}_0 - \mathcal{H}\|_1 \geq \|\mathcal{X}_0\|_* + \lambda\|\mathcal{E}_0\|_1 + \frac{1}{2}(1 - \lambda\sqrt{n_3})\|\mathcal{P}_{T^\perp}\mathcal{H}\|_* + \frac{\lambda}{4}\|\mathcal{P}_\Omega\mathcal{H}\|_1,$$

and the last two terms are strictly positive when $\mathcal{H} \neq \mathbf{0}$. Thus, the proof is completed. □

Lemma 1 implies that it suffices to produce a dual certificate $\mathcal{W}$ obeying

$$\begin{cases} \mathcal{W} \in T^\perp, \\ \|\mathcal{W}\| \leq \frac{1}{2}, \\ \|\mathcal{P}_\Omega(\mathcal{U} *_D \mathcal{V}^T + \mathcal{W} - \lambda\mathrm{sgn}(\mathcal{S}_0)\|_F \leq \frac{\lambda}{4}, \\ \|\mathcal{P}_{\Omega^\perp}(\mathcal{U} *_D \mathcal{V}^T + \mathcal{W}\|_\infty \leq \frac{\lambda}{2}. \end{cases} \quad (30)$$

‘‘‘

3.2 Dual Certification via The Golfing Scheme

The remaining work is to construct the aforementioned dual certificates. Before introducing our construction, we first assume that $\Omega \sim \mathrm{Ber}(\rho)$, or equivalently that $\Omega^c \sim \mathrm{Ber}(1-\rho)$. Now the distribution of Ω^c is the same as that of $\Omega^c =$

$\Omega_1 \cup \Omega_2 \cup \cdots \cup \Omega_{j_0}$, where each Ω_j follows the Bernoulli model with parameter q, that is,

$$\mathbb{P}\left((i,j,k) \in \Omega\right) = \mathbb{P}(\text{Bin}(j^0, q) = 0) = (1-q)^{j_0},$$

so that the two models are the same if $\rho = (1-q)^{j_0}$. Note that because of overlaps between the Ω_j's, $q \geq (1-\rho)/j_0$. Now, we construct a dual certificate

$$\mathcal{W} = \mathcal{W}^{\mathcal{L}} + \mathcal{W}^{\mathcal{S}},$$

where each component is as follows:

1) Construction of $\mathcal{W}^{\mathcal{L}}$ via the Golfing scheme. Let $j_0 \geq 1$, and let $\Omega_j, 1 \leq j \leq j_0$, be defined as aforementioned so that $\Omega^c = \cup_{1 \leq j \leq j_0} \Omega_j$. Then define

$$\mathcal{W}^{\mathcal{L}} = \mathcal{P}_{T^\perp} \mathbf{Y}_{j_0}, \tag{31}$$

where

$$\mathcal{Y}_j = \mathcal{Y}_{j-1} + q^{-1} \mathcal{P}_{\Omega_j} \mathcal{P}_T \left(\mathcal{U} *_D \mathcal{V}^T - \mathcal{Y}_{j-1}\right), \mathcal{Y}_0 = \mathbf{0}. \tag{32}$$

2) Construction of $\mathcal{W}^{\mathcal{S}}$ via the Method of Least Squares. Assume that $\|\mathcal{P}_\Omega \mathcal{P}_T\| \leq \frac{1}{2}$. Then, $\|\mathcal{P}_\Omega \mathcal{P}_T \mathcal{P}_\Omega\| < \frac{1}{4}$ and thus, the operator $\mathcal{P}_\Omega - \mathcal{P}_\Omega \mathcal{P}_T \mathcal{P}_\Omega$ mapping Ω onto itself is invertible, and its inverse is denoted by $(\mathcal{P}_\Omega - \mathcal{P}_\Omega \mathcal{P}_T \mathcal{P}_\Omega)^{-1}$. We then set

$$\mathcal{W}^{\mathcal{S}} = \lambda \mathcal{P}_{T^\perp} \left(\mathcal{P}_\Omega - \mathcal{P}_\Omega \mathcal{P}_T \mathcal{P}_\Omega\right)^{-1} \text{sgn}(\mathcal{E}_0). \tag{33}$$

This is equivalent to

$$\mathcal{W}^{\mathcal{S}} = \lambda \mathcal{P}_{T^\perp} \sum_{k \geq 0} (\mathcal{P}_\Omega \mathcal{P}_T \mathcal{P}_\Omega)^k \text{sgn}(\mathcal{E}_0). \tag{34}$$

Since both $\mathcal{W}^{\mathcal{L}}, \mathcal{W}^{\mathcal{S}} \in T^\perp$ and $\mathcal{P}_\Omega \mathcal{W}^{\mathcal{S}} = \lambda \mathcal{P}_\Omega (\mathcal{I} - \mathcal{P}_T) \left(\mathcal{P}_\Omega - \mathcal{P}_\Omega \mathcal{P}_T \mathcal{P}_\Omega\right)^{-1} \text{sgn}(\mathcal{E}_0) = \lambda \text{sgn}(\mathcal{E}_0)$, we shall establish that $\mathcal{W}^{\mathcal{L}} + \mathcal{W}^{\mathcal{S}}$ is a valid dual certificate if it obeys

$$\begin{cases} \|\mathcal{W}^{\mathcal{L}} + \mathcal{W}^{\mathcal{S}}\| < \frac{1}{2}, \\ \left\|\mathcal{P}_\Omega \left(\mathrm{U} *_D \mathrm{V}^T + \mathcal{W}^{\mathcal{L}}\right)\right\|_F \leq \frac{\lambda}{4}, \\ \left\|\mathcal{P}_{\Omega^\perp} \left(\mathrm{U} *_D \mathrm{V}^T + \mathcal{W}^{\mathcal{L}} + \mathcal{W}^{\mathcal{S}}\right)\right\|_\infty \leq \frac{\lambda}{2}. \end{cases} \tag{35}$$

This can be done by using the following two lemmas.

Lemma 2. *Assume that $\Omega \sim$ Ber(ρ) with $\rho \leq \rho_s$ for some $\rho_s > 0$. Set $j_0 = 2\lceil \log n \rceil$ (use $\log n_{(1)}$ for rectangular matrices). Then, the $\mathcal{W}^{\mathcal{L}}$ in Eq. (31) obeys*

(a) $\|\mathcal{W}^{\mathcal{L}}\| < 1/4$,

(b) $\left\|\mathcal{P}_\Omega \left(\mathrm{U} *_D \mathrm{V}^T + \mathcal{W}^{\mathcal{L}}\right)\right\|_F \leq \frac{\lambda}{4}$,

(c) $\left\|\mathcal{P}_{\Omega^\perp} \left(\mathrm{U} *_D \mathrm{V}^T + \mathcal{W}^{\mathcal{L}}\right)\right\|_\infty \leq \frac{\lambda}{4}$.

Lemma 3. *Assume $\Omega \sim$ Ber(ρ_s), and the sign of $\mathcal{S}_0$ are independent and identically distributed symmetric (and independent of Ω). Then, the tensor $\mathcal{W}^{\mathcal{S}}$ with Eq. (33) obeys*

(a) $\|\mathcal{W}^{\mathcal{S}}\| < 1/4$,

(b) $\|\mathcal{P}_{\Omega^\perp} \mathcal{W}^{\mathcal{S}}\|_\infty < \lambda/4$.

3.3 Proofs of Dual Certification

Before proving Lemma 2 and 3, we shall list the following five useful lemmas. The proofs of these lemmas are presented in the next chapter.

Lemma 4. *For the Bernoulli sign variable $\mathcal{M} \in \mathbb{R}^{n \times n \times n_3}$ defined as*

$$\mathcal{W}_{ijk} = \begin{cases} 1, & w.p. \quad \rho/2, \\ 0, & w.p. \quad 1-\rho, \\ -1, & w.p. \quad \rho/2, \end{cases} \tag{36}$$

where $\rho > 0$, there exists a function $\phi(\rho)$ satisfying $\lim_{\rho \to 0^+} \phi(\rho) = 0$, such that the following statement holds with large probability

$$\|\mathcal{M}\| \leq \phi(\rho)\sqrt{nn_3}. \tag{37}$$

Lemma 5. *Suppose $\Omega \sim$ Ber(ρ). Then with high probability,*

$$\|\mathcal{P}_T - \rho^{-1} \mathcal{P}_T \mathcal{P}_\Omega \mathcal{P}_T\| \leq \epsilon, \tag{38}$$

provided that $\rho \geq C_0 \epsilon^{-2} \beta \mu R \log(n)/(n)$ for some numerical constant $C_0 > 0$. For the tensor of rectangular frontal slice, we need $\rho \geq C_0 \epsilon^{-2} \beta \mu R \log(n_{(1)})/(n_{(2)})$, where $n_{(1)} = \max\{n_1, n_2\}, n_{(2)} = \min\{n_1, n_2\}$.

Lemma 6. *Assume that $\Omega \sim$ Ber(ρ), then $\|\mathcal{P}_\Omega \mathcal{P}_T\|^2 \leq \rho + \epsilon$, provided that $1 - \rho \geq C\epsilon^{-2}(\mu R \log(n)/n)$, where C is as in Lemma 5. For the tensor with frontal slice, the modification is as in Lemma 5.*

Lemma 7. *Suppose $\mathcal{Z} \in T$ is a fixed tensor, and $\Omega \sim$ Ber(ρ_0). Then with high probability,*

$$\|\mathcal{Z} - \rho^{-1} \mathcal{P}_T \mathcal{P}_\Omega \mathcal{Z}\|_\infty \leq \epsilon \|\mathcal{Z}\|_\infty, \tag{39}$$

provided that $\rho \geq C_0 \epsilon^{-2} \beta \mu R \log(n)/(n)$ for some numerical constant $C_0 > 0$. For the tensor of rectangular frontal slice, we need $\rho \geq C_0 \epsilon^{-2} \beta \mu R \log(n_{(1)})/(n_{(2)})$.

Lemma 8. *Suppose $\mathcal{Z}$ is fixed, and $\Omega \sim$ Ber(ρ_0). Then with high probability,*

$$\|\left(\mathcal{I} - \rho^{-1} \mathcal{P}_\Omega\right) \mathcal{Z}\| \leq \sqrt{\frac{C_0 n \log(n)}{\rho}} \|\mathcal{Z}\|_\infty, \tag{40}$$

provided that $\rho \geq C_0 \log(n)/(n)$ for some small numerical constant $C_0 > 0$. For the tensor of rectangular frontal slice, we need $\rho \geq C_0 \log(n_{(1)})/(n_{(2)})$.

Proof of Lemma 2

Proof. We first introduce some notations. Setting

$$\mathcal{Z}_j = \mathcal{U} *_D \mathcal{V}^T - \mathcal{P}_T \mathcal{Y}_j,$$

thus $\mathcal{Z}_j \in T$ for all $j \geq 0$. From the definition of $\mathcal{Y}_j$ (32), and $\mathcal{Y}_j \in \Omega^\perp$, we have

$$\mathcal{Z}_j = (\mathcal{P}_T - q^{-1} \mathcal{P}_T \mathcal{P}_{\Omega_j} \mathcal{P}_T) \mathcal{Z}_{j-1},$$
$$\mathcal{Y}_j = \mathcal{Y}_{j-1} + q^{-1} \mathcal{P}_{\Omega_j} \mathcal{Z}_{j-1}.$$

Therefore, when

$$q \geq C_0 \epsilon^{-2} \mu R \log(n_{(1)})/(n_{(2)}), \tag{41}$$

we have

$$\|\mathcal{Z}_j\|_\infty \le \epsilon\|\mathcal{Z}_{j-1}\|_\infty \le \epsilon^j\|\mathcal{U} *_D \mathcal{V}^T\|_\infty \tag{42}$$

by Lemma 7. When q obeys Eq. (41), we have

$$\|\mathcal{Z}_j\|_F \le \epsilon\|\mathcal{Z}_{j-1}\|_F \le \epsilon^j\|\mathcal{U} *_D \mathcal{V}^T\|_F \le \epsilon^j\sqrt{R} \tag{43}$$

by Lemma 5. We assume $\epsilon \le e^{-1}$.

proof of (a). Since $\mathcal{Y}_{j_0} = \sum_j q^{-1}\mathcal{P}_{\Omega_j}\mathcal{Z}_{j-1}$, we have

$$\begin{aligned}
\|\mathcal{W}^{\mathcal{L}}\| &= \|\mathcal{P}_{T^\perp}\mathcal{Y}_{j_0}\|_\infty \le \sum_j \|q^{-1}\mathcal{P}_{T^\perp}\mathcal{P}_{\Omega_j}\mathcal{Z}_{j-1}\| \\
&\le \sum_j \|\mathcal{P}_{T^\perp}(q^{-1}\mathcal{P}_{\Omega_j}\mathcal{Z}_{j-1} - \mathcal{Z}_{j-1})\| \\
&\le \sum_j \|q^{-1}\mathcal{P}_{\Omega_j}\mathcal{Z}_{j-1} - \mathcal{Z}_{j-1}\| \\
&\le C_1\sqrt{\frac{n_{(1)}\log(n_{(1)})}{q}}\sum_j \|\mathcal{Z}_{j-1}\|_\infty \\
&\le C_1\sqrt{\frac{n_{(1)}\log(n_{(1)})}{q}}\sum_j \epsilon^j\|\mathcal{U} *_D \mathcal{V}^T\|_\infty \\
&\le \frac{C_1}{(1-\epsilon)}\sqrt{\frac{n_{(1)}\log(n_{(1)})}{q}}\|\mathcal{U} *_D \mathcal{V}^T\|_\infty.
\end{aligned} \tag{44}$$

The fourth step is according to Lemma 8 and the fifth step can be directly obtained from Eq. (42). Now by using Eq. (41) and tensor incoherence condition, we get

$$\|\mathcal{W}^{\mathcal{L}}\| \le C_2\epsilon$$

for some numerical constant C_2.

proof of (b). Since $\mathcal{P}_\Omega\mathcal{Y}_{j_0} = 0$,

$$\begin{aligned}
\mathcal{P}_\Omega(\mathcal{U} *_D \mathcal{V}^T + \mathcal{W}^{\mathcal{L}}) &= \mathcal{P}_\Omega(\mathcal{U} *_D \mathcal{V}^T + \mathcal{P}_{T^\perp}\mathcal{Y}_{j_0}) \\
&= \mathcal{P}_\Omega(\mathcal{U} *_D \mathcal{V}^T + \mathcal{P}_T\mathcal{Y}_{j_0}) = \mathcal{P}_\Omega(\mathcal{Z}_{j_0}).
\end{aligned}$$

By using Eqs. (41), we can get

$$\|\mathcal{P}_\Omega(\mathcal{Z}_{j_0})\|_F \le \|\mathcal{Z}_{j_0}\|_F \le \epsilon^{j_0}\sqrt{R}.$$

Since $\epsilon \le e^{-1}$, $j_0 \ge 2\log(n_{(1)})$ and $\epsilon^{j_0} \le 1/(n_{(1)})^2$, and this proves the claim.

proof of (c). We have $\mathcal{U} *_D \mathcal{V}^T + \mathcal{W}^{\mathcal{L}} = \mathcal{Z}_{j_0} + \mathcal{Y}_{j_0}$ and know that $\mathcal{Y}_{j_0}$ is supported on Ω^c. Therefore, since $\|\mathcal{Z}_{j_0}\|_\infty \le \|\mathcal{Z}_{j_0}\|_F \le \lambda/8$, it suffices to show that $\|\mathcal{Y}_{j_0}\|_\infty \le \frac{\lambda}{8}$. To this end, we deduce

$$\begin{aligned}
\|\mathcal{Y}_{j_0}\|_\infty &\le q^{-1}\sum_j \|\mathcal{P}_{\Omega_j}\mathcal{Z}_{j_0}\|_\infty \le q^{-1}\sum_j \|\mathcal{Z}_{j_0}\|_\infty \\
&\le q^{-1}\sum_j \epsilon^j\|\mathcal{U} *_D \mathcal{V}^T\|_\infty.
\end{aligned}$$

Since $\|\mathcal{U} *_D \mathcal{V}^T\|_\infty \le \sqrt{\mu n^{-2} r}$, this gives

$$\|\mathcal{Y}_{j_0}\|_\infty \le C^{'}\frac{\epsilon^2}{\sqrt{\mu r(\log(n))^2}} \tag{45}$$

for some numerical constant $C^{'}$ whenever q obeys Eq. (41). By setting $\lambda = 1/\sqrt{n_{(1)}}$, $\|\mathcal{Y}_{j_0}\|_\infty \le \lambda/8$ if

$$\epsilon \le C\left(\frac{\mu r(\log(n_{(1)}))^2}{n_{(2)}}\right)^{\frac{1}{4}}.$$

We have seen that (a) and (b) are satisfied if ϵ is sufficiently small and $j_0 \ge 2\log(n_{(1)})$. For (c), we can take ϵ on the order of $\left(\mu r(\log(n_{(1)}))^2/(n_{(2)})\right)^{\frac{1}{4}}$, which could be sufficiently small as well provided that ρ_r in Eq. (41) in the manuscript is sufficiently small. Note that everything is consistent, since $C_0\epsilon^{-2}\mu r\log(n_{(1)})/(n_{(2)}) < 1$.

□

Proof of Lemma 3

Proof. We denote $\mathcal{M} = \mathrm{sgn}(\mathcal{E}_0)$ distributed as

$$\mathcal{M}_{ijk} = \begin{cases} 1, & w.p. \quad \rho/2, \\ 0, & w.p. \quad 1-\rho, \\ -1, & w.p. \quad \rho/2. \end{cases} \tag{46}$$

Note that for any $\sigma > 0$, $\{\|\mathcal{P}_\Omega\mathcal{P}_T\| \le \sigma\}$ holds with high probability provided that ρ is sufficiently small, see Lemma 5.

1. Proof of (a).

By construction,

$$\begin{aligned}
\mathcal{W}^{\mathcal{S}} &:= \lambda\mathcal{P}_{T^\perp}\mathcal{M} + \lambda\mathcal{P}_{T^\perp}\sum_{k\ge1}(\mathcal{P}_\Omega\mathcal{P}_T\mathcal{P}_\Omega)^k\mathcal{M} \\
&:= \mathcal{P}_{T^\perp}\mathcal{W}_0^{\mathcal{S}} + \mathcal{P}_{T^\perp}\mathcal{W}_1^{\mathcal{S}}.
\end{aligned} \tag{47}$$

Note that $\|\mathcal{P}_{T^\perp}\mathcal{W}_0^{\mathcal{S}}\| \le \|\mathcal{W}_0^{\mathcal{S}}\| = \lambda\|\mathcal{M}\|$ and $\|\mathcal{P}_{T^\perp}\mathcal{W}_1^{\mathcal{S}}\| \le \|\mathcal{W}_1^{\mathcal{S}}\| = \lambda\|\mathcal{R}(\mathcal{M})\|$, where $\mathcal{R} = \sum_{k\ge1}(\mathcal{P}_\Omega\mathcal{P}_T\mathcal{P}_\Omega)^k$. Now, we will respectively show that $\lambda\|\mathcal{M}\|$ and $\lambda\|\mathcal{R}(\mathcal{M})\|$ are small enough when ρ is sufficiently small for $\lambda = 1/\sqrt{n}$. Therefore, $\|\mathcal{W}^{\mathcal{S}}\| \le 1/4$.

1) Bound $\lambda\|\mathcal{M}\|$**.** By using Lemma 4 directly, we have that $\lambda\|\mathcal{M}\| \le \phi(\rho)$ is sufficiently small given $\lambda = 1/\sqrt{n}$ and ρ is sufficiently small.

2) Bound $\|\mathcal{R}(\mathcal{M})\|$**.** For simplicity, let $\mathcal{Z} = \mathcal{R}(\mathcal{M})$, we have

$$\|\mathcal{Z}\| = \|\overline{\mathbf{Z}}\| = \sup_{x\in\mathbb{S}^{nr_3-1}}\|\overline{\mathbf{Z}}x\|_2. \tag{48}$$

The optimal x to Eq. (48) is an eigenvector of $\overline{\mathbf{Z}}{*}\overline{\mathbf{Z}}$. Since $\overline{\mathbf{Z}}$ is a block diagonal matrix, the optimal x has a block sparse structure, i.e., $x \in B = \{x \in \mathbb{R}^{nr_3} | x = [x_1^T, \cdots, x_i^T, \cdots, x_{r_3}^T]$, with $x_i \in \mathbb{R}^n$, and there exist j such that $x_j \ne 0$ and $x_i \ne 0, i \ne j\}$. Note that $\|x\|_2 = \|x_j\|_2 = 1$. Let N be the $1/2$-net for $\mathbb{S}^{n-1}$ of size at most 5^n (see Lemma 5.2 in [4]). Then the $1/2$-net, denote as N', for B has the size at most $r_3.5^n$. We have

$$\begin{aligned}
\|\mathcal{R}(\mathcal{M})\| &= \|\mathrm{bdiag}(\overline{\mathcal{R}(\mathcal{M})})\| = \sup_{x,y\in B}\left\langle x, \mathrm{bdiag}(\overline{\mathcal{R}(\mathcal{M})})y\right\rangle \\
&= \sup_{x,y\in B}\left\langle xy^*, \mathrm{bdiag}(\overline{\mathcal{R}(\mathcal{M})})\right\rangle \\
&= \sup_{x,y\in B}\left\langle \mathrm{bdiag}^*(xy^*), \overline{\mathcal{R}(\mathcal{M})}\right\rangle,
\end{aligned} \tag{49}$$

where bdiag*, the joint operator of bdiag (see definition in the manuscript), maps the block diagonal matrix xy^* to a tensor of size $n \times n \times n_3$. Let $\mathcal{Z}' = \text{bdiag}^*(xy^*)$ and $\mathcal{Z} = \mathcal{Z}' \times_3 D$. We have

$$\begin{aligned}\|\mathcal{R}(\mathcal{M})\| &= \sup_{x,y\in B} \left\langle \mathcal{Z}', \overline{\mathcal{R}(\mathcal{M})} \right\rangle = \sup_{x,y\in B} \langle \mathcal{Z}', \mathcal{R}(\mathcal{M})\rangle \\ &= \sup_{x,y\in B} \langle \mathcal{R}(\mathcal{Z}), \mathcal{M}\rangle = \sup_{x,y\in N'} 4\langle \mathcal{R}(\mathcal{Z}), \mathcal{M}\rangle .\end{aligned} \tag{50}$$

For a fixed pair (x, y) of unit-normed vectors, define the random variable

$$X(x, y) = 4\langle \mathcal{R}(\mathcal{Z}), \mathcal{M}\rangle . \tag{51}$$

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

$$\mathbb{P}(|X(x, y)| > t|\Omega) \le 2\exp\left(\frac{-2t^2}{\|4\mathcal{R}(\mathcal{Z})\|_F^2}\right). \tag{52}$$

Note that $\|4\mathcal{R}(\mathcal{Z})\|_F^2 \le 4\|\mathcal{R}\|\|\mathcal{Z}\|_F = 4\|\mathcal{R}\|\|\mathcal{Z}'\|_F = 4\|\mathcal{R}\|$. Therefore, we have

$$\mathbb{P}\left(\sup_{x,y\in N'} |X(x, y)| > t|\Omega\right) \le 2|N'|^2 \exp\left(\frac{-t^2}{8\|\mathcal{R}\|^2}\right). \tag{53}$$

Hence,

$$\mathbb{P}(\|\mathcal{R}(\mathcal{M})\| > t|\Omega) \le 2|N'|^2 \exp\left(\frac{-t^2}{8\|\mathcal{R}\|^2}\right). \tag{54}$$

On the event $\{\|\mathcal{P}_\Omega\mathcal{P}_T\| \le \sigma\}$, $\|\mathcal{R}\| \le \sum_{k\ge1} \sigma^{2k} = \frac{\sigma^2}{1-\sigma^2}$, therefore, unconditionally,

$$\begin{aligned}&\mathbb{P}(\|\mathcal{R}(\mathcal{M})\| > t) \le 2|N'|^2 \exp\left(\frac{-\gamma^2t^2}{8}\right) \\ &+ \mathbb{P}(\|\mathcal{P}_\Omega\mathcal{P}_T\| \ge \sigma), \gamma = \frac{1-\sigma^2}{\sigma^2} \\ &= 2r_3^2.5^{2n} \exp\left(\frac{-\gamma^2t^2}{8}\right) + \mathbb{P}(\|\mathcal{P}_\Omega\mathcal{P}_T\| \ge \sigma).\end{aligned} \tag{55}$$

Let $t = c\sqrt{n}$, where c can be a small absolute constant. Then the above inequality implies that $\|\mathcal{R}(\mathcal{M})\| \le t$ with high probability.

2. Proof of (b). Observe that

$$\mathcal{P}_{\Omega^\perp}\mathcal{W}^\mathcal{S} = -\lambda\mathcal{P}_{\Omega^\perp}\mathcal{P}_T(\mathcal{P}_\Omega - \mathcal{P}_\Omega\mathcal{P}_T\mathcal{P}_\Omega)^{-1}\mathcal{M}. \tag{56}$$

Note for $(i, j, k) \in \Omega^c, \mathcal{W}^\mathcal{S}_{ijk} = \langle \mathcal{W}, \mathfrak{e}_{ijk}\rangle$, and we have $\mathcal{W}^\mathcal{S}_{ijk} = \lambda\langle \mathcal{Q}(i, j, k), \mathcal{W}\rangle$, where $\mathcal{Q}(i, j, k)$ is the tensor $-(\mathcal{P}_\Omega - \mathcal{P}_\Omega\mathcal{P}_T\mathcal{P}_\Omega)^{-1}\mathcal{P}_\Omega\mathcal{P}_T(\mathfrak{e}_{ijk})$. Conditional on $\Omega = \text{supp}(\mathcal{M})$, the signs of $\mathcal{M}$ are independent and identically distributed symmetric, and the Hoeffding's inequality gives

$$\mathbb{P}\left(\|\mathcal{W}^\mathcal{S}_{ijk}\| > t\lambda|\Omega\right) \le 2\exp\left(-\frac{2t^2}{\|\mathcal{Q}(i, j, k)\|_F^2}\right), \tag{57}$$

and

$$\begin{aligned}&\mathbb{P}\left(\sup_{i,j,k} \|\mathcal{W}^\mathcal{S}_{ijk}\| > t\lambda/n_3|\Omega\right) \\ &\le 2n^2n_3 \exp\left(-\frac{2t^2}{\sup_{i,j,k} \|\mathcal{Q}(i, j, k)\|_F^2}\right),\end{aligned} \tag{58}$$

By using Eq. (12), we have

$$\|\mathcal{P}_\Omega\mathcal{P}_T(\mathfrak{e}_{ijk})\|_F \le \|\mathcal{P}_\Omega\mathcal{P}_T\|\|\mathcal{P}_T(\mathfrak{e}_{ijk})\|_F \le \sigma\sqrt{\frac{2\mu R}{n}}, \tag{59}$$

on the event $\{\|\mathcal{P}_\Omega\mathcal{P}_T\| \le \sigma\}$. On the same event, we have $\|(\mathcal{P}_\Omega - \mathcal{P}_\Omega\mathcal{P}_T\mathcal{P}_\Omega)^{-1}\|(1-\sigma^2)^{-1}$ and thus $\|\mathcal{Q}(i, j, k)\|_F^2 \le \frac{2\sigma^2}{(1-\sigma^2)^2}\frac{\mu R}{n}$. Then, unconditionally,

$$\begin{aligned}\mathbb{P}\left(\sup_{i,j,k} |\mathcal{W}^\mathcal{S}_{ijk}| > t\lambda\right) &\le 2n^2n_3 \exp\left(-\frac{n\gamma^2t^2}{\mu R}\right) \\ &+ \mathbb{P}(\|\mathcal{P}_\Omega\mathcal{P}_T\| \ge \sigma),\end{aligned} \tag{60}$$

where $\gamma^2 = \frac{(1-\sigma^2)^2}{2\sigma^2}$. This proves the claim when $\mu R \le \rho_r' n \log(n)^{-1}$ and ρ_r' is sufficiently small.

□

3.4 Proof of Some Lemmas

Before the proof, we introduce a theorem.

Theorem 5. (Noncommutative Bernstein Inequality) Let $\mathrm{X}_1, \mathrm{X}_2, \cdots, \mathrm{X}_L$ be independent zero-mean random matrices of dimension $d_1 \times d_2$. Suppose $\|\mathrm{X}_k\| \le M$ and

$$\rho_k^2 = \max\{\|\mathbb{E}[\mathrm{X_k X_k^T}]\|, \|\mathbb{E}[\mathrm{X_k^T X_k}]\|\} \tag{61}$$

almost surely for all k. Then for any $\tau > 0$,

$$\mathbb{P}\left[\|\sum_{k=1}^{L} \mathrm{X_k}\| > \tau\right] \le (d_1 + d_2)\exp\left(\frac{-\tau^2/2}{\sum_{\mathrm{k=1}}^{\mathrm{L}} \rho_{\mathrm{k}}^2 + \mathrm{M}\tau/3}\right) \tag{62}$$

This theorem is a corollary of a Chernoff bound for finite dimension operators developed by [5]. An extension of this theorem [6] states that if

$$\max\left\{\|\sum_{\mathrm{k=1}}^{\mathrm{L}} \mathrm{X_k X_k^T}\|, \|\sum_{\mathrm{k=1}}^{\mathrm{L}} \mathrm{X_k^T X_k}\|\right\} \le \sigma^2 \tag{63}$$

and let

$$\tau = \sqrt{4c\sigma^2 \log(d_1 + d_2)} + cM \log(d_1 + d_2) \tag{64}$$

for any $c > 0$. Then Eq. (62) becomes

$$\mathbb{P}\left[\|\sum_{k=1}^{L} \mathrm{X}_k\| > \tau\right] \le (d_1 + d_2)^{-(c-1)}. \tag{65}$$

Proof of Lemma 4

Proof. The proof has three steps.

We first introduce some notations. Let f^* be the i-th row of $D^T \in \mathbb{R}^{n_3\times r_3}$, and $\mathrm{M}^H = [\mathrm{M}_1^H; \mathrm{M}_2^H; \cdots; \mathrm{M}_n^H] \in \mathbb{R}^{nr_3\times n}$ be a matrix unfolded by $\mathcal{M}$, where $\mathrm{M}_i^H \in \mathbb{R}^{r_3\times n}$ is the i-th horizontal slice of $\mathcal{M}$, i.e., $[\mathrm{M}_i^H]_{kj} = \mathcal{M}_{ijk}$. Consider that $\overline{\mathcal{M}} = \mathcal{M} \times_3 D^T$, we have $\overline{M}_i = [f_i^*\mathrm{M}_1^H; f_i^*\mathrm{M}_2^H; \cdots; f_i^*\mathrm{M}_n^H]$, where $\overline{M}_i \in \mathbb{R}^{n\times n}$ is the i-th frontal slice of $\mathcal{M}$. Note that

$$\|\mathcal{M}\| = \|\overline{\mathrm{M}}\| = \max_{i=1,\cdots,r_3} \|\overline{\mathrm{M}}_i\|. \tag{66}$$

Let N be the $1/2$-net for $\mathbb{S}^{n-1}$ of size at most 5^n (see Lemma 5.2 in [4]). Then Lemma 5.3 in [4] gives

$$\|\overline{\mathrm{M}}_i\| \le 2 \max_{x \in N} \|\overline{\mathrm{M}}_i x\|_2. \tag{67}$$

So we consider to bound $\|\overline{\mathrm{M}}_i x\|_2$.

We can express $\|\overline{\mathrm{M}}_i x\|_2^2$ as a sum of independent random variables

$$\|\overline{\mathrm{M}}_i x\|_2^2 = \sum_{j=1}^{n} (\mathrm{f}_i^* M_j^H x)^2 := \sum_{j=1}^{n} z_j^2, \tag{68}$$

where $z_j = \langle M_j^H, \mathrm{f}_i x^* \rangle, j = 1, \cdots, n$ are independent sub-gaussian random variables with $\mathbb{E}(z_j^2) = \rho \|\mathrm{f}_i x^*\|_f^2 = \rho r_3$. Using Eq. (36), we have

$$|[M_j^H]_{kl}| = \begin{cases} 1, & w.p.\ \rho, \\ 0, & w.p.\ 1-\rho. \end{cases} \tag{69}$$

Thus, the sub-gaussian norm of $[M_j^H]_{kl}$, denoted as $\|\cdot\|_{\psi_2}$, is

$$\|[M_j^H]_{kl}\|_{\psi_2} = \sup_{p \ge 1} p^{-0.5} (\mathbb{E}[|[M_j^H]_{kl}|^p])^{1/p} = \sup_{p \ge 1} p^{-0.5} \rho^{1/p}. \tag{70}$$

Define the function $\phi(x) = x^{-1/2} \rho^{1/x}$ on $[1, +\infty)$. The only stationary point occurs at $x^* = \log \rho^{-2}$. Thus,

$$\begin{aligned} \phi(x) &\le \max\{\phi(1), \phi(x^*)\} \\ &= \max\left(\rho, (\log \rho^{-2})^{-0.5} \rho^{1/\log \rho^{-2}}\right) := \psi(\rho). \end{aligned} \tag{71}$$

Therefore, $\|[M_j^H]_{kl}\|_{\psi_2} \le \psi(\rho)$. Consider that z_j is a sum of independent centered sub-gaussian random variables $[M_j^H]_{kl}$'s, bu using Lemma 5.9 in [4], we have $\|z_j\|_{\psi_2}^2 \le c_1(\psi(\rho))^2 r_3$, where c_1 is an absolute constant. Therefore, by Remark 5.18 and Lemma 5.14 in [4], $z_j^2 - \rho r_3$ are independent centered sub-exponential random variables with $\|z_j^2 - \rho r_3\|_{\psi_1} \le 2\|z_j\|_{\psi_1}^2 \le 4\|z_j\|_{\psi_2}^2 \le 4c_1(\psi(\rho))^2 r_3$.

Now, we use an exponential deviation inequality, Corollary 5.17 in [4], to control the sum of Eq. (68). We have

$$\begin{aligned} \mathbb{P}(|\|\overline{\mathrm{M}}_i \boldsymbol{x}\|_2^2 - \rho n r_3| \ge tn) &= \mathbb{P}\left(\left|\sum_{j=1}^{n} (z_j^2 - \rho r_3)\right|\right) \\ &\le 2\exp\left(-c_2 n \min\left(\left(\frac{t}{4c_1(\psi(\rho))^2 r_3}\right)^2, \frac{t}{4c_1(\psi(\rho))^2 r_3}\right)\right), \end{aligned} \tag{72}$$

where $c_2 > 0$. Let $t = c_3(\psi(\rho))^2 r_3$ for some absolute constant c_3, we have

$$\begin{aligned} &\mathbb{P}(|\|\overline{\mathrm{M}}_i \boldsymbol{x}\|_2^2 - \rho n r_3| \ge c_3(\psi(\rho))^2 n r_3) \\ &\le 2\exp\left(-c_2 n \min\left(\left(\frac{c_3}{4c_1}\right)^2, \frac{c_3}{4c_1}\right)\right). \end{aligned} \tag{73}$$

Step 3: Union bound. Taking the union bound over all $\boldsymbol{x}$ in the Net N of cardinality $|N| \le 5^n$, we obtain

$$\begin{aligned} &\mathbb{P}\left(\left|\max_{\boldsymbol{x} \in N} \|\overline{\mathrm{M}}_i \boldsymbol{x}\|_2^2 - \rho n r_3\right| \ge c_3(\psi(\rho))^2 n r_3\right) \\ &\le 2 \cdot 5^n \cdot \exp\left(-c_2 n \min\left(\left(\frac{c_3}{4c_1}\right)^2, \frac{c_3}{4c_1}\right)\right). \end{aligned} \tag{74}$$

Furthermore, taking the union over all $i = 1, \cdots, r_3$, we have

$$\begin{aligned} &\mathbb{P}\left(\max_i \left|\max_{\boldsymbol{x} \in N} \|\overline{\mathbf{M}}_i \boldsymbol{x}\|_2^2 - \rho n r_3\right| \ge c_3(\psi(\rho))^2 n r_3\right) \\ &\le 2 \cdot 5^n \cdot r_3 \cdot \exp\left(-c_2 n \min\left(\left(\frac{c_3}{4c_1}\right)^2, \frac{c_3}{4c_1}\right)\right). \end{aligned} \tag{75}$$

This implies that, with high probability (when the constant c_3 is large enough),

$$\max_i \max_{\boldsymbol{x} \in N} \|\overline{\mathrm{M}}_i \boldsymbol{x}\|_2^2 \le \left(\rho + c_3(\psi(\rho))^2\right) n r_3 \tag{76}$$

Let $\phi(\rho) = 2\sqrt{\rho + c_3(\psi(\rho))^2}$ and it satisfies $\lim_{\rho \to 0^+} \phi(\rho) = 0$ by using Eq. (71). The proof is completed by further combing Eq. (66), (67) and (76). ☐

Proof of Lemma 5

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

$$\begin{aligned} &(\rho^{-1} \mathcal{P}_T \mathcal{P}_\Omega \mathcal{P}_T - \mathcal{P}_T)\mathcal{Z} \\ &= \sum_{ijk} (\rho^{-1} \delta_{ijk} - 1) \langle \mathfrak{e}_{ijk}, \mathcal{P}_T \mathcal{Z} \rangle \mathcal{P}_T(\mathfrak{e}_{ijk}) \\ &:= \sum_{ijk} \mathcal{H}_{ijk}(\mathcal{Z}) \end{aligned} \tag{77}$$

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

$$\overline{H}_{ijk}(\overline{Z}) = (\rho^{-1} \delta_{ijk} - 1) \langle \mathfrak{e}_{ijk}, \mathcal{P}_T(\mathcal{Z}) \rangle \,\mathrm{bdiag}(\overline{\mathcal{P}_T(\mathfrak{e}_{ijk})}). \tag{78}$$

By the above definitions, we have $\mathcal{H}_{ijk} = \overline{H}_{ijk}$ and $\|\sum_{ijk} \mathcal{H}_{ijk}\| = \|\sum_{ijk} \overline{H}_{ijk}\|$. Also, $\overline{H}_{ijk}$ is self-adjoint and $\mathbb{E}[\overline{H}_{ijk}] = 0$. To prove the result by the non-commutative Bernstein inequality, we need to bound $\|\overline{H}_{ijk}\|$ and $\|\sum_{ijk} \mathbb{E}[\overline{H}_{ijk}^2]\|$. First, we have

$$\begin{aligned} \|\overline{H}_{ijk}\| &= \sup_{\|\overline{Z}\|_F = 1} \|\overline{H}_{ijk}(\overline{Z})\|_F \\ &\le \sup_{\|\overline{Z}\|_F = 1} \|\mathcal{P}_T(\mathfrak{e}_{ijk})\|_F \|\mathrm{bdiag}(\overline{\mathcal{P}_T(\mathfrak{e}_{ijk})})\|_F \|\mathcal{Z}\|_F \\ &= \sup_{\|\overline{Z}\|_F = 1} \|\mathcal{P}_T(\mathfrak{e}_{ijk})\|_F^2 \|\overline{Z}\|_F \le \frac{2\mu R}{n\rho}, \end{aligned} \tag{79}$$

where the last inequality use Eq. (12). On the other hand, by direct computation, we have $\mathcal{H}_{ijk}^2(\overline{Z}) = (\rho^{-1}\delta_{ijk} - 1)^2 \langle \mathfrak{e}_{ijk}, \mathcal{P}_T(\mathcal{Z}) \rangle \langle \mathfrak{e}_{ijk}, \mathcal{P}_T(\mathfrak{e}_{ijk}) \rangle \,\mathrm{bdiag}(\overline{\mathcal{P}_T(\mathfrak{e}_{ijk})})$. Note that $\mathbb{E}[(\rho^{-1}\delta_{ijk}-1)^2] \le \rho^{-1}$. We have

$$\begin{aligned}
&\left\|\sum_{ijk}\mathbb{E}[\overline{\mathrm{H}}_{ijk}^2(\overline{\mathrm{Z}})]\right\|_F \\
&\le \rho^{-1}\|\sum_{ijk}\langle\mathfrak{e}_{ijk},\mathcal{P}_T(\mathcal{Z})\rangle\langle\mathfrak{e}_{ijk},\mathcal{P}_T(\mathfrak{e}_{ijk})\rangle \\
&\qquad \mathrm{bdiag}(\overline{\mathcal{P}_T(\mathfrak{e}_{ijk})})\|_F \\
&\le \rho^{-1}\|\mathcal{P}_T(\mathfrak{e}_{ijk})\|_F^2\left\|\sum_{ijk}\langle\mathfrak{e}_{ijk},\mathcal{P}_T(\mathcal{Z})\rangle\right\|_F \\
&\le \rho^{-1}\|\mathcal{P}_T(\mathfrak{e}_{ijk})\|_F^2\|\mathcal{P}_T(\mathcal{Z})\|_F \\
&\le \rho^{-1}\|\mathcal{P}_T(\mathfrak{e}_{ijk})\|_F^2\|\mathcal{Z}\|_F \\
&\le \rho^{-1}\|\mathcal{P}_T(\mathfrak{e}_{ijk})\|_F^2\|\overline{\mathrm{Z}}\|_F \le \frac{2\mu R}{n\rho}\|\overline{\mathbf{Z}}\|_F.
\end{aligned} \tag{80}$$

By Theorem 5, we have

$$\begin{aligned}
&\mathbb{P}[\|\rho^{-1}\mathcal{P}_T\mathcal{P}_\Omega\mathcal{P}_T-\mathcal{P}_T\|>\epsilon] \\
&=\mathbb{P}\left[\left\|\sum_{ijk}\mathcal{H}_{ijk}\right\|>\epsilon\right]=\mathbb{P}\left[\left\|\sum_{ijk}\overline{\mathrm{H}}_{ijk}\right\|>\epsilon\right] \\
&\le 2nr_3\exp\left(-\frac{3}{8}\cdot\frac{\epsilon^2}{2\mu R/(n\rho)}\right)\le 2(n)^{1-3C_0/16},
\end{aligned} \tag{81}$$

where the last inequality uses $\rho \ge C_0\epsilon^{-2}\mu R\log(n)/(n)$. Thus, $\|\rho^{-1}\mathcal{P}_T\mathcal{P}_\Omega\mathcal{P}_T-\mathcal{P}_T\|\le\epsilon$ holds with high probability for some numerical constant C_0. □

Proof of Lemma 6

Proof. From the proof of Lemma 5 (i.e., the last subsection), we have

$$\|\mathcal{P}_T-(1-\rho)^{-1}\mathcal{P}_T\mathcal{P}_{\Omega^\perp}\mathcal{P}_T\|\le\epsilon, \tag{82}$$

provided that $1-\rho\ge C_0\epsilon^{-2}(\mu R\log(n)/n)$. Note that $\mathcal{I}=\mathcal{P}_\Omega+\mathcal{P}_{\Omega^\perp}$, we have

$$\|\mathcal{P}_T-(1-\rho)^{-1}\mathcal{P}_T\mathcal{P}_{\Omega^\perp}\mathcal{P}_T\|=(1-\rho)^{-1}(\mathcal{P}_T\mathcal{P}_\Omega\mathcal{P}_T-\rho\mathcal{P}_T). \tag{83}$$

Then, by the triangular inequality

$$\|\mathcal{P}_T\mathcal{P}_\Omega\mathcal{P}_T\|\le\epsilon(1-\rho)+\rho\|\mathcal{P}_T\|=\rho+\epsilon(1-\rho). \tag{84}$$

This proof is completed by using $\|\mathcal{P}_\Omega\mathcal{P}_T\|^2=\|\mathcal{P}_T\mathcal{P}_\Omega\mathcal{P}_T\|$. □

Proof of Lemma 7

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

$$\rho^{-1}\mathcal{P}_\Omega\mathcal{P}_T(\mathcal{Z})=\sum_{ijk}\rho^{-1}\delta_{ijk}z_{ijk}\mathcal{P}_T(\mathfrak{e}_{ijk}).$$

The (a,b,c)-th entry of $\rho^{-1}\mathcal{P}_\Omega\mathcal{P}_T(\mathcal{Z})-\mathcal{Z}$ can be written as a sum of independent random variables, i.e.,

$$\begin{aligned}
&\left\langle\rho^{-1}\mathcal{P}_\Omega\mathcal{P}_T(\mathcal{Z})-\mathcal{Z},\mathfrak{e}_{abc}\right\rangle \\
&=\sum_{ijk}(\rho^{-1}\delta_{ijk}-1)z_{ijk}\langle\mathcal{P}_T(\mathfrak{e}_{ijk}),\mathfrak{e}_{abc}\rangle \\
&:=\sum_{ijk}t_{ijk},
\end{aligned} \tag{85}$$

where t_{ijk}'s are independent and $\mathbb{E}(t_{ijk})=0$. Now next bound $|t_{ijk}|$ and $|\sum_{ijk}\mathbb{E}[t_{ijk}^2]|$. First

$$\begin{aligned}
|t_{ijk}|&\le\rho^{-1}\|\mathcal{Z}\|_\infty\|\mathcal{P}_T(\mathfrak{e}_{ijk})\|_F\|\mathcal{P}_T(\mathfrak{e}_{abc})\|_F \\
&\le\frac{2\mu R}{n\rho}\|\mathcal{Z}\|_\infty.
\end{aligned} \tag{86}$$

Second, we have

$$\begin{aligned}
\left|\sum_{ijk}\mathbb{E}[t_{ijk}^2]\right|&\le\rho^{-1}\|\mathcal{Z}\|_\infty^2\sum_{ijk}\langle\mathcal{P}_T(\mathfrak{e}_{ijk}),\mathfrak{e}_{abc}\rangle \\
&=\rho^{-1}\|\mathcal{Z}\|_\infty^2\sum_{ijk}\langle\mathfrak{e}_{ijk},\mathcal{P}_T(\mathfrak{e}_{abc})\rangle \\
&=\rho^{-1}\|\mathcal{Z}\|_\infty^2\|\mathcal{P}_T(\mathfrak{e}_{abc})\|_F^2\le\frac{2\mu R}{n\rho}\|\mathcal{Z}\|_\infty^2.
\end{aligned} \tag{87}$$

Let $\epsilon\le1$. By Theorem 5, we have

$$\begin{aligned}
&\mathbb{P}\left[|[\rho^{-1}\mathcal{P}_T\mathcal{P}_\Omega(\mathcal{Z})-\mathcal{Z}]_{abc}|\ge\epsilon\|\mathcal{Z}\|_\infty\right] \\
&=\mathbb{P}\left[\left|\sum_{ijk}[t_{ijk}]\right|\ge\epsilon\|\mathcal{Z}\|_\infty\right] \\
&\le 2\exp\left(-\frac{3}{8}\cdot\frac{\epsilon^2\|\mathcal{Z}\|_\infty^2}{2\mu R\|\mathcal{Z}\|_\infty^2/(n\rho)}\right)\le 2n^{-\frac{3}{16}C_0},
\end{aligned} \tag{88}$$

where the last inequality uses $\rho\ge C_0\epsilon^{-2}\mu R\log(n)/n$. Thus, $\|\rho^{-1}\mathcal{P}_T\mathcal{P}_\Omega(\mathcal{Z})-\mathcal{Z}\|_\infty\le\epsilon\|\mathcal{Z}\|_\infty$ holds with high probability for some numerical constant C_0. □

Proof of Lemma 8

Proof. Denote the tensor $\mathcal{H}_{ijk}=(1-\rho^{-1}\delta_{ijk})z_{ijk}\mathfrak{e}_{ijk}$. Then we have

$$(\mathcal{I}-\rho^{-1}\mathcal{P}_\Omega)\mathcal{Z}=\sum_{ijk}\mathcal{H}_{ijk}. \tag{89}$$

Note that δ_{ijk}'s are independent random scalars. Thus, $\mathcal{H}_{ijk}$'s are independent random tensors and $\overline{\mathrm{H}}_{ijk}$'s are independent random matrices. Observe that $\mathbb{E}[\overline{\mathrm{H}}_{ijk}]=\mathbf{0}$ and $\|\overline{\mathrm{H}}_{ijk}\|\le\rho^{-1}\|\mathcal{Z}\|_\infty$, we have

$$\begin{aligned}
&\left\|\sum_{ijk}\mathbb{E}[\overline{\mathrm{H}}_{ijk}^*\overline{\mathrm{H}}_{ijk}]\right\|=\left\|\sum_{ijk}\mathbb{E}[\mathcal{H}_{ijk}^*\mathcal{H}_{ijk}]\right\| \\
&=\left\|\sum_{ijk}\mathbb{E}[(1-\rho^{-1}\delta_{ijk})^2]z_{ijk}^2(\dot{\mathfrak{e}}_j *_L \dot{\mathfrak{e}}_j^*)\right\| \\
&=\left\|\frac{1-\rho}{\rho}\sum_{ijk}z_{ijk}^2(\dot{\mathfrak{e}}_j *_D \dot{\mathfrak{e}}_j^*)\right\|\le\frac{nn_3}{\rho}\|\mathcal{Z}\|_\infty^2.
\end{aligned} \tag{90}$$

A similar calculation yields $\left\|\sum_{ijk}\mathbb{E}[\overline{\mathrm{H}}_{ijk}^*\overline{\mathrm{H}}_{ijk}]\right\|\le\rho^{-1}nn_3\|\boldsymbol{\mathcal{Z}}\|_\infty^2$. Let $t=\sqrt{C_0nn_3\log(nn_3)/\rho}\|\mathcal{Z}\|_\infty$. When

$\rho \geq C_0 log(n)/n$, we apply Theorem 5 and obtain

$$\mathbb{P}[\|(\mathcal{I}-\rho^{-1}\mathcal{P}_\Omega)\mathcal{Z}\| > t] = \mathbb{P}\left[\left\|\sum_{ijk}\mathcal{H}_{ijk}\right\| > t\right]$$

$$= \mathbb{P}\left[\left\|\sum_{ijk}\overline{\mathrm{H}}_{ijk}\right\| > t\right] \tag{91}$$

$$\leq 2nr_3 \exp\left(-\frac{3}{8}\cdot\frac{C_0 n n_3 \log(nn_3)\|\mathcal{Z}\|_\infty^2/\rho}{nn_3\|\mathcal{Z}\|_\infty^2/\rho}\right)$$

$$\leq 2(nr_3)^{1-3C_0/8}.$$

Thus, $\|(\mathcal{I}-\rho^{-1}\mathcal{P}_\Omega)\mathcal{Z}\| > t$ holds with high probability for some numerical constant C_0. □

4 The Proof of Exact Recovery Theorem about TC Model

The following fact is used frequently in this section.

Lemma 9. *Suppose $\Omega \sim \mathrm{Ber(p)}$. Then with high probability,*

$$\|\mathcal{P}_T\mathcal{R}_\Omega\mathcal{P}_T - \mathcal{P}_T\| \leq \epsilon, \tag{92}$$

provided that $p \geq c_0\epsilon^{-2}(\mu R\log(n))/(n)$ for some numerical constant $c_0 > 0$. For the tensor of rectangular frontal slices, we need $p \geq c_0\epsilon^{-2}(\mu R\log(n_{(1)}))/(n_{(2)})$, where $n_{(1)} = \max\{n_1, n_2\}, n_{(2)} = \min\{n_1, n_2\}$.

Proof. By replacing $\rho^{-1}\mathcal{P}$ with $\mathcal{R}_\Omega$ in Lemma 5, this Lemma holds. □

Lemma 10. *Suppose that $\boldsymbol{\mathcal{Z}}$ is fixed, and $\boldsymbol{\Omega} \sim \mathrm{Ber(p)}$. Then, with high probability,*

$$\|(\mathcal{R}_\Omega - \mathcal{I})\mathcal{Z}\| \leq c\left(\frac{\log(n)}{p}\|\mathcal{Z}\|_\infty + \frac{\log(n)}{p}\|\mathcal{Z}\|_{\infty,2}\right), \tag{93}$$

for some numerical constant $c > 0$.

Lemma 11. *Suppose that $\mathcal{Z} \in T$ is a fixed tensor and $\Omega \sim \mathrm{Ber(p)}$. Then, with high probability,*

$$\|\mathcal{P}_T\mathcal{R}_\Omega\mathcal{Z} - \mathcal{Z}\|_{\infty,2} \leq \frac{1}{2}\sqrt{\frac{n}{\mu R}}\|\mathcal{Z}\|_\infty + \frac{1}{2}\|\mathcal{Z}\|_{\infty,2}, \tag{94}$$

provided that $p \geq c_0\mu R\log(n)/(n)$.

Lemma 12. *Suppose that $\mathcal{Z} \in T$ is a fixed tensor and$\Omega \sim \mathrm{Ber(p)}$. Then, with high probability,*

$$\|\mathcal{Z} - \mathcal{P}_T\mathcal{R}_\Omega(\mathcal{Z})\|_\infty \leq \epsilon\|\mathcal{Z}\|_\infty, \tag{95}$$

provided that $p \geq c_0\epsilon^{-2}(\mu R\log(n))/(n)$ (for the tensor of rectangular frontal slice, $p \geq c_0\epsilon^{-2}(\mu R\log(n_{(1)}))/(n_{(2)})$ for some numerical constant $c_0 > 0$.)

Proof. By replacing $\rho^{-1}\mathcal{P}$ with $\mathcal{R}_\Omega$ in Lemma 7, this Lemma holds. □

4.1 The Proof of Exact Recovery Theorem about TC Model

Proposition 1. *The tensor $\mathcal{T}_0$ is the unique optimal solution of TC model (14) if the following conditions hold: 1. $\|\mathcal{P}_T\mathcal{R}_\Omega\mathcal{P}_T - \mathcal{P}_T\| \leq \frac{1}{2}$.*

2. There exists a dual certificate $\mathcal{W} \in \mathbb{R}^{n_1\times n_2\times n_3}$ which satisfies $\mathcal{P}_\Omega(\mathcal{W}) = \mathcal{W}$ and

(a) $\|\mathcal{P}_{\Omega^\perp}(\mathcal{W})\| \leq \frac{1}{2}$.

(b) $\|\mathcal{P}_\Omega(\mathcal{W} - \mathcal{U}*_D\mathcal{V}^T)\| \leq \frac{1}{4}\sqrt{\frac{p}{r_3}}$.

Proof. Consider any feasible solution $\mathcal{X}$ to TC model (14). Let $\mathcal{G}$ be an $n\times n\times n_3$ tensor which satisfies $\|\mathcal{P}_{\Omega^\perp}\mathcal{G}\| = 1$ and $\langle\mathcal{P}_{\Omega^\perp}\mathcal{G}, \mathcal{P}_{\Omega^\perp}(\mathcal{T}-\mathcal{T}_0)\rangle = \|\mathcal{P}_{\Omega^\perp}(\mathcal{T}-\mathcal{T}_0)\|_*$. Such $\mathcal{G}$ always exists by duality between the tensor nuclear norm and tensor spectral norm. Note that $\mathcal{U}*_D\mathcal{V}^T + \mathcal{P}_{\Omega^\perp}\mathcal{G}$ is a subgradient of $\mathcal{Z}$ and $\mathcal{Z} = \mathcal{T}_0$, we have

$$\|\mathcal{T}\|_* - \|\mathcal{T}_0\|_* \geq \left\langle\mathcal{U}*_D\mathcal{V}^T + \mathcal{P}_{\Omega^\perp}\mathcal{G}, \mathcal{T}-\mathcal{T}_0\right\rangle. \tag{96}$$

We also have $\langle\mathcal{W}, \mathcal{T}-\mathcal{T}_0\rangle = \langle\mathcal{P}_\Omega\mathcal{W}, \mathcal{P}_\Omega(\mathcal{T}-\mathcal{T}_0)\rangle = 0$ since $\mathcal{P}_\Omega(\mathcal{W}) = \mathcal{W}$. It follows that

$$\begin{aligned}
&\|\mathcal{T}\|_* - \|\mathcal{T}_0\|_* \geq \left\langle\mathcal{U}*_D\mathcal{V}^T + \mathcal{P}_{\Omega^\perp}\mathcal{G} - \mathcal{W}, \mathcal{T}-\mathcal{T}_0\right\rangle \\
&= \|\mathcal{P}_{\Omega^\perp}(\mathcal{T}-\mathcal{T}_0)\|_* + \left\langle\mathcal{U}*_D\mathcal{V}^T - \mathcal{P}_T\mathcal{W}, \mathcal{T}-\mathcal{T}_0\right\rangle \\
&\quad - \langle\mathcal{P}_{T^\perp}\mathcal{W}, \mathcal{T}-\mathcal{T}_0\rangle \\
&\geq \|\mathcal{P}_{\Omega^\perp}(\mathcal{T}-\mathcal{T}_0)\|_* + \|\mathcal{U}*_D\mathcal{V}^T - \mathcal{P}_T\mathcal{W}\|_F\|\mathcal{P}_T(\mathcal{T}-\mathcal{T}_0)\|_F \\
&\quad - \|\mathcal{P}_{T^\perp}\mathcal{W}\|\|\mathcal{P}_{T^\perp}(\mathcal{T}-\mathcal{T}_0)\|_* \\
&\geq \frac{1}{2}\|\mathcal{P}_{\Omega^\perp}(\mathcal{T}-\mathcal{T}_0)\|_* - \frac{1}{4}\sqrt{\frac{p}{r_3}}\|\mathcal{P}_T(\mathcal{T}-\mathcal{T}_0)\|_F
\end{aligned} \tag{97}$$

where the last inequality uses Conditions (1) and (2) in the proposition. Now, by using Lemma 13 below, we have

$$\begin{aligned}
\|\mathcal{T}\|_* - \|\mathcal{T}_0\|_* &\geq \frac{1}{2}\|\mathcal{P}_{\Omega^\perp}(\mathcal{T}-\mathcal{T}_0)\|_* \\
&\quad - \frac{1}{4}\sqrt{\frac{p}{r_3}}\sqrt{\frac{2r_3}{p}}\|\mathcal{P}_T\mathcal{T}-\mathcal{T}_0)\|_* \\
&> \frac{1}{8}\|\mathcal{P}_{\Omega^\perp}(\mathcal{X}-\mathcal{T}_0)\|_*.
\end{aligned} \tag{98}$$

Note that the right hand side of the above inequality is strictly positive for all $\mathcal{T}$ with $\mathcal{P}_\Omega(\mathcal{T}-\mathcal{T}_0) = 0$ and $\mathcal{T} \neq \mathcal{T}_0$. Otherwise, we must have $\mathcal{P}_T(\mathcal{T}-\mathcal{T}_0) = \mathcal{T}-\mathcal{T}_0$ and $\mathcal{P}_T\mathcal{R}_\Omega\mathcal{P}_T(\mathcal{T}-\mathcal{M}) = 0$, contradicting the assumption $\|\mathcal{P}_T\mathcal{R}_\Omega\mathcal{P}_T - \mathcal{P}_T\| \leq \frac{1}{2}$. Therefore, $\mathcal{T}_0$ is the unique optimum. □

Lemma 13. *If $\|\mathcal{P}_T\mathcal{R}_\Omega\mathcal{P}_T - \mathcal{P}_T\| \leq \frac{1}{2}$, then we have*

$$\|\mathcal{P}_T\mathcal{Z}\|_F \leq \sqrt{\frac{2r_3}{p}}\|\mathcal{P}_{T^\perp}\mathcal{Z}\|_*, \forall\mathcal{Z} \in \{\mathcal{Z}' : \mathcal{P}_\Omega(\mathcal{Z}') = 0\}. \tag{99}$$

Proof. We deduce

$$\begin{aligned}&\|\sqrt{p}\mathcal{R}_\Omega\mathcal{P}_T\mathcal{Z}\|_F\\&=\sqrt{\langle(\mathcal{P}_T\mathcal{R}_\Omega\mathcal{P}_T-\mathcal{P}_T)\mathcal{Z},\mathcal{P}_T\mathcal{Z}\rangle+\langle\mathcal{P}_T\mathcal{Z},\mathcal{P}_T\mathcal{Z}\rangle}\\&=\sqrt{\|\mathcal{P}_T\mathcal{Z}\|_F^2-\|\mathcal{P}_T\mathcal{R}_\Omega\mathcal{P}_T-\mathcal{P}_T\|\|\mathcal{P}_T\mathcal{Z}\|_F^2}\\&\geq\frac{1}{\sqrt{2}}\|\mathcal{P}_T\mathcal{Z}\|_F\end{aligned}\tag{100}$$

where the last inequality uses $\|\mathcal{P}_T\mathcal{R}_\Omega\mathcal{P}_T-\mathcal{P}_T\|\leq\frac{1}{2}$. On the other hand, $\mathcal{P}_\Omega(\mathcal{Z})=0$ implies that $\mathcal{R}_\Omega(\mathcal{Z})=0$ and thus

$$\begin{aligned}\|\sqrt{p}\mathcal{R}_\Omega\mathcal{P}_T\mathcal{Z}\|_F&=\|\sqrt{p}\mathcal{R}_\Omega\mathcal{P}_{T^\perp}\mathcal{Z}\|_F\\&\leq\frac{1}{\sqrt{p}}\|\mathcal{P}_T\mathcal{Z}\|_F\leq\sqrt{\frac{r_3}{p}}\|\mathcal{P}_T\mathcal{Z}\|_*,\end{aligned}\tag{101}$$

where the last inequality uses

$$\|\mathcal{A}\|_F=\|\overline{A}\|_F\leq\|\overline{A}\|_*\leq\|\mathcal{A}\|_*.\tag{102}$$

The proof is completed by combining Eq. (100) and (101). □

New we give the completed proof of the Exact Recovery Theorem (i.e., Theorem 4 in the manuscript) about TC model.

Proof. First, as shown in Lemma 9, the Condition 1 of Proposition 1 holds with high probability. Now we construct a dual certificate $\mathcal{W}$ which satisfies Condition 2 in Proposition 1. We do this using the Golfing Scheme. For the choice of p in Theorem 3 in the manuscript, we have

$$p\geq\frac{c_0\mu R(\log(n))^2}{n}\geq\frac{1}{n},\tag{103}$$

for some sufficiently large $c_0>0$. Set $t_0:=20\log(n)$. Assume that the set Ω of observed entries is generated from $\Omega=\cup_{t=1}^{t_0}\Omega_t$, where each t and tensor index (i,j,k), $\mathbb{P}[(i,j,k)\in\Omega_t]=q:=1-(1-p)^{1/t_0}$ and is independent of all others. Clearly this Ω has the same distribution as the original model. Let $\mathcal{W}_0:=0$ and for $t=1,\cdots,t_0$, define

$$\mathcal{W}_t=\mathcal{W}_{t-1}+\mathcal{R}_{\Omega_t}\mathcal{P}_T(\mathcal{U}*_D\mathcal{V}^T-\mathcal{P}_\Omega\mathcal{W}_{t-1}),\tag{104}$$

where the operator $\mathcal{R}_{\Omega_t}$ is defined analogously to $\mathcal{R}_\Omega$ as $\mathcal{R}_{\Omega_t}(\mathcal{Z}):=\sum_{ijk}q^{-1}1_{(i,j,k)\in\Omega_t}z_{ijk}\mathfrak{e}_{ijk}$. Then the dual certificate is given by $\mathcal{W}:=\mathcal{W}_{t_0}$. We have $\mathcal{P}_\Omega(\mathcal{W})=\mathcal{W}$ by construction. To prove Theorem 2, we only need to show that $\mathcal{W}$ satisfies Conditions 2 in Proposition 1 w.h.p.

Validating Condition 2(b). Denote $\mathcal{D}_t:=\mathcal{U}*_D\mathcal{V}^T-\mathcal{P}_T\mathcal{W}_k$ for $t=0,\cdots,t_0$. By the definition of $\mathcal{W}_k$, we have $\mathcal{D}_0=\mathcal{U}*_D\mathcal{V}^T$ and

$$\mathcal{D}_t=(\mathcal{P}_T-\mathcal{P}_T\mathcal{R}_{\Omega_t}\mathcal{P}_T)\mathcal{D}_{t-1}.\tag{105}$$

Obviously $\mathcal{D}_t\in T$ for all $t>0$. Note that Ω_t is independent of $\mathcal{D}_{t-1}$ and by the choice of p in Theorem 3 in the manuscript, we have

$$q\geq\frac{p}{t_0}\geq\frac{c_0\mu R\log(n)}{n}.\tag{106}$$

Applying Lemma 9 with Ω replaced by Ω_t, we obtain that w.h.p.

$$\|\mathcal{D}_t\|_F\leq\|\mathcal{P}_T-\mathcal{P}_T\mathcal{R}_{\Omega_t}\mathcal{P}_T\|\|\mathcal{D}_{t-1}\|_F\leq\frac{1}{2}\|\mathcal{D}_{t-1}\|_F\tag{107}$$

for each t. Applying the above inequality recursively with $t=t_0,t_0-1,\cdots,1$ gives

$$\begin{aligned}\|\mathcal{P}_T\mathcal{Y}-\mathcal{U}*_D\mathcal{V}^T\|_F&=\|\mathcal{D}_{t_0}\|_F\leq(\frac{1}{2})^{t_0}\|\mathcal{U}*_D\mathcal{V}^T\|_F\\&\leq\frac{1}{4nn_3}\sqrt{R}\leq\frac{1}{4\sqrt{nn_3}}\leq\frac{1}{4}\sqrt{\frac{p}{r_3}},\end{aligned}\tag{108}$$

where the last inequality uses Eq. (103) and $r_3\leq n_3$.

Validating Condition 2(a). Note that $\mathcal{W}=\sum_{t=1}^{t_0}\mathcal{R}_{\Omega_t}\mathcal{P}_T\mathcal{D}_{t-1}$ by construction. We have

$$\begin{aligned}\|\mathcal{P}_{\Omega^\perp}\mathcal{W}\|_F&\leq\sum_{t=1}^{t_0}\|\mathcal{P}_{T^\perp}(\mathcal{R}_{\Omega_t}\mathcal{P}_T-\mathcal{P}_T)\mathcal{D}_{t-1}\|\\&\leq\sum_{t=1}^{t_0}\|(\mathcal{R}_{\Omega_t}-\mathcal{I})\mathcal{P}_T\mathcal{D}_{t-1}\|.\end{aligned}\tag{109}$$

Applying Lemma 10 with Ω replaced by Ω_t to the above inequality, we get that w.h.p.

$$\begin{aligned}\|\mathcal{P}_{\Omega^\perp}\mathcal{Y}\|_F&\leq c\sum_{t=1}^{t_0}\left(\frac{\log(n)}{q}\|\mathcal{D}_{t-1}\|_\infty+\sqrt{\frac{\log(n)}{q}}\|\mathcal{D}_{t-1}\|_{\infty,2}\right)\\&\leq\frac{c}{\sqrt{c_0}}\sum_{t=1}^{t_0}\left(\frac{n}{\mu R}\|\mathcal{D}_{t-1}\|_\infty+\sqrt{\frac{n}{\mu R}}\|\mathcal{D}_{t-1}\|_{\infty,2}\right)\end{aligned}\tag{110}$$

where the last inequality uses Eq. (106). Now we bound $\|\mathcal{D}_{t-1}\|_\infty$and $\|\mathcal{D}_{t-1}\|_{\infty,2}$. Using Eq. (105) and repeatedly applying Lemma 4 with Ω replaced as Ω_t, we obtain that w.h.p.

$$\begin{aligned}&\|\mathcal{D}_{t-1}\|_\infty\\&=\|(\mathcal{P}_T-\mathcal{P}_T\mathcal{R}_{\Omega_{t-1}}\mathcal{P}_T)\cdots(\mathcal{P}_T-\mathcal{P}_T\mathcal{R}_{\Omega_1}\mathcal{P}_T)\mathcal{D}_0\|_\infty\\&\leq(\frac{1}{2})^{t-1}\|\mathcal{U}*_D\mathcal{V}^T\|_\infty.\end{aligned}\tag{111}$$

By Lemma 11 with Ω replaced by Ω_t, we obtain that w.h.p.

$$\begin{aligned}\|\mathcal{D}_{t-1}\|_{\infty,2}&=\|(\mathcal{P}_T-\mathcal{P}_T\mathcal{R}_{\Omega_{t-1}}\mathcal{P}_T)\mathcal{D}_{t-2}\|_{\infty,2}\\&\leq\frac{1}{2}\sqrt{\frac{n}{\mu R}}\|\mathcal{D}_{t-2}\|_\infty+\frac{1}{2}\|\mathcal{D}_{t-2}\|_{\infty,2}.\end{aligned}\tag{112}$$

Using Eq. (105) and combining the last two display equations gives w.h.p.

$$\begin{aligned}\|\mathcal{D}_{t-1}\|_{\infty,2}\leq\ &t(\frac{1}{2})^{t-1}\sqrt{\frac{nn_3}{\mu R}}\|\mathcal{U}*_D\mathcal{V}^T\|_\infty\\&+(\frac{1}{2})^{t-1}\|\mathcal{U}*_D\mathcal{V}^T\|_{\infty,2}.\end{aligned}\tag{113}$$

Substituting back to Eq. (110), we get w.h.p.

$$\begin{aligned}\|\mathcal{P}_{\Omega^\perp}\mathcal{Y}\|_F &\le \frac{c}{\sqrt{c_0}}\frac{n}{\mu R}\|\mathcal{U}*_D\mathcal{V}^T\|_\infty\sum_{t=1}^{t_0}(t+1)(\frac{1}{2})^{t+1}\\ &+\frac{c}{\sqrt{c_0}}\sqrt{\frac{n}{\mu R}}\|\mathcal{U}*_D\mathcal{V}^T\|_{\infty,2}\sum_{t=1}^{t_0}(\frac{1}{2})^{t+1}\\ &\le\frac{6c}{\sqrt{c_0}}\frac{n}{\mu R}\|\mathcal{U}*_D\mathcal{V}^T\|_\infty\\ &+\frac{2c}{\sqrt{c_0}}\sqrt{\frac{n}{\mu R}}\|\mathcal{U}*_D\mathcal{V}^T\|_{\infty,2}.\end{aligned}\tag{114}$$

Now we proceed to bound $\|\mathcal{U}*_D\mathcal{V}^T\|_\infty$ and $\|\mathcal{U}*_D\mathcal{V}^T\|_{\infty,2}$. First, by the definition of t-product, we have

$$\begin{aligned}&\|\mathcal{U}*_D\mathcal{V}^T\|_\infty=\max_{i,j}\left\|\sum_{t=1}^{r}\mathcal{U}(i,t,:)*_D\mathcal{V}(j,t,:)\right\|_\infty\\ &\le\max_{i,j}\sum_{t=1}^{r}\|\mathcal{U}(i,t,:)\|_F\|\mathcal{V}(j,t,:)\|_F\\ &\le\max_{i,j}\sum_{t=1}^{r}\frac{1}{2}\left(\|\mathcal{U}(i,t,:)\|_F^2+\|\mathcal{V}(j,t,:)\|_F^2\right)\\ &=\max_{i,j}\frac{1}{2}\left(\|\mathcal{U}^T*_D\mathring{\mathfrak{e}}_i\|_F^2+\|\mathcal{V}^T*_D\mathring{\mathfrak{e}}_j\|_F^2\right)\le\frac{\mu R}{n},\end{aligned}\tag{115}$$

Also, we have

$$\begin{aligned}&\|\mathcal{U}*_D\mathcal{V}^T\|_{\infty,2}\\ &\le\max\left\{\max_i\|\mathring{\mathfrak{e}}_i^T*_D\mathcal{U}*_D\mathcal{V}\|_F,\max_i\|\mathcal{U}*_D\mathcal{V}*_D\mathring{\mathfrak{e}}_j\|_F\right\}\\ &\le\sqrt{\frac{\mu R}{n}}.\end{aligned}\tag{116}$$

It follows that w.h.p.

$$\|\mathcal{P}_{T^\perp}\mathcal{Y}\|\le\frac{6c}{\sqrt{c_0}}+\frac{2c}{\sqrt{c_0}}\le\frac{1}{2},\tag{117}$$

provided that c_0 is sufficiently large. This completes the proof of Theorem 3 in the manuscript. □

4.2 Proof of Some Lemmas

Proof of Lemma 10

Proof. Denote the tensor $\mathcal{H}_{ijk}=(1-\rho^{-1}\delta_{ijk})z_{ijk}\mathfrak{e}_{ijk}$. Then we have

$$(\mathcal{I}-\mathcal{R}_\Omega)\mathcal{Z}=\sum_{ijk}\mathcal{H}_{ijk}.\tag{118}$$

Note that δ_{ijk}'s are independent random scalars. Thus, $\mathcal{H}_{ijk}$'s are independent random tensors and $\overline{\mathrm{H}}_{ijk}$'s are independent random matrices. Observe that $\mathbb{E}[\overline{\mathrm{H}}_{ijk}]=\mathbf{0}$ and $\|\overline{\mathrm{H}}_{ijk}\|\le\rho^{-1}\|\mathcal{Z}\|_\infty$, we have

$$\begin{aligned}\left\|\sum_{ijk}\mathbb{E}[\overline{\mathrm{H}}_{ijk}^*\overline{\mathrm{H}}_{ijk}]\right\|&=\left\|\sum_{ijk}\mathbb{E}[\mathcal{H}_{ijk}^*\mathcal{H}_{ijk}]\right\|\\ &=\left\|\sum_{ijk}\mathbb{E}[(1-\rho^{-1}\delta_{ijk})^2]z_{ijk}^2(\mathring{\mathfrak{e}}_j*_L\mathring{\mathfrak{e}}_j^*)\right\|\\ &=\left\|\frac{1-\rho}{\rho}\sum_{ijk}z_{ijk}^2(\mathring{\mathfrak{e}}_j*_L\mathring{\mathfrak{e}}_j^*)\right\|\\ &\le\rho^{-1}\max_j\left|\sum_{i,k}z_{ijk}^2\right|\le\rho^{-1}\|\mathcal{Z}\|_{\infty,2}^2.\end{aligned}\tag{119}$$

A similar calculation yields $\left\|\sum_{ijk}\mathbb{E}[\overline{\mathrm{H}}_{ijk}^*\overline{\mathrm{H}}_{ijk}]\right\|\le\rho^{-1}\|\mathcal{Z}\|_{\infty,2}^2$. Then the proof is completed by applying the matrix Bernstein inequality in Theorem 5. □

Proof of Lemma 11

Proof. For fixed $\mathcal{Z}\in T$ and fixed $b\in[n]$, the b-th column of the tensor $\mathcal{P}_T\mathcal{R}_\Omega(\mathcal{Z})-\mathcal{Z}$ can be written as

$$\begin{aligned}&(\mathcal{P}_T\mathcal{R}_\Omega(\mathcal{Z})-\mathcal{Z})*_D\mathring{\mathfrak{e}}_b\\ &=\sum_{ijk}(\rho^{-1}-1)\delta_{ijk}z_{ijk}\mathcal{P}_T(\mathfrak{e}_{ijk})*_D\mathring{\mathfrak{e}}_b:=\sum_{ijk}\mathcal{H}_{ijk},\end{aligned}\tag{120}$$

where $\mathcal{H}_{ijk}$'s are independent column tensor in $\mathbb{R}^{n\times1\times n_3}$ and $\mathbb{P}[\mathcal{H}_{ijk}]=\mathbf{0}$. Let $\boldsymbol{h}_{ijk}\in\mathbb{R}^{nn_3}$ be the column vector obtained by vectorizing $\mathcal{H}_{ijk}$. Then we have

$$\begin{aligned}\|\boldsymbol{h}_{ijk}\|&\le\rho^{-1}|z_{ijk}|\|\mathcal{P}_T(\mathfrak{e}_{ijk})*_L\mathring{\mathfrak{e}}_b\|_F\\ &\le\rho^{-1}\|\mathcal{Z}\|_\infty\sqrt{\frac{2\mu R}{n}}\le\frac{1}{c_0\log(n)}\sqrt{\frac{2n}{\mu R}}\|\mathcal{Z}\|_\infty.\end{aligned}\tag{121}$$

We also have

$$\begin{aligned}\left|\sum_{ijk}\mathbb{E}[\boldsymbol{h}_{ijk}^*\boldsymbol{h}_{ijk}]\right|&=\left|\sum_{ijk}\mathbb{E}[\|\mathcal{H}_{ijk}\|_F^2]\right|\\ &\le\frac{1-\rho}{\rho}\sum_{ijk}z_{ijk}^2\|\mathcal{P}_T(\mathfrak{e}_{ijk})*_D\mathring{\mathfrak{e}}_b\|_F^2.\end{aligned}\tag{122}$$

Note that

$$\begin{aligned}&\|\mathcal{P}_T(\mathfrak{e}_{ijk})*_D\mathring{\mathfrak{e}}_b\|_F^2\\ =&\|\mathcal{U}*_D\mathcal{U}^T*_D\mathring{\mathfrak{e}}_i*_D\dot{\mathfrak{e}}_k*_D\mathring{\mathfrak{e}}_j^**_D\mathring{\mathfrak{e}}_b\\ &-(\mathcal{I}-\mathcal{U}*_D\mathcal{U}^T)*_D\mathring{\mathfrak{e}}_i*_D\dot{\mathfrak{e}}_k*_D\mathring{\mathfrak{e}}_j^**_D\mathcal{V}*_D\mathcal{V}^T*_D\mathring{\mathfrak{e}}_b\|_F\\ \le&\|\mathcal{U}*_D\mathcal{U}^T*_D\mathring{\mathfrak{e}}_i*_D\dot{\mathfrak{e}}_k\|_F\|\mathring{\mathfrak{e}}_j^**_L\mathring{\mathfrak{e}}_b\|_F\\ &+\|(\mathcal{I}-\mathcal{U}*_D\mathcal{U}^T)*_D\mathring{\mathfrak{e}}_i*_D\dot{\mathfrak{e}}_k\|\|\mathring{\mathfrak{e}}_j^**_D\mathcal{V}*_D\mathcal{V}^T*_D\mathring{\mathfrak{e}}_b\|_F\\ \le&\sqrt{\frac{\mu R}{n}}\|\mathring{\mathfrak{e}}_j^**_D\mathring{\mathfrak{e}}_b\|_F+\|\mathring{\mathfrak{e}}_j^**_D\mathcal{V}*_D\mathcal{V}^T*_D\mathring{\mathfrak{e}}_b\|_F.\end{aligned}\tag{123}$$

It follows that

$$\left|\sum_{ijk}\mathbb{E}[\boldsymbol{h}_{ijk}^*\boldsymbol{h}_{ijk}]\right| = \frac{2}{\rho}\sum_{ijk} z_{ijk}^2\frac{\mu R}{n}\|\mathring{\mathfrak{e}}_j^* *_L \mathring{\mathfrak{e}}_b\|_F^2$$
$$+\frac{2}{\rho}\sum_{ijk} z_{ijk}^2\|\mathring{\mathfrak{e}}_j^* *_D \mathcal{V} *_D \mathcal{V}^T *_D \mathring{\mathfrak{e}}_b\|_F^2$$
$$=\frac{2\mu R}{\rho n}\sum_{ijk} z_{ijk}^2 + \frac{2}{p}\sum_j \|\mathring{\mathfrak{e}}_j^* *_D \mathcal{V} *_D \mathcal{V}^T *_D \mathring{\mathfrak{e}}_b\|_F^2 \sum_{ik} z_{ijk}^2$$
$$\leq \frac{2\mu R}{\rho n}\|\mathcal{Z}\|_{\infty,2}^2 + \frac{2}{\rho}\|\mathcal{V} *_D \mathcal{V}^T *_D \mathring{\mathfrak{e}}_b\|_F^2\|\mathcal{Z}\|_{\infty,2}^2$$
$$\leq \frac{4\mu R}{\rho n}\|\mathcal{Z}\|_{\infty,2}^2 \leq \frac{4}{c_0\log(n)}\|\boldsymbol{\mathcal{Z}}\|_{\infty,2}^2. \tag{124}$$

We can bound $\|\sum_{ijk}\mathbb{E}[\boldsymbol{h}_{ijk}\boldsymbol{h}_{ijk}^*]\|$ by the same quantity in a similar manner. Treating $\boldsymbol{h}_{ijk}$'s as $nn_3\times 1$ matrices and applying the matrix Bernstein inequality in Theorem 5 gives that w.h.p.

$$\begin{aligned}&\|(\mathcal{P}_T\mathcal{R}_\Omega(\mathcal{Z})-\mathcal{Z}) *_D \mathring{\mathfrak{e}}_b\|_F\\ &=\left\|\sum_{ijk}\mathcal{H}_{ijk}\right\| = \left\|\sum_{ijk}\boldsymbol{h}_{ijk}\right\|\\ &\leq \frac{C}{c_0}\sqrt{\frac{2n}{\mu R}}\|\mathcal{Z}\|_\infty + 4\sqrt{\frac{C}{c_0}}\|\mathcal{Z}\|_{\infty,2}\\ &\leq \frac{1}{2}\sqrt{\frac{2n}{\mu R}}\|\mathcal{Z}\|_\infty + \frac{1}{2}\|\mathcal{Z}\|_{\infty,2},\end{aligned} \tag{125}$$

provided that c_0 in the lemma statement is large enough. In a similar way, we prove that $\|\mathring{\mathfrak{e}}_b^* *_L (\mathcal{P}_T\mathcal{R}_\Omega(\mathcal{Z})-\mathcal{Z})\|_F$ is bounded by the same quantity w.h.p. The lemma follows from a union bound over all $(a,b)\in[n]\times[n]$. □

5 Algorithm Details

5.1 Details of Solving ATNN-TC model (14) and ATNN-RPCA model (15)

We first write the augmented Lagrangian function of the ATNN-TC model (14) as:

$$\min_{\mathcal{B},\mathrm{D},\boldsymbol{\mathcal{E}},\Lambda,\mathcal{P}_\Omega(\mathcal{E})=0}\|\mathcal{B}\|_* + \frac{\mu}{2}\|\mathcal{Y}-\mathcal{B}\times_3\mathrm{D}-\mathcal{E}+\Lambda/\mu\|_F^2, \tag{126}$$

where μ is the penalty parameter and Λ is the lagrange multiplier. For Eq. (126), we need to solve the following three subproblems in Eq. (127).

$$\begin{cases}\mathcal{B} := \min\limits_{\mathcal{B}}\|\mathcal{B}\|_* + \dfrac{\mu}{2}\|\mathcal{Y}-\mathcal{B}\times_3\mathrm{D}-\mathcal{E}+\Lambda/\mu\|_F^2,\\ \mathrm{D} := \min\limits_{\mathrm{D}^T\mathrm{D}=\mathrm{I}}\|\mathcal{Y}-\mathcal{B}\times_3\mathrm{D}-\mathcal{E}+\Lambda/\mu\|_F^2,\\ \mathcal{E} := \min\limits_{\mathcal{P}_\Omega(\mathcal{E})=0}\|\mathcal{Y}-\mathcal{B}\times_3\mathrm{D}-\mathcal{E}+\Lambda/\mu\|_F^2,\\ \Lambda = \Lambda+\mu(\mathcal{Y}-\mathcal{B}\times_3\mathrm{D}-\mathcal{E}).\end{cases} \tag{127}$$

Update $\mathcal{B}$. According to TNN definition, solving $\mathcal{B}$ can be rewritten as the following r_3 equations:

$$\arg\min_{\mathcal{B}(:,:,i)}\|\mathcal{B}(:,:,i)\|_* + \frac{\mu}{2}\|\mathcal{B}(:,:,i)-\mathcal{Z}(:,:,i)\|_F^2, \tag{128}$$

for each $i=1,\cdots,r_3$, where $\mathcal{Z} := (\mathcal{Y}-\mathcal{E}+\Lambda/\mu\times_3\mathrm{D}^T$.

Then the each $\mathcal{B}(:,:,i), i=1,\cdots,r_3$ can be updated by the soft-thresholding operator SVD$\tau(\cdot)$ [7]:

$$\mathcal{B}(:,:,i) = \mathrm{U}\mathcal{S}_{1/\mu}(\Sigma)\mathrm{V}^T, \text{where } \mathcal{Z}(:,:,i) \overset{\text{svd}}{=} \mathrm{U}\Sigma\mathrm{V}^T, \tag{129}$$

where $\mathcal{S}_\tau$ is the soft threshold operator $\mathcal{S}$ defined by [8].

Update D. The subproblem about updating D in Eq. (127) can be rewritten as:

$$\arg\max_{\mathrm{D}^T\mathrm{D}=\mathrm{I}}\left\langle \mathrm{unfold}_3(\mathcal{Y}-\mathcal{E}+\Lambda/\mu)^T\mathrm{unfold}_3(\mathrm{bfold}(\mathcal{B}),\mathrm{D}\right\rangle. \tag{130}$$

According to Theorem 1 in [9], we can get the solution of D in Eq. (127) as follows:

$$\begin{cases}[\mathrm{U},\Sigma,\mathrm{V}] = \mathrm{svd}(\mathrm{O}),\\ \mathrm{D} = \mathrm{U}\mathrm{V}^T,\end{cases} \tag{131}$$

where $\mathrm{O} := \mathrm{unfold}_3(\mathcal{Y}-\mathcal{E}+\Lambda/\mu)^T\mathrm{unfold}_3(\mathrm{bfold}(\mathcal{B})$.

Update $\mathcal{E}$. The solution of $\mathcal{E}$ in Eq. (127) is:

$$\mathcal{E} = \mathcal{P}_{\Omega^\perp}(\mathcal{Y}-\mathcal{B}\times_3\mathrm{D}+\Lambda/\mu) \tag{132}$$

The augmented Lagrangian function of the ATNN-RPCA model (15) as:

$$\min_{\mathcal{B},\mathrm{D},\mathcal{E},\Lambda}\|\mathcal{B}\|_* + \lambda\|\mathcal{E}\|_1 + \frac{\mu}{2}\|\mathcal{Y}-\mathcal{B}\times_3\mathrm{D}-\mathcal{E}+\Lambda/\mu\|_F^2, \tag{133}$$

The solution to Eq. (133) is also divided into four steps, which involve the solution process of sub-problems $\mathcal{B}$, D, and $\mathcal{E}$. This process is completely consistent with ATNN-TC model, with the only difference being in the optimization problem of $\mathcal{E}$.

The $\mathcal{E}$ sub-problems of Eq. (133) is:

$$\min_{\mathcal{E}}\lambda\|\mathcal{E}\|_1 + \frac{\mu}{2}\|\mathcal{Y}-\mathcal{B}\times_3\mathrm{D}-\mathcal{E}+\Lambda/\mu\|_F^2. \tag{134}$$

The solution of Eq. (134) is:

$$\mathcal{E} := \mathcal{S}_{\lambda/\mu}(\|\mathcal{Y}-\mathcal{B}\times_3\mathrm{D}+\Lambda/\mu\|_F^2). \tag{135}$$

5.2 Convergence Analysis of the Algorithm 1 and Algorithm 2

Since program (126) and (126) are a non-convex programs, we cannot directly apply the theory of convex optimization [10] to provide a proof of their global convergence. Here, we can establish the convergence of these two algorithms by relying on the following two lemmas.

Lemma 14. *The sequence of dual variable Λ in the program (126) and (126) is bounded.*

Proof. According to the optimality principle, we have

$$\begin{aligned}\mathbf{0} &\in \partial(\|\mathcal{B}^{k+1}\|_*) - \Lambda^k\times_3 {\mathrm{D}^{k+1}}^T\\ &\quad - \mu_k\left(\mathcal{Y}-\mathcal{B}^{k+1}\times_3\mathrm{D}^{k+1}-\mathcal{E}^{k+1}\right)\times_3 {\mathrm{D}^{k+1}}^T,\\ \mathbf{0} &\in \partial(\lambda\|\mathcal{E}^{k+1}\|_1) - \Lambda^k - \mu_k\left(\mathcal{Y}-\mathcal{B}^{k+1}\times_3\mathrm{D}^{k+1}-\mathcal{E}^{k+1}\right).\end{aligned} \tag{136}$$

Combining this with the update criterion of the Λ^k in Algorithm 1 and 2, we have

$$\begin{aligned}\Lambda^{k+1} \times_3 \mathrm{D}^{k+1^T} \in \partial(\|\mathcal{B}^{k+1}\|_*), \\ \Lambda^{k+1} \in \partial(\|\mathcal{E}^{k+1}\|_1).\end{aligned} \tag{137}$$

Note the fact that the dual norm of $\|\cdot\|_*$ and $\|\cdot\|_1$ are $\|\cdot\|_2$ and $\|\cdot\|_\infty$, respectively, and $\|\cdot\|_2 = \lambda^{-1}\|\cdot\|_\infty$ by the definition in [7]. Thus, using Theorem 4 in [11], we get that $\mathbf{\Lambda}^{k+1}$ are bounded. □

Lemma 15. *The accumulation point* $(\mathcal{B}^k, \mathrm{D}^k, \mathcal{E}^k)$ *generated by Eq. (127) is a feasible solution of ATNN-TC model (14) and ATNN-RPCA model (15).*

Proof. Based on the general ADMM principle, we have

$$\|\Lambda^{k+1} - \Lambda^k\|_F = \mu^k\|\mathcal{Y} - \mathcal{B}^{k+1} \times_3 \mathrm{D}^{k+1} - \mathcal{E}^{k+1}\|_F \tag{138}$$

Since $\{\mu^k\}$ is an increasing sequence and $\lim_{k\to+\infty} \mu^k = +\infty$, and according to Lemma 14, we have

$$\lim_{k\to+\infty} \|\mathcal{Y} - \mathcal{B}^{k+1} \times_3 \mathrm{D}^{k+1} - \mathcal{E}^{k+1}\|_F = 0 \tag{139}$$

This completes the proof. □

Next, we give the following convergence theorem about the ADMM algorithm to solve the Eq. (126) and (133)

Theorem 6. *The sequence* $(\mathcal{T}^k = \mathcal{B}^k \times_3 \mathrm{D}^k, \mathcal{E}^k)$ *generated by the ADMM converge to the optimal solution of ATNN-TC model (14) and ATNN-RPCA model (15).*

Proof. Suppose $(\mathcal{T}^*, \mathcal{E}^*)$ are the optimal solution of ADMM. Since the decomposition form $\mathcal{T}^* = \mathcal{B}^k_{\mathrm{D}^k} \times_3 \mathrm{D}_k$ will not lose information of $\mathcal{T}^*$, where D_k is the solution of ATNN-TC model (14) and ATNN-RPCA model (15) in the k-th iteration.

Based on Eq. (137) and the definition of subgradient, we have

$$\begin{aligned}&\|\mathcal{B}^k\|_* + \lambda\|\mathcal{E}^k\|_1 \le \|\mathcal{B}^k_{\mathrm{D}^k}\|_* - \langle\Lambda^k \times_3 \mathrm{D}^{k^T}, \mathcal{B}^*_{\mathrm{D}^k} - \mathcal{B}^k\rangle \\ &+ \lambda\|\mathcal{E}^*\|_1 - \langle\Lambda^k, \mathcal{E}^* - \mathcal{E}^k\rangle \\ &= \|\mathcal{M}^k_{\mathrm{D}^k}\|_* + \lambda\|\mathcal{E}^*\|_1 \\ &- \langle\Lambda^k, \mathcal{B}^*_{\mathrm{D}^k} \times_3 \mathrm{D}^k - \mathcal{B}^k \times_3 \mathrm{D}^k\rangle - \langle\Lambda^k, \mathcal{E}^* - \mathcal{E}^k\rangle \\ &= \|\mathcal{B}^k_{\mathrm{D}^k}\|_* + \lambda\|\mathcal{E}^*\|_1 - \langle\Lambda^k, \mathcal{Y} - \mathcal{B}^k \times_3 \mathrm{D}^k - \mathcal{E}^k\rangle \\ &= \|\mathcal{B}^k_{\mathrm{D}^k}\|_* + \lambda\|\mathcal{E}^*\|_1 + \langle\Lambda^k, \mathcal{M}^k \times_3 \mathrm{D}^k + \mathcal{E}^k - \mathcal{Y}\rangle\end{aligned}$$

Combining the above equation with Lemma 15, we further have

$$\lim_{k\to+\infty} \|\mathcal{B}^k\|_* + \lambda\|\mathcal{E}^k\|_1 = \lim_{k\to+\infty} \|\mathcal{M}^*_{\mathrm{D}^k}\|_* + \lambda\|\mathcal{E}^*\|_1. \tag{140}$$

According to the optimality criterion, we have

$$\begin{aligned}\|\mathcal{B}^*_{\mathrm{D}^k}\|_* + \lambda\|\mathcal{E}^*\|_1 &\le \|\mathcal{B}^*_{\mathrm{D}^k}\|_* + \lambda\|\mathcal{E}^k\|_1 \\ &\le \|\mathcal{B}^k\|_* + \lambda\|\mathcal{E}^k\|_1.\end{aligned} \tag{141}$$

Taking the limit of k on both sides of Eq. (141), we can get

$$\begin{aligned}\lim_{k\to+\infty} \|\mathcal{B}^k\|_* &= \lim_{k\to+\infty} \|\mathcal{B}^*_{\mathrm{D}^k}\|_*, \\ \lim_{k\to+\infty} \|\mathcal{E}^k\|_1 &= \lim_{k\to+\infty} \|\mathcal{E}^*\|_1.\end{aligned} \tag{142}$$

Based on the above equation, we can deduce that $\lim_{k\to+\infty} \mathcal{E}^k = \lim_{k\to+\infty} \mathcal{E}^*$. Moreover, as per Lemma 15, we know that $\mathcal{B}^k, \mathrm{D}^k, \mathcal{E}^k$ are all feasible solutions of the ATNN-TC model (14) and ATNN-RPCA model (15). Consequently, we can derive that

$$\begin{aligned}\lim_{k\to+\infty} \mathcal{B}^k \times_3 \mathrm{D}^k &= \lim_{k\to+\infty} \mathcal{Y} - \mathcal{E}^k \\ &= \mathcal{Y} - \lim_{k\to+\infty} \mathcal{E}^k = \mathcal{Y} - \mathcal{E}^* = \mathcal{B}^*.\end{aligned} \tag{143}$$

This completes the proof. □

6 More Experiments about ATNN-based Models

In the manuscript, due to page limitations, we only include the recovery results for the TRPCA task with a sparse noise variance of 0.6 and the recovery results for the TC task with an observation rate of 0.05. Here, we present more experimental results. The experimental numerical results of all compared methods for the TRPCA task under various sparse noises are provided in Tables 1 and 2. The experimental numerical results of all compared methods for the TC task under various sparse noises are provided in Tables 3, 4, 5 and 6.

From the results in Tables 1 and 2, it can be observed that despite our method only utilizing low-rank prior on spectral bands, it outperforms methods such as CTV and TCTV, which simultaneously incorporate spatial smoothness and spectral low-rankness priors. This demonstrates the effectiveness of the proposed ATNN norm. Additionally, our method exhibits significant advantages in terms of speed. In certain cases with sparse noise, the running time of our method is even lower than that of RPCA methods based on matrix nuclear norm. It should be noted that the running time of our method varies in different scenarios of sparse noise. This is because the chosen rank, i.e., r_3, is different for different sparse noise scenarios. In more complex scenarios where the proportion of valid information in the data is lower and the proportion of erroneous information is higher, assigning a high value to r_3 would not only fail to learn an effective COM but also increase the algorithm's running time. Therefore, for sparse noise with large variance, a smaller rank, i.e., r_3, should be chosen.

Compared to the TRPCA task, the tensor completion (TC) task has received more extensive attention since it has the higher practical value. As a result, many strong comparative methods have emerged, such as S2NTNN based on nonlinear transformations using neural networks, KBR model based on Tucker and CP joint decomposition, and the recently proposed TCTV that integrates both low-rankness and local smoothness properties. Nevertheless, from Tables 3 to (6), we can observe that the performance of the proposed ATNN is comparable to these three state-of-the-art tensor methods. Considering the recoverability theory and running time of our proposed model, the ATNN model demonstrates strong competitiveness.

References

[1] Canyi Lu, Jiashi Feng, Yudong Chen, Wei Liu, Zhouchen Lin, and Shuicheng Yan. Tensor robust principal component analysis with a new tensor nuclear norm. *IEEE*

Table 1: Quantitative comparison of all RPCA-based competing methods on **WDC** dataset under salt-and-pepper noise with various variance. The best and second results are highlighted in bold italics and underline.

Variance	Metric	Observed	RPCA	SNN	KBR	TNN	CTNN	CTV	TCTV	Ours
0.1	MPSNR	14.26	45.35	47.68	35.65	47.27	28.79	50.11	49.46	**50.19**
	MSSIM	0.2830	0.9980	0.9990	0.9720	0.9940	0.8820	0.9980	0.9950	**0.9990**
	MFSIM	0.7200	0.9980	0.9990	0.9820	0.9960	0.9260	0.9990	0.9970	**0.9990**
	ERGAS	836.82	28.99	22.41	71.06	34.20	149.31	16.35	26.38	**15.16**
	MSAM	43.10	1.49	1.04	4.52	2.50	7.82	1.23	2.08	**1.02**
	Times	/	**27.18**	291.36	219.16	808.74	339.07	122.16	607.60	43.39
0.2	MPSNR	11.24	43.99	46.27	34.98	44.99	27.17	48.70	47.78	**49.05**
	MSSIM	0.1370	0.9970	0.9980	0.9680	0.9920	0.8190	0.9980	0.9950	**0.9980**
	MFSIM	0.5850	0.9970	0.9980	0.9800	0.9950	0.8970	0.9990	0.9970	**0.9990**
	ERGAS	1184.15	33.34	25.98	76.09	38.53	179.40	18.51	28.87	**17.25**
	MSAM	48.61	1.65	1.25	4.84	2.87	11.65	1.34	2.28	**1.10**
	Times	/	**32.05**	460.46	352.29	1115.37	498.99	210.54	838.54	37.67
0.3	MPSNR	9.48	42.33	44.55	34.05	41.94	25.06	47.21	45.70	**47.62**
	MSSIM	0.0830	0.9960	0.9970	0.9620	0.9880	0.7030	0.9980	0.9930	**0.9980**
	MFSIM	0.5020	0.9960	0.9980	0.9760	0.9920	0.8560	0.9980	0.9960	**0.9992**
	ERGAS	1450.57	39.61	30.98	84.55	45.26	228.60	21.92	32.33	**20.12**
	MSAM	50.91	1.88	1.70	5.49	3.53	16.90	**1.50**	2.56	**1.23**
	Times	/	37.83	474.72	340.09	1096.57	498.69	210.93	839.23	**32.36**
0.4	MPSNR	8.23	39.98	42.20	32.63	37.17	22.39	45.35	43.09	**45.73**
	MSSIM	0.0540	0.9940	0.9940	0.9480	0.9710	0.5230	0.9940	0.9900	**0.9970**
	MFSIM	0.4470	0.9940	0.9960	0.9670	0.9840	0.7950	0.9950	0.9940	**0.9980**
	ERGAS	1675.42	49.52	39.29	98.13	62.72	312.13	29.41	37.60	**24.61**
	MSAM	52.03	2.21	2.61	6.43	5.59	24.02	1.78	3.04	**1.48**
	Times	/	**26.52**	482.14	253.40	1108.16	456.48	211.40	873.60	42.30
0.5	MPSNR	7.26	36.58	38.01	28.55	28.92	19.70	41.07	39.49	**42.82**
	MSSIM	0.0370	0.9840	0.9780	0.8710	0.8210	0.3430	0.9890	0.9820	**0.9950**
	MFSIM	0.4090	0.9880	0.9870	0.9230	0.9220	0.7240	0.9910	0.9900	**0.9970**
	ERGAS	1873.12	68.70	58.84	153.34	147.67	427.94	41.11	48.65	**32.55**
	MSAM	52.52	2.83	5.19	9.27	13.59	31.54	2.53	4.14	**2.20**
	Times	/	42.45	481.57	289.15	1017.68	454.77	212.01	876.65	**32.83**
0.6	MPSNR	6.47	32.09	26.02	22.65	19.62	17.21	33.85	31.95	**39.82**
	MSSIM	0.0260	0.9520	0.7170	0.6430	0.3720	0.2030	0.9450	0.9080	**0.9910**
	MFSIM	0.3820	0.9700	0.8770	0.8390	0.7290	0.6530	0.9670	0.9500	**0.9940**
	ERGAS	2052.29	112.91	211.96	303.49	432.62	570.94	**87.23**	103.47	**45.81**
	MSAM	52.68	4.43	21.52	12.82	29.85	37.78	7.70	9.43	**2.47**
	Times	/	29.00	736.19	167.21	419.22	485.74	170.22	815.04	**21.34**

Table 2: Quantitative comparison of all RPCA-based competing methods on **PaviaU** dataset under salt-and-pepper noise with various variance. The best and second results are highlighted in bold italics and underline.

Variance	Metric	Observed	RPCA	SNN	KBR	TNN	CTNN	CTV	TCTV	Ours
0.1	MPSNR	14.52	46.51	43.33	33.69	46.83	27.42	47.80	<u>47.89</u>	**50.20**
	MSSIM	0.2750	0.9970	0.9970	0.9670	0.9900	0.8640	<u>0.9980</u>	0.9930	**0.9987**
	MFSIM	0.7230	0.9970	0.9970	0.9800	0.9930	0.9050	<u>0.9980</u>	0.9950	**0.9987**
	ERGAS	699.00	34.96	35.94	81.40	44.20	159.33	<u>16.07</u>	34.02	**15.38**
	MSAM	39.55	0.107	1.12	4.25	3.81	6.05	<u>1.06</u>	3.26	**1.04**
	Times	/	**5.01**	124.39	59.76	71.26	123.12	39.55	140.24	<u>38.57</u>
0.2	MPSNR	11.51	43.21	41.14	32.77	44.48	25.73	<u>46.54</u>	46.46	**49.03**
	MSSIM	0.1280	0.9940	0.9950	0.9610	0.9870	0.7720	<u>0.9970</u>	0.9910	**0.9985**
	MFSIM	0.5780	0.9940	0.9950	0.9760	0.9910	0.8700	<u>0.9988</u>	0.9940	**0.9991**
	ERGAS	989.25	62.63	45.73	90.50	49.20	191.17	<u>18.71</u>	36.61	**17.15**
	MSAM	45.48	1.26	1.38	4.97	4.17	8.96	<u>1.12</u>	3.40	**1.10**
	Times	/	**5.76**	183.52	71.26	80.19	126.71	<u>30.18</u>	352.36	34.28
0.3	MPSNR	9.75	39.10	38.69	30.91	41.67	23.48	<u>44.77</u>	44.20	**47.34**
	MSSIM	0.0750	0.9890	0.9900	0.9430	0.9830	0.6020	<u>0.9960</u>	0.9900	**0.9979**
	MFSIM	0.4910	0.9890	0.9920	0.9640	0.9890	0.8140	<u>0.9980</u>	0.9930	**0.9987**
	ERGAS	1210.8	95.28	58.77	109.37	55.73	244.06	<u>22.76</u>	40.25	**20.73**
	MSAM	47.77	1.64	1.70	5.88	4.70	13.66	<u>1.45</u>	3.67	**1.29**
	Times	/	**4.68**	122.71	48.22	71.34	124.86	41.42	139.44	<u>24.56</u>
0.4	MPSNR	8.51	34.25	36.22	30.50	36.84	20.47	<u>42.47</u>	41.89	**45.68**
	MSSIM	0.0480	0.9750	0.9830	0.9390	0.9660	0.3730	<u>0.9940</u>	0.9860	**0.9971**
	MFSIM	0.4340	0.9790	0.9880	0.9580	0.9790	0.7230	<u>0.9970</u>	0.9910	**0.9982**
	ERGAS	1398.3	130.85	75.09	112.64	69.88	341.05	<u>29.31</u>	45.29	**24.49**
	MSAM	48.76	2.18	2.14	5.94	6.31	21.62	<u>1.73</u>	4.01	**1.46**
	Times	/	**5.41**	181.80	69.68	81.77	122.99	39.48	155.01	<u>24.38</u>
0.5	MPSNR	7.53	29.55	33.84	24.34	26.62	17.70	<u>39.22</u>	38.35	**42.68**
	MSSIM	0.0320	0.9330	0.9710	0.6920	0.7440	0.2100	<u>0.9860</u>	0.9770	**0.9949**
	MFSIM	0.3940	0.9590	0.9810	0.8130	0.8820	0.6280	<u>0.9930</u>	0.9850	**0.9968**
	ERGAS	1564.1	174.67	94.08	221.35	173.54	467.76	<u>41.93</u>	55.62	**33.38**
	MSAM	49.08	3.23	2.79	8.35	15.95	30.13	<u>2.24</u>	4.93	**1.96**
	Times	/	**4.63**	160.38	41.61	91.40	93.01	30.50	140.86	<u>19.64</u>
0.6	MPSNR	6.74	24.99	31.34	20.92	17.09	15.38	<u>31.91</u>	29.63	**38.86**
	MSSIM	0.0220	0.8260	<u>0.9490</u>	0.4470	0.2340	0.1160	0.8870	0.8550	**0.9847**
	MFSIM	0.3660	0.9170	<u>0.9700</u>	0.7080	0.6410	0.5480	0.9460	0.9190	**0.9908**
	ERGAS	1713.1	243.65	117.44	328.18	513.59	610.53	<u>94.59</u>	123.22	**52.78**
	MSAM	49.04	5.74	<u>4.79</u>	8.61	33.63	36.86	6.10	11.54	**4.44**
	Times	/	**6.59**	121.03	58.63	120.21	130.77	41.85	172.51	<u>19.53</u>

Table 3: Quantitative comparison of all competing methods on **Akiyo** dataset under difference sampling ratio (SR). The best and second results are highlighted in bold italics and underline, respectively.

SR	Metric	LRMC	HaLRTC	KBR	TNN	CTNN	FTNN	OITNN	TCTV	S2NTNN	Ours
0.05	MPSNR	10.80	17.66	29.77	31.95	28.64	22.74	32.69	33.42	33.16	**33.73**
	MSSIM	0.2620	0.5300	0.9110	0.9340	0.8460	0.7090	0.9530	0.9530	0.9519	**0.9566**
	MFSIM	0.6590	0.7480	0.9440	0.9620	0.9190	0.8440	0.9700	0.9690	0.9716	**0.9778**
	ERGAS	706.08	322.44	79.83	63.42	91.09	190.88	58.52	53.17	55.06	**52.01**
	MSAM	19.94	7.17	2.53	2.40	4.05	6.87	1.98	2.13	2.02	**1.84**
	Times	**8.06**	61.04	696.93	217.47	188.90	1204.61	397.54	874.80	99.96	79.89
0.1	MPSNR	22.75	21.68	**38.94**	34.95	32.11	27.88	36.01	37.54	36.40	37.84
	MSSIM	0.6760	0.6670	**0.9870**	0.9630	0.9200	0.8480	0.9760	0.9800	0.9757	0.9807
	MFSIM	0.8520	0.8120	**0.9910**	0.9780	0.9550	0.9130	0.9840	0.9870	0.9848	0.9890
	ERGAS	183.5	201.9	**28.8**	45.7	61.4	104.2	40.9	33.9	38.3	32.8
	MSAM	5.35	5.22	**0.95**	1.76	2.85	4.18	1.42	1.32	1.45	1.13
	Times	**10.51**	42.92	689.19	197.99	175.24	870.32	347.57	876.51	99.67	92.87
0.2	MPSNR	39.04	25.23	**45.17**	39.09	36.70	33.73	40.23	41.95	39.66	41.07
	MSSIM	0.9860	0.7960	**0.9960**	0.9840	0.9690	0.9680	0.9890	0.9920	0.9877	0.9910
	MFSIM	0.9920	0.8810	**0.9970**	0.9900	0.9820	0.9790	0.9930	0.9940	0.9919	0.9950
	ERGAS	29.30	134.11	**14.50**	29.39	36.70	51.59	26.06	21.17	26.48	22.41
	MSAM	1.00	3.94	**0.52**	1.16	1.71	1.46	0.94	0.82	1.03	0.78
	Times	**10.46**	35.44	835.68	220.64	179.44	655.12	381.47	922.82	100.95	158.19
0.3	MPSNR	44.39	27.67	**48.81**	42.08	40.02	36.89	43.23	44.82	42.02	45.17
	MSSIM	0.9950	0.8670	**0.9980**	0.9910	0.9850	0.9830	0.9940	0.9950	0.9924	0.9960
	MFSIM	0.9970	0.9210	**0.9990**	0.9940	0.9910	0.9890	0.9960	0.9970	0.9949	0.9970
	ERGAS	16.61	101.26	**9.41**	21.35	25.26	36.21	18.88	15.52	20.35	14.04
	MSAM	0.58	3.21	**0.37**	0.86	1.17	0.98	0.70	0.61	0.79	0.55
	Times	**9.02**	25.10	888.15	207.23	181.30	545.51	382.85	905.03	102.42	189.51

Table 4: Quantitative comparison of all competing methods on **Carphone** dataset under difference sampling ratio (SR). The best and second results are highlighted in bold italics and underline, respectively.

SR	Metric	LRMC	HaLRTC	KBR	TNN	CTNN	FTNN	OITNN	TCTV	S2NTNN	Ours
0.05	MPSNR	11.58	14.20	26.49	26.27	25.06	25.43	27.14	**29.10**	27.33	27.44
	MSSIM	0.2710	0.3440	0.8160	0.7650	0.7260	0.7770	0.8340	**0.8740**	0.8090	0.8095
	MFSIM	0.6470	0.6410	0.8920	0.8820	0.8590	0.8810	0.9060	**0.9240**	0.9025	0.9056
	ERGAS	676.72	499.98	122.16	127.62	144.13	139.69	115.92	**91.71**	112.74	110.17
	MSAM	22.09	13.58	5.69	6.96	7.69	7.83	5.90	**5.06**	6.01	5.87
	Times	**6.92**	21.12	798.21	493.22	195.56	1135.70	472.97	1103.24	100.76	80.11
0.1	MPSNR	21.88	19.79	32.00	28.23	27.84	28.16	29.31	31.29	30.31	30.38
	MSSIM	0.6230	0.5890	**0.9260**	0.8240	0.8160	0.8560	0.8800	0.9110	0.8843	0.8870
	MFSIM	0.8160	0.7800	**0.9550**	0.9110	0.9050	0.9210	0.9310	0.9470	0.9371	0.9380
	ERGAS	210.44	262.79	**65.14**	102.32	104.97	102.30	90.66	71.57	80.76	80.88
	MSAM	9.36	9.89	**3.30**	5.76	5.89	5.83	4.79	4.03	4.31	4.34
	Times	**7.91**	42.07	606.52	158.53	136.74	722.46	275.82	759.54	98.32	88.81
0.2	MPSNR	30.97	19.73	**36.63**	30.94	31.27	31.04	32.04	33.59	33.34	33.71
	MSSIM	0.9080	0.5540	0.9660	0.8880	0.8970	0.9130	0.9230	0.9413	0.9321	0.9310
	MFSIM	0.9520	0.7610	**0.9800**	0.9420	0.9460	0.9510	0.9560	0.9640	0.9617	0.9650
	ERGAS	76.79	264.66	**38.82**	75.34	71.01	73.78	66.53	54.29	56.86	53.33
	MSAM	3.98	10.08	**2.17**	4.37	4.15	4.24	3.65	3.08	3.15	3.06
	Times	**9.27**	21.94	849.07	416.23	166.91	608.54	337.33	884.53	100.88	151.87
0.3	MPSNR	34.84	23.64	**39.58**	33.01	33.80	33.21	34.11	35.83	35.38	36.40
	MSSIM	0.9550	0.7280	**0.9800**	0.9230	0.9360	0.9420	0.9470	0.9610	0.9531	0.9580
	MFSIM	0.9750	0.8510	**0.9890**	0.9600	0.9660	0.9670	0.9700	0.9770	0.9733	0.9780
	ERGAS	50.08	168.64	**27.80**	59.55	53.23	57.66	52.57	42.77	45.02	40.18
	MSAM	2.61	7.11	**1.62**	3.51	3.17	3.33	2.95	2.49	2.56	2.33
	Times	**12.92**	29.82	881.66	409.21	170.54	462.46	342.69	870.83	93.99	196.46

Table 5: Quantitative comparison of all competing methods on **WDC** dataset under difference sampling ratio (SR). The best and second results are highlighted in bold italics and underline, respectively.

SR	Metric	LRMC	HaLRTC	KBR	TNN	CTNN	FTNN	OITNN	TCTV	S2NTNN	Ours
0.01	MPSNR	14.70	18.12	20.27	22.86	22.91	21.42	25.75	26.30	**27.49**	26.62
	MSSIM	0.0480	0.2890	0.3590	0.5510	0.5130	0.5490	0.7350	0.7360	**0.8165**	0.7630
	MFSIM	0.4840	0.5410	0.6100	0.7850	0.7550	0.7780	0.8590	0.8520	**0.9031**	0.8830
	ERGAS	773.63	512.73	399.62	299.11	295.56	402.06	219.65	201.04	**174.34**	197.95
	MSAM	21.45	17.85	16.99	16.71	14.53	21.45	13.06	12.23	**10.73**	12.15
	Times	**14.00**	88.58	1349.8	470.75	472.85	2472.0	957.36	1720.7	162.30	113.20
0.05	MPSNR	18.54	22.01	31.42	30.06	33.36	34.70	32.29	33.33	37.30	**38.06**
	MSSIM	0.4620	0.6670	0.9020	0.8800	0.9430	0.9530	0.9270	0.9390	0.9749	**0.9790**
	MFSIM	0.7610	0.8350	0.9420	0.9350	0.9660	0.9710	0.9570	0.9640	0.9846	**0.9870**
	ERGAS	521.83	374.09	117.56	132.83	90.49	83.78	106.21	91.52	56.89	**52.93**
	MSAM	17.24	23.33	7.37	10.46	6.64	6.82	8.27	7.64	4.71	**4.35**
	Times	**24.38**	54.38	1589.7	1019.6	378.99	4376.2	838.68	2116.5	168.75	149.70
0.1	MPSNR	27.62	29.55	**44.11**	33.58	39.48	39.75	36.40	37.68	40.57	43.47
	MSSIM	0.8620	0.9110	**0.9930**	0.9400	0.9680	0.9810	0.9670	0.9740	0.9868	0.9920
	MFSIM	0.9300	0.9460	0.9960	0.9660	0.9870	0.9880	0.9800	0.9840	0.9918	**0.9962**
	ERGAS	215.13	176.03	**28.86**	90.43	43.49	51.46	68.77	57.41	40.11	29.80
	MSAM	11.86	8.78	**2.64**	7.95	3.72	4.88	6.05	5.34	3.57	2.74
	Times	**20.87**	54.75	1585.2	1000.7	358.34	3021.0	828.60	2088.7	203.58	245.32
0.2	MPSNR	46.38	25.21	**50.52**	38.23	47.35	45.45	41.29	42.58	43.73	48.91
	MSSIM	0.9950	0.7040	**0.9980**	0.9750	0.9970	0.9930	0.9860	0.9890	0.9931	0.9974
	MFSIM	0.9970	0.8330	**0.9990**	0.9850	0.9980	0.9950	0.9910	0.9930	0.9958	0.9984
	ERGAS	24.94	224.65	**14.87**	55.10	19.77	31.12	41.80	35.07	28.33	17.32
	MSAM	2.45	11.85	**1.52**	5.33	1.96	3.21	4.04	3.50	2.65	1.69
	Times	**36.58**	53.34	1893.5	462.46	383.57	3128.3	881.75	2203.5	163.28	280.35

Table 6: Quantitative comparison of all competing methods on **Cloth** dataset under difference sampling ratio (SR). The best and second results are highlighted in bold italics and underline, respectively.

SR	Metric	LRMC	HaLRTC	KBR	TNN	CTNN	FTNN	OITNN	TCTV	S2NTNN	Ours
0.01	MPSNR	11.82	16.46	17.47	18.03	18.16	16.16	19.27	**22.68**	20.43	18.71
	MSSIM	0.0280	0.3050	0.2790	0.2270	0.2750	0.2430	0.3360	**0.5850**	0.3895	0.3425
	MFSIM	0.4230	0.5090	0.5390	0.6320	0.6440	0.7010	0.6810	**0.8340**	0.8017	0.7046
	ERGAS	904.39	539.99	487.47	458.03	453.74	549.95	394.66	**264.38**	344.39	420.81
	MSAM	24.56	17.90	21.91	21.14	20.91	22.47	17.24	**11.03**	14.02	17.64
	Times	**10.54**	110.61	1223.6	412.56	124.18	1930.5	517.37	1430.0	82.32	28.04
0.05	MPSNR	13.10	19.00	24.14	23.46	25.70	25.26	24.01	**28.39**	27.44	25.81
	MSSIM	0.1900	0.3570	0.6420	0.6010	0.7360	0.7250	0.6510	**0.8440**	0.7589	0.7340
	MFSIM	0.6250	0.6100	0.8800	0.8670	0.9170	0.9110	0.8790	**0.9540**	0.9491	0.9270
	ERGAS	783.44	417.77	223.50	240.91	183.60	193.82	226.05	**135.61**	151.35	182.67
	MSAM	22.17	15.20	9.51	12.01	8.92	8.82	10.94	**6.50**	7.21	8.65
	Times	**10.95**	92.65	1292.4	441.03	136.16	2054.4	391.88	1488.6	86.35	121.61
0.1	MPSNR	15.96	20.62	31.87	26.71	29.23	29.49	27.45	31.78	**32.21**	29.65
	MSSIM	0.3970	0.4540	0.9080	0.7640	0.8720	0.8680	0.8020	**0.9140**	0.9061	0.8510
	MFSIM	0.7950	0.7080	0.9800	0.9340	0.9640	0.9600	0.9420	0.9780	**0.9802**	0.9660
	ERGAS	574.58	346.60	90.79	164.37	119.83	118.33	150.88	92.23	**89.16**	116.70
	MSAM	18.93	12.61	**4.52**	8.90	5.84	5.97	8.02	4.81	4.77	6.18
	Times	**10.61**	73.22	1285.5	437.32	131.46	1553.0	397.20	1459.8	88.22	182.20
0.2	MPSNR	24.78	23.25	**38.54**	31.09	34.28	34.38	31.87	35.87	37.75	35.21
	MSSIM	0.7310	0.6320	**0.9730**	0.8890	0.9420	0.9440	0.9070	0.9580	0.9672	0.9440
	MFSIM	0.9460	0.8400	**0.9950**	0.9740	0.9810	0.9850	0.9780	0.9910	0.9938	0.9893
	ERGAS	233.65	255.15	**44.78**	99.67	68.33	69.51	91.39	58.83	49.15	66.54
	MSAM	9.24	9.55	**2.66**	5.89	3.83	3.90	5.35	3.39	2.90	3.80
	Times	**9.85**	49.62	1221.9	356.16	117.28	1052.4	412.30	1402.2	88.44	226.42

transactions on pattern analysis and machine intelligence, 42(4):925–938, 2019.

[2] Tamara G Kolda and Brett W Bader. Tensor decompositions and applications. *SIAM review*, 51(3):455–500, 2009.

[3] Emmanuel J Candès, Xiaodong Li, Yi Ma, and John Wright. Robust principal component analysis? *Journal of the ACM (JACM)*, 58(3):1–37, 2011.

[4] Roman Vershynin. Introduction to the non-asymptotic analysis of random matrices. *arXiv preprint arXiv:1011.3027*, 2010.

[5] Rudolf Ahlswede and Andreas Winter. Strong converse for identification via quantum channels. *IEEE Transactions on Information Theory*, 48(3):569–579, 2002.

[6] Joel A Tropp. User-friendly tail bounds for sums of random matrices. *Foundations of computational mathematics*, 12:389–434, 2012.

[7] Jianfeng Cai, Emmanuel J Candès, and Zuowei Shen. A singular value thresholding algorithm for matrix completion. *SIAM Journal on optimization*, 20(4):1956–1982, 2010.

[8] David L Donoho. De-noising by soft-thresholding. *IEEE transactions on information theory*, 41(3):613–627, 1995.

[9] Jiangjun Peng, Qi Xie, Qian Zhao, Yao Wang, Leung Yee, and Deyu Meng. Enhanced 3dtv regularization and its applications on hsi denoising and compressed sensing. *IEEE Transactions on Image Processing*, 29:7889–7903, 2020.

[10] Stephen Boyd, Neal Parikh, Eric Chu, Borja Peleato, Jonathan Eckstein, et al. Distributed optimization and statistical learning via the alternating direction method of multipliers. *Foundations and Trends® in Machine learning*, 3(1):1–122, 2011.

[11] Zhouchen Lin, Minming Chen, and Yi Ma. The augmented lagrange multiplier method for exact recovery of corrupted low-rank matrices. *arXiv preprint arXiv:1009.5055*, 2010.