# Supplementary Material For " Exact Decomposition of Joint Low-rankness and Local Smoothness Plus Sparse Matrices"

Jiangjun Peng, Yao Wang, *Member, IEEE*, Hongying Zhang, Jianjun Wang, *Member, IEEE*, and Deyu Meng, *Member, IEEE*

**Abstract**—In the document, we give the detailed proofs of our theoretical assertions (i.e., Theorem 1) presented in the manuscript in Section 1. Next, we give the detailed proof of Theorem 2 presented in the manuscript in Section 2. Furthermore, in Section 3, we suggest two extended models based on the proposed 3DCTV regularizer, which tends to get better performance in certain scenarios. The more evaluations on the hyper-parameter robustness issue, experiments on extending CTV to its tensor form and image inpainting experiments to validate the generalization of 3DCTV are provided in Sections 4, 5 and 6, respectively.

## 1 Main Proofs

### 1.1 The Equivalent Model of 3DCTV-RPCA

The 3DCTV-RPCA model is:

$$\begin{aligned} \min_{\mathbf{X},\mathbf{S}} \quad & \sum_{i=1}^{3} \|\mathbf{G}_i\|_* + 3\lambda\|\mathbf{S}\|_1 \\ \text{s.t.} \quad & \mathbf{M} = \mathbf{X} + \mathbf{S}, \\ & \mathbf{G}_i = \nabla_i(\mathbf{X}), i = 1, 2, 3, \end{aligned} \tag{1}$$

where $\mathbf{X} \in \mathbb{R}^{hw \times p}$, and $\nabla_i(\cdot), i = 1, 2, 3$ are the differential operators. Since such differential operators are linear, the three gradient maps $\mathbf{G}_i = \nabla_i \mathbf{X}, i = 1, 2, 3$ can be transformed in the form of matrix product, i.e.,

$$\begin{aligned} \mathbf{G}_1 &= \mathbf{A}_1\mathbf{X}_0, \mathbf{A}_1 = \mathbf{circ}(-1, 1, \overbrace{0, \ldots, 0}^{n_1-2}), \\ \mathbf{G}_2 &= \mathbf{A}_2\mathbf{X}_0, \mathbf{A}_2 = \mathbf{circ}(-1, \overbrace{0, \ldots, 0}^{h}, 1, \overbrace{0, \ldots, 0}^{n_1-h-2}), \\ \mathbf{G}_3 &= \mathbf{X}_0\mathbf{B}_1, \mathbf{B}_1 = \mathbf{circ}(-1, \overbrace{0, \ldots, 0}^{n_1-2}, 1), \end{aligned} \tag{2}$$

where "**circ**" denotes the circulant matrix, and $n_1 = h \times w, n_2 = p$. Thus, our 3DCTV-RPCA model can be transformed into the following form:

$$\begin{aligned} \min_{\mathbf{X},\mathbf{E}} \quad & \|\mathbf{A}_1\mathbf{X}\|_* + \|\mathbf{A}_2\mathbf{X}\|_* + \|\mathbf{X}\mathbf{B}_1\|_* + 3\lambda\|\mathbf{S}\|_1 \\ \text{s.t.} \quad & \mathbf{M} = \mathbf{X} + \mathbf{S}. \end{aligned} \tag{3}$$

Note that if $\mathbf{A}_1, \mathbf{A}_2$ and $\mathbf{B}_1$ in Eq. (3) are expanded/reduced by $N$ times, then the trade-off coefficient $\lambda$ needs to be expanded/reduced in the same proportion. Therefore, we make the $\ell_2$-norm of $\mathbf{A}_1, \mathbf{A}_2$, and $\mathbf{B}_1$ be one by setting $\mathbf{A}_1 := \mathbf{A}_1/\|\mathbf{A}_1\|_2, \mathbf{A}_2 := \mathbf{A}_2/\|\mathbf{A}_2\|_2$ and $\mathbf{B}_1 := \mathbf{B}_1/\|\mathbf{B}_1\|_2$.

### 1.2 Mathematical Preliminaries

Suppose we are given a large data matrix $\mathbf{M}$, and know that it can be decomposed as:

$$\mathbf{M} = \mathbf{X}_0 + \mathbf{S}_0, \tag{4}$$

where $\mathbf{X}_0, \mathbf{S}_0$ are the joint low-rankness and local smoothness component, and the sparse component, respectively. Theorem 1 in the main text has asserted that by solving the 3DCTV-RPCA model, we can get the exact decomposition $(\mathbf{X}_0, \mathbf{S}_0)$. Before proving theorem 1, it is helpful to review some basic concepts and introduce some notations.

For a give scalar $x$, we use $\mathrm{sgn}(x)$ to denote the sign of $x$. By extension, $\mathrm{sgn}(\mathbf{S})$ is the matrix whose entries are the signs of those of $\mathbf{S}$. We recall that any sub-gradient of $\ell_1$ norm at $\mathbf{S_0}$ supported on $\Omega$, is of the form

$$\mathrm{sgn}(\mathbf{S}_0) + \mathbf{F},$$

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

*J. Peng and H. Zhang are with School of Mathematics and Statistics and Ministry of Education Key Lab of Intelligent Networks and Network Security, Xi'an Jiaotong University, Xi'an 710049, Shaan'xi, China. (email: andrew.pengjj@gmail.com, zhyemily@mail.xjtu.edu.cn).*
*Y. Wang is with the Center for Intelligent Decision-making and Machine Learning, School of Management, Xian Jiaotong University, Xi'an, Shaan'xi, China. (email: yao.s.wang@gmail.com).*
*J. Wang is with the College of Artificial Intelligence, Southwest University, Chongqing, 400715, China. (email: wjj@swu.edu.cn).*
*D. Meng is with the School of Mathematics and Statistics and Ministry of Education Key Lab of Intelligent Networks and Network Security, Xian Jiaotong University, Xian, Shaan'xi, China and Macau Institute of Systems Engineering, Macau University of Science and Technology, Taipa, Macau, China. (email: dymeng@mail.xjtu.edu.cn).*

We then assume that each gradient map $\mathbf{G}_i, (i = 1, 2, 3)$ of rank $r$ has the singular value decomposition $\mathbf{U}_i\Sigma_i\mathbf{V}_i^T$, where $\mathbf{U}_i \in \mathbb{R}^{n_1\times r}$ and $\mathbf{V}_i \in \mathbb{R}^{n_2\times r}$. Then according to the chain rule of derivation, we can get

$$\begin{aligned}\frac{\partial\|\mathbf{G}_i\|_*}{\partial\mathbf{X}_0} &= \frac{\partial\|\mathbf{A}_i\mathbf{X}_0\|_*}{\partial\mathbf{X}_0} = \mathbf{A}_i^T\mathbf{U}_i\mathbf{V}_i^T + \mathbf{A}_i^T\mathbf{W}_i, i = 1, 2,\\ \frac{\partial\|\mathbf{G}_i\|_*}{\partial\mathbf{X}_0} &= \frac{\partial\|\mathbf{X}_0\mathbf{B}_1\|_*}{\partial\mathbf{X}_0} = \mathbf{U}_i\mathbf{V}_i^T\mathbf{B}_1^T + \mathbf{W}_i\mathbf{B}_1^T, i = 3,\end{aligned} \tag{5}$$

where $\mathbf{U}_i^T\mathbf{W}_i = \mathbf{0}, \mathbf{W}_i\mathbf{V}_i = \mathbf{0}$ and $\|\mathbf{W}_i\| \leq 1, (i = 1, 2, 3)$. We use $T_i, (i = 1, 2, 3)$ to denote the linear space of matrices, i.e.,

$$T_i := \{\mathbf{U}_i\mathbf{X}^T + \mathbf{Y}\mathbf{V}_i^T, \mathbf{X} \in \mathbb{R}^{n_1\times r}, \mathbf{Y} \in \mathbb{R}^{n_2\times r}\}, (i = 1, 2, 3)$$

and $T_i^\perp (i = 1, 2, 3)$ to denote its orthogonal complement. For any matrix $\mathbf{M}$, the projections onto $T_i$ and $T_i^\perp$ are

$$\begin{aligned}\mathcal{P}_{T_i}\mathbf{M} &= \mathbf{U}_i\mathbf{U}_i^T\mathbf{M} + \mathbf{M}\mathbf{V}_i\mathbf{V}_i^T - \mathbf{U}_i\mathbf{U}_i^T\mathbf{M}\mathbf{V}_i\mathbf{V}_i^T,\\ \mathcal{P}_{T_i^\perp}\mathbf{M} &= (\mathbf{I} - \mathbf{U}_i\mathbf{U}_i^T)\mathbf{M}(\mathbf{I} - \mathbf{V}_i\mathbf{V}_i^T), (i = 1, 2, 3),\end{aligned} \tag{6}$$

where $\mathbf{I}$ is the identity matrix.

Therefore, for any matrix of the form $\hat{\mathbf{e}}_k\hat{\mathbf{e}}_j^T$, it is easy to see that

$$\begin{aligned}&\|\mathcal{P}_{T_i^\perp}\hat{\mathbf{e}}_k\hat{\mathbf{e}}_j^T\|_F^2 = \|(\mathbf{I} - \mathbf{U}_i\mathbf{U}_i^T)\hat{\mathbf{e}}_k\|^2\|(\mathbf{I} - \mathbf{V}_i\mathbf{V}_i^T)\hat{\mathbf{e}}_j\|^2\\ &\geqslant (1 - \mu r/n_1)\,(1 - \mu r/n_2)\,.\end{aligned}$$

We assume $\mu r/n_{(1)} \leq 1$, and $n_{(1)} = \max\{n_1, n_2\}, n_{(2)} = \min\{n_1, n_2\}$. Since $\|\mathcal{P}_{T_i^\perp}\hat{\mathbf{e}}_k\hat{\mathbf{e}}_j^T\|_F^2 + \|\mathcal{P}_{T_i}\hat{\mathbf{e}}_k\hat{\mathbf{e}}_j^T\|_F^2 = 1$, this gives

$$\|\mathcal{P}_{T_i}\hat{\mathbf{e}}_k\hat{\mathbf{e}}_j^T\|_F \leq \sqrt{\frac{2\mu r}{n_{(2)}}}, (i = 1, 2, 3). \tag{7}$$

The following incoherence conditions will be used in our proofs:

$$\max_k \|\mathbf{U}_i^T\hat{e}_k\|^2 \leq \frac{\mu r}{n_1}, (i = 1, 2, 3), \tag{8}$$

$$\max_k \|\mathbf{V}_i^T\hat{e}_k\|^2 \leq \frac{\mu r}{n_2}, (i = 1, 2, 3), \tag{9}$$

and

$$\|\mathbf{U}_i\mathbf{V}_i^T\|_\infty \leq \sqrt{\frac{\mu r}{n_1 n_2}}, (i = 1, 2, 3). \tag{10}$$

### 1.3 An Elimination Lemma

We begin with a helpful definition.

**Definition 1:** We say that $\mathbf{S}^{'}$ is a trimmed version of $\mathbf{S}$ if $supp(\mathbf{S}^{'}) \subset supp(\mathbf{S})$ and $\mathbf{S}_{ij}^{'} = \mathbf{S}_{ij}$ whenever $\mathbf{S}_{ij}^{'} \neq 0$.

In other words, a trimmed version of $\mathbf{S}$ is obtained by setting some of the entries of $\mathbf{S}$ to be zeros. And then, the following elimination theorem asserts that if the Eq. (3) recovers the low-rank and sparse components of $\mathbf{M}_0 = \mathbf{X}_0 + \mathbf{S}_0$, it also correctly recovers the components of $\mathbf{M}_0^{'} = \mathbf{L}_0 + \mathbf{S}_0^{'}$, where $\mathbf{S}_0^{'}$ is a trimmed version of $\mathbf{S}_0$.

**Lemma 1:** Suppose the solution to Eq. (3) is exact with input $\mathbf{M}_0 = \mathbf{X}_0 + \mathbf{S}_0$, and consider $\mathbf{M}_0^{'} = \mathbf{X}_0 + \mathbf{S}_0^{'}$, where $\mathbf{S}_0^{'}$ is a trimmed version of $\mathbf{S}_0$. Then, the solution to the Eq. (3) is exact as well with input $\mathbf{M}_0^{'} = \mathbf{X}_0 + \mathbf{S}_0^{'}$.

**Proof:** We first write $\mathbf{S}_0^{'} = \mathbf{P}_{\Omega_0}\mathbf{S}_0$ for some $\Omega_0 \subset [n] \times [n]$ and let $(\widehat{\mathbf{X}}, \widehat{\mathbf{S}})$ be the solution to Eq. (3) with input $\mathbf{M}_0^{'} = \mathbf{X}_0 + \mathbf{S}_0^{'}$. Then

$$\begin{aligned}&\sum_{i=1}^{2}\|\mathbf{A}_i\widehat{\mathbf{X}}\|_* + \|\widehat{\mathbf{X}}\mathbf{B}_1\|_* + 3\lambda\|\widehat{\mathbf{S}}\|_1\\ \leq &\sum_{i=1}^{2}\|\mathbf{A}_i\mathbf{X}_0\|_* + \|\mathbf{X}_0\mathbf{B}_1\|_* + 3\lambda\|\mathcal{P}_{\Omega_0}\mathbf{S}_0\|_1\end{aligned}$$

and therefore,

$$\begin{aligned}&\sum_{i=1}^{2}\|\mathbf{A}_i\widehat{\mathbf{X}}\|_* + \|\widehat{\mathbf{X}}\mathbf{B}_1\|_* + 3\lambda\left(\|\widehat{\mathbf{S}}\|_1 + \|\mathcal{P}_{\Omega_0^\perp}\mathbf{S}_0\|_1\right)\\ \leq &\sum_{i=1}^{2}\|\mathbf{A}_i\mathbf{X}_0\|_* + \|\mathbf{X}_0\mathbf{B}_1\|_* + 3\lambda\|\mathbf{S}_0\|_1.\end{aligned}$$

Note that $(\widehat{\mathbf{X}}, \widehat{\mathbf{S}} + \mathcal{P}_{\Omega_0^\perp}\mathbf{S}_0)$ is a feasible solution to Eq. (3) under the condition that the measurement matrix satisfies $\mathbf{M}_0 = \mathbf{X}_0 + \mathbf{S}_0$, and $\|\widehat{\mathbf{S}} + \mathcal{P}_{\Omega_0^\perp}\mathbf{S}_0\|_1 \leq \|\widehat{\mathbf{S}}\|_1 + \|\mathcal{P}_{\Omega_0^\perp}\mathbf{S}_0\|_1$, we thus have

$$\begin{aligned}&\sum_{i=1}^{2}\|\mathbf{A}_i\widehat{\mathbf{X}}\|_* + \|\widehat{\mathbf{X}}\mathbf{B}_1\|_* + 3\lambda\left(\|\widehat{\mathbf{S}} + \mathcal{P}_{\Omega_0^\perp}\mathbf{S}_0\|_1\right)\\ \leq &\sum_{i=1}^{2}\|\mathbf{A}_i\widehat{\mathbf{X}}\|_* + \|\widehat{\mathbf{X}}\mathbf{B}_1\|_* + 3\lambda\left(\|\widehat{\mathbf{S}}\|_1 + \|\mathcal{P}_{\Omega_0^\perp}\mathbf{S}_0\|_1\right)\\ \leq &\sum_{i=1}^{2}\|\mathbf{A}_i\mathbf{X}_0\|_* + \|\mathbf{X}_0\mathbf{B}_1\|_* + 3\lambda\|\mathbf{S}_0\|_1.\end{aligned}$$

The right-hand side above, however, is the optimal value, and by uniqueness of the optimal solution, we must have each $\mathbf{A}_i\widehat{\mathbf{X}} = \mathbf{A}_i\mathbf{X}_0$, and $\widehat{\mathbf{X}}\mathbf{B}_1 = \mathbf{X}_0\mathbf{B}_1$, and $\widehat{\mathbf{S}} + \mathcal{P}_{\Omega_0^\perp}\mathbf{S}_0 = \mathbf{S}_0$. So we further have $\widehat{\mathbf{X}} = \mathbf{X}_0$ and $\widehat{\mathbf{S}} = \mathcal{P}_{\Omega_0}\mathbf{S}_0 = \mathbf{S}^{'}$. This proves the claim. ■

### 1.4 Dual Certificates

**Lemma 2:** Assume that $\|\mathcal{P}_\Omega\mathcal{P}_{T_i}\| \leq 1, (i = 1, 2, 3)$. Then $(\mathbf{X}_0, \mathbf{S}_0)$ is the unique solution to the Eq. (3) if there exist a pair $(\{\mathbf{W}_i\}_{i=1}^3, \mathbf{F})$ obeying

$$\begin{aligned}&\sum_{i=1}^{2}\{\mathbf{A}_i^T\mathbf{U}_i\mathbf{V}_i^T + \mathbf{A}_i^T\mathbf{W}_i\} + \mathbf{U}_3\mathbf{V}_3^T\mathbf{B}_1^T + \mathbf{W}_3\mathbf{B}_1^T\\ &= 3\lambda(\mathrm{sgn}(\mathbf{S}_0) + \mathbf{F})\end{aligned} \tag{11}$$

with $\mathcal{P}_{T_i}(\mathbf{W}_i) = 0, \|\mathbf{W}_i\| < 1, (i = 1, 2, 3), \mathcal{P}_\Omega(\mathbf{F}) = 0$ and $\|\mathbf{F}\|_\infty < 1$.

**Lemma 3:** Assume that $\|\mathcal{P}_\Omega\mathcal{P}_{T_i}\| \leq 1/2, (i = 1, 2, 3)$. Then $(\mathbf{X}_0, \mathbf{S}_0)$ is the unique solution to the Eq. (3) if there exist a pair $(\{\mathbf{W}_i\}_{i=1}^3, \mathbf{F})$ obeying

$$\begin{aligned}&\sum_{i=1}^{2}\{\mathbf{A}_i^T\mathbf{U}_i\mathbf{V}_i^T + \mathbf{A}_i^T\mathbf{W}_i\} + \mathbf{U}_3\mathbf{V}_3^T\mathbf{B}_1^T + \mathbf{W}_3\mathbf{B}_1^T\\ &= 3\lambda(\mathrm{sgn}(\mathbf{S}_0) + \mathbf{F} + P_\Omega\mathbf{D})\end{aligned} \tag{12}$$

with $\mathcal{P}_{T_i}(\mathbf{W}_i) = 0, \|\mathbf{W}_i\| < 1/2, (i = 1, 2, 3), \|\mathbf{F}\|_\infty < 1$ and $\|P_\Omega\mathbf{D}\|_\infty \leq 1/4$.

The above lemmas imply that to prove our final exact recovery result, it is sufficient to produce two dual certificates $\mathbf{W}_1, \mathbf{W}_2$ obeying

$$\begin{cases} \mathbf{W}_i \in T_i^{\perp}, \\ \|\mathbf{W}_i\| < \frac{1}{2}, \\ \|\mathcal{P}_{\Omega}\left(\mathbf{A}_i^T\mathbf{U}_i\mathbf{V}_i^T + \mathbf{A}_i^T\mathbf{W}_i - \lambda\mathrm{sgn}(\mathbf{S}_0)\right)\|_F \leq \frac{\lambda}{4}, \\ \|\mathcal{P}_{\Omega^{\perp}}\left(\mathbf{A}_i^T\mathbf{U}_i\mathbf{V}_i^T + \mathbf{A}_i^T\mathbf{W}_i\right)\|_{\infty} < \frac{\lambda}{2}, \end{cases} \tag{13}$$

and a dual certificate $\mathbf{W}_3$ obeying

$$\begin{cases} \mathbf{W}_3 \in T_3^{\perp}, \\ \|\mathbf{W}_3\| < \frac{1}{2}, \\ \|\mathcal{P}_{\Omega}\left(\mathbf{U}_3\mathbf{V}_3^T\mathbf{B}_1^T + \mathbf{W}_3\mathbf{B}_1^T - \lambda\mathrm{sgn}(\mathbf{S}_0)\right)\|_F \leq \frac{\lambda}{4}, \\ \|\mathcal{P}_{\Omega^{\perp}}\left(\mathbf{U}_3\mathbf{V}_3^T\mathbf{B}_1^T + \mathbf{W}_3\mathbf{B}_1^T\right)\|_{\infty} < \frac{\lambda}{2}. \end{cases}$$

### 1.5 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 $\Omega^c$ is the same as that of $\Omega^c = \Omega_1 \cup \Omega_2 \cup \cdots \cup \Omega_{j_0}$, where each $\Omega_j$ follows the Bernoulli model with parameter $q$, that is,

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

so that the two models are the same if

$$\rho = (1-q)^{j_0}.$$

Now, taking dual certificate $\mathbf{W}_1$ as example, we can decompose it as

$$\mathbf{W}_1 = \mathbf{W}^L + \mathbf{W}^S,$$

where each component can be constructed in the following way.

**Construction of $\mathbf{W}^L$ via the Golfing scheme.** Let $j_0 \geqslant 1$, 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

$$\mathbf{W}^L = \mathcal{P}_{T_1^{\perp}}\mathbf{Y}_{j_0}, \tag{14}$$

where

$$\mathbf{Y}_j = \mathbf{Y}_{j-1} + q^{-1}\mathcal{P}_{\Omega_j}\mathcal{P}_{T_1}\left(\mathbf{U}_1\mathbf{V}_1^T - \mathbf{Y}_{j-1}\right), \mathbf{Y}_0 = 0. \tag{15}$$

**Construction of $\mathbf{W}^S$ via the Method of Least Squares.** Assume that $\|\mathcal{P}_{\Omega}\mathcal{P}_{T_1}\| \leq \frac{1}{2}$. Then, $\|\mathcal{P}_{\Omega}\mathcal{P}_{T_1}\mathcal{P}_{\Omega}\| < \frac{1}{4}$ and thus, the operator $\mathcal{P}_{\Omega} - \mathcal{P}_{\Omega}\mathcal{P}_{T_1}\mathcal{P}_{\Omega}$ mapping $\Omega$ onto itself is invertible, and its inverse is denoted by $\left(\mathcal{P}_{\Omega} - \mathcal{P}_{\Omega}\mathcal{P}_{T_1}\mathcal{P}_{\Omega}\right)^{-1}$. We then set

$$\mathbf{W}^S = \lambda\mathcal{P}_{T_1^{\perp}}\left(\mathcal{P}_{\Omega} - \mathcal{P}_{\Omega}\mathcal{P}_{T_1}\mathcal{P}_{\Omega}\right)^{-1}\left(\mathrm{sgn}(\mathbf{S_0})\right). \tag{16}$$

This is equivalent to

$$\mathbf{W}^S = \lambda\mathcal{P}_{T_1^{\perp}}\sum_{k\geqslant 0}\left(\mathcal{P}_{\Omega}\mathcal{P}_{T_1}\mathcal{P}_{\Omega}\right)^k\left(\mathrm{sgn}(\mathbf{S})\right). \tag{17}$$

Since both $\mathbf{W}^L$ and $\mathbf{W}^S$ belong to $T_1^{\perp}$ and $\mathcal{P}_{\Omega}\mathbf{W}^S = \lambda\mathcal{P}_{\Omega}(\mathcal{I} - \mathcal{P}_{T_1})\left(\mathcal{P}_{\Omega} - \mathcal{P}_{\Omega}\mathcal{P}_{T_1}\mathcal{P}_{\Omega}\right)^{-1}\left(\mathrm{sgn}(\mathbf{S_0})\right) = \lambda\mathrm{sgn}(\mathbf{S}_0)$, we shall establish that $\mathbf{W}^L + \mathbf{W}^S$ is a valid dual certificate if it obeys

$$\begin{cases} \|\mathbf{W}^L + \mathbf{W}^S\| < \frac{1}{2}, \\ \|\mathcal{P}_{\Omega}\left(\mathbf{A}_1^T\mathbf{U}_1\mathbf{V}_1^T + \mathbf{A}_1^T\mathbf{W}^L\right)\|_F \leq \frac{\lambda}{4}, \\ \|\mathcal{P}_{\Omega^{\perp}}\left(\mathbf{A}_1^T\mathbf{U}_1\mathbf{V}_1^T + \mathbf{A}_1^T\mathbf{W}^L + \mathbf{A}_1^T\mathbf{W}^S\right)\|_{\infty} \leq \frac{\lambda}{2}. \end{cases} \tag{18}$$

This can be done by using the following two lemmas.

**Lemma 4:** Assume that $\Omega \sim \mathrm{Ber}(\rho)$ 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 $\mathbf{W}^L$ in Eq. (14) obeys

(a) $\|\mathbf{W}^L\| < 1/4$,
(b) $\|\mathcal{P}_{\Omega}\left(\mathbf{A}_1^T\mathbf{U}_1\mathbf{V}_1^T + \mathbf{A}_1^T\mathbf{W}^L\right)\|_F \leq \lambda/4$,
(c) $\|\mathcal{P}_{\Omega^{\perp}}\left(\mathbf{A}_1^T\mathbf{U}_1\mathbf{V}_1^T + \mathbf{A}_1^T\mathbf{W}^L\right)\|_{\infty} < \lambda/4$.

**Lemma 5:** Assume $\Omega \sim \mathrm{Ber}(\rho_s)$, and the sign of $\mathbf{S}_0$ are independent and identically distributed symmetric (and independent of $\Omega$). Then, the matrix $\mathbf{W}^S$ with Eq. (16) obeys

(a) $\|\mathbf{W}^S\| < 1/4$,
(b) $\|\mathcal{P}_{\Omega^{\perp}}(\mathbf{A}_1^T\mathbf{W}^S)\|_{\infty} < \lambda/4$.

### 1.6 Proofs of Dual Certification

Before proving Lemma 2, we shall list the following three useful lemmas.

**Lemma 6:** (Lemma 4.1 in [1]): Suppose $\Omega_0 \sim \mathrm{Ber}(\rho_0)$. Then with high probability,

$$\|\mathcal{P}_T - \rho_0^{-1}\mathcal{P}_{T_1}\mathcal{P}_{\Omega_0}\mathcal{P}_{T_1}\| \leq \epsilon, \tag{19}$$

provided that $\rho_0 \geq C_0\epsilon^{-2}\beta\mu r\log n_{(1)}/n_{(2)}$ for some numerical constant $C_0 > 0$.

**Lemma 7:** (Lemma 3.1 in [2]) : Suppose $\mathbf{Z} \in T_1$ is a fixed matrix, and $\Omega_0 \sim \mathrm{Ber}(\rho_0)$. Then with high probability,

$$\|\mathbf{Z} - \rho_0^{-1}\mathcal{P}_{T_1}\mathcal{P}_{\Omega_0}\mathbf{Z}\|_{\infty} \leq \epsilon\|\mathbf{Z}\|_{\infty}, \tag{20}$$

provided that $\rho_0 \geq C_0\epsilon^{-2}\mu r\log n_{(1)}/n_{(2)}$ for some numerical constant $C_0 > 0$.

**Lemma 8:** (Lemma 6.3 in [1] and Lemma 3.2 in [2]): Suppose $\mathbf{Z}$ is fixed, and $\Omega_0 \sim \mathrm{Ber}(\rho_0)$. Then with high probability,

$$\|\left(\mathbf{I} - \rho_0^{-1}\mathcal{P}_{\Omega_0}\right)\mathbf{Z}\| \leq C_1\sqrt{\frac{\beta n_{(1)}\log n_{(1)}}{\rho_0}}\|\mathbf{Z}\|_{\infty}, \tag{21}$$

provided that $\rho_0 \geq C_1\mu\log n_{(1)}/n_{(2)}$ for some small numerical constant $C_1 > 0$.

#### 1.6.1 *Proof of Lemma 4*

**Proof:** We first introduce some notations. Setting

$$\mathbf{Z}_j = \mathbf{U}_1\mathbf{V}_1^T - \mathcal{P}_{T_1}\mathbf{Y}_j,$$

thus $\mathbf{Z}_j \in T_1$ for all $j \geq 0$. From the definition of $\mathbf{Y}_j$ (15), and $\mathbf{Y}_j \in \Omega^{\perp}$, we have

$$\begin{aligned} \mathbf{Z}_j &= (\mathcal{P}_{T_1} - q^{-1}\mathcal{P}_{T_1}\mathcal{P}_{\Omega_j}\mathcal{P}_{T_1})\mathbf{Z}_{j-1}, \\ \mathbf{Y}_j &= \mathbf{Y}_{j-1} + q^{-1}\mathcal{P}_{\Omega_j}\mathbf{Z}_{j-1}. \end{aligned}$$

Therefore, when

$$q \geq C_0\epsilon^{-2}\mu r\log n_{(1)}/n_{(2)}, \tag{22}$$

we have

$$\|\mathbf{Z}_j\|_\infty \leq \epsilon \|\mathbf{Z}_{j-1}\|_\infty \tag{23}$$

by Lemma 7. In particular, this gives that with high probability,

$$\|\mathbf{Z}_j\|_\infty \leq \epsilon^j \|\mathbf{U}_1\mathbf{V}_1^T\|_\infty. \tag{24}$$

When $q$ obeys Eq. (22), we have

$$\|\mathbf{Z}_j\|_F \leq \epsilon \|\mathbf{Z}_{j-1}\|_F \tag{25}$$

by Lemma 6. This further gives that with high probability,

$$\|\mathbf{Z}_j\|_F \leq \epsilon^j \|\mathbf{U}_1\mathbf{V}_1^T\|_F \leq \epsilon^j \sqrt{r}. \tag{26}$$

We assume $\epsilon \leq e^{-1}$.

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

$$\begin{aligned}
\|\mathbf{W}^L\| = \|\mathbf{P}_{T_1^\perp}\mathbf{Y}_{j_0}\|_\infty &\leq \sum_j \|q^{-1}\mathcal{P}_{T_1^\perp}\mathcal{P}_{\Omega_j}\mathbf{Z}_{j-1}\| \\
&\leq \sum_j \|\mathcal{P}_{T_1^\perp}(q^{-1}\mathcal{P}_{\Omega_j}\mathbf{Z}_{j-1} - \mathbf{Z}_{j-1})\| \\
&\leq \sum_j \|q^{-1}\mathcal{P}_{\Omega_j}\mathbf{Z}_{j-1} - \mathbf{Z}_{j-1}\| \\
&\leq C_1\sqrt{\frac{\beta n_{(1)}\log n_{(1)}}{q}} \sum_j \|\mathbf{Z}_{j-1}\|_\infty \\
&\leq C_1\sqrt{\frac{\beta n_{(1)}\log n_{(1)}}{q}} \sum_j \epsilon^j \|\mathbf{U}_1\mathbf{V}_1^T\|_\infty \\
&\leq \frac{C_1}{(1-\epsilon)}\sqrt{\frac{\beta n_{(1)}\log n_{(1)}}{q}} \|\mathbf{U}_1\mathbf{V}_1^T\|_\infty.
\end{aligned} \tag{27}$$

The fourth step is according to Lemma 8 and the fifth step can be directly obtained from Eq. (23). Now by using Eq. (22) and (10), we get

$$\|\mathbf{W}^L\| \leq C_2\epsilon$$

for some numerical constant $C_2$.

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

$$\begin{aligned}
\mathcal{P}_\Omega(\mathbf{U}_1\mathbf{V}_1^T + \mathbf{W}^L) &= \mathcal{P}_\Omega(\mathbf{U}_1\mathbf{V}_1^T + \mathcal{P}_{T_1^\perp}\mathbf{Y}_{j_0}) \\
&= \mathcal{P}_\Omega(\mathbf{U}_1\mathbf{V}_1^T - \mathcal{P}_{T_1}\mathbf{Y}_{j_0}) = \mathcal{P}_\Omega(\mathbf{Z}_{j_0}).
\end{aligned}$$

By using Eqs. (22) and (26) , we can get

$$\|\mathcal{P}_\Omega(\mathbf{Z}_{j_0})\|_F \leq \|\mathbf{Z}_{j_0}\|_F \leq \epsilon^{j_0}\sqrt{r}.$$

Since $\epsilon \leq e^{-1}$, $j_0 \geq 2\log n_{(1)}$ and $\epsilon^{j_0} \leq 1/n_{(1)}^2$, then $\epsilon^{j_0}\sqrt{r} \leq \lambda/4$ is satisfied with probability at least $1 - n_{(1)}^{-\beta}$ for all $\beta > 2$ . Since

$$\begin{aligned}
\|\mathcal{P}_\Omega(\mathbf{A}_1^T(\mathbf{U}_1\mathbf{V}_1^T + \mathbf{W}^L))\|_F &\leq \|\mathbf{A}_1^T(\mathbf{U}_1\mathbf{V}_1^T + \mathbf{W}^L)\|_F \\
&\leq \|\mathbf{A}_1^T\|\|(\mathbf{U}_1\mathbf{V}_1^T + \mathbf{W}^L)\|_F \\
&= \|(\mathbf{U}_1\mathbf{V}_1^T + \mathbf{W}^L)\|_F \\
&\leq \|\mathbf{Z}_{j_0}\|_F,
\end{aligned}$$

this then proves the claim.

**proof of (c)**. According to the setting of $\mathbf{A}_1$ in Eq. (3), we have

$$\|\mathbf{A}_1^T(\mathbf{U}_1\mathbf{V}_1^T + \mathbf{W}^L)\|_\infty \leq \|\mathbf{U}_1\mathbf{V}_1^T + \mathbf{W}^L\|_\infty. \tag{28}$$

Thus, we only need to show $\|\mathbf{U}_1\mathbf{V}_1^T + \mathbf{W}^L\|_\infty \leq \lambda/4$. We have $\mathbf{U}_1\mathbf{V}_1^T + \mathbf{W}^L = \mathbf{Z}_{j_0} + \mathbf{Y}_{j_0}$ and know that $\mathbf{Y}_{j_0}$ is supported on $\Omega^c$. Therefore, since $\|\mathbf{Z}_{j_0}\|_\infty \leq \|\mathbf{Z}_{j_0}\|_F \leq \lambda/8$, it suffices to show that $\|\mathbf{Y}_{j_0}\|_\infty \leq \frac{\lambda}{8}$. To this end, we deduce

$$\begin{aligned}
\|\mathbf{Y}_{j_0}\|_\infty &\leq q^{-1}\sum_j \|\mathcal{P}_{\Omega_j}\mathbf{Z}_{j-1}\|_\infty \\
&\leq q^{-1}\sum_j \|\mathbf{Z}_{j-1}\|_\infty \\
&\leq q^{-1}\sum_j \epsilon^j \|\mathbf{U}_1\mathbf{V}_1^T\|_\infty.
\end{aligned}$$

Since $\|\mathbf{U}_1\mathbf{V}_1^T\|_\infty \leq \sqrt{\frac{\mu r}{n_1 n_2}}$, this gives

$$\|\mathbf{Y}_{j_0}\|_\infty \leq \frac{C'\epsilon^2}{\sqrt{\mu r n_{(2)}^{-1} n_{(1)}(\log n_{(1)})^2}} \tag{29}$$

for some numerical constant $C'$ whenever $q$ obeys Eq. (22). By setting $\lambda = 1/\sqrt{n_{(1)}}$, $\|\mathbf{Y}_{j_0}\|_\infty \leq \lambda/8$ if

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

We have seen that (a) and (b) are satisfied if $\epsilon$ is sufficiently small and $j_0 \geq 2\log n_{(1)}$. For (c), we can take $\epsilon$ 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 $\rho_r$ in Eq. (17) in the manuscript is sufficiently small. Note that everything is consistent, since $C_0\epsilon^{-2}\mu r \log n_{(1)}/n_{(2)} < 1$. ■

#### 1.6.2 *Proof of Lemma 5*

**Proof:** Following the proof of Lemma 2.9 in [2], we have

(a) $\|\mathbf{W}^S\|_F < 1/4$,
(b) $\|\mathbf{W}^S\|_\infty < \lambda/4$.

According to the setting of $\mathbf{A}_1$, we further have

$$\begin{aligned}
\|\mathcal{P}_{\Omega^\perp}(\mathbf{A}_1^T\mathbf{W}^S)\|_\infty &\leq \|\mathbf{A}_1^T\mathbf{W}^S\|_\infty \\
&\leq \|\mathbf{W}^S\|_\infty < \lambda/4.
\end{aligned} \tag{30}$$

Thus, this lemma can be established. ■

Using the similar techniques, we can construct the other two dual certificates $\mathbf{W}_2, \mathbf{W}_3$ to make the Eq. (11) hold true. Therefore, Theorem 1 presented in the manuscript can be established.

### 1.7 Proofs of Some Auxiliary Lemmas

#### 1.7.1 *Proof of Lemma 2*

**Proof:** It is easy to see that for any $\mathbf{H} \neq \mathbf{0}$, $(\mathbf{X}_0 + \mathbf{H}, \mathbf{S}_0 - \mathbf{H})$ is also a feasible solution. We show that its objective is larger than that at $(\mathbf{X}_0, \mathbf{S}_0)$, hence proving that $(\mathbf{X}_0, \mathbf{S}_0)$ is the unique solution. To do this, let $\mathbf{U}_i\mathbf{V}_i^T + \mathbf{W}_i^0$ be an arbitrary sub-gradient of the nuclear norm at $\mathbf{G}_i, (i = 1, 2, 3)$, and $\mathrm{sgn}(\mathbf{S}_0) + \mathbf{F}_0$ be an arbitrary sub-gradient of the $\ell_1$ norm at $\mathbf{S}_0$. Then we have

$$\begin{aligned}
&\sum_{i=1}^2 \|\mathbf{A}_i\mathbf{X}_0 + \mathbf{A}_i\mathbf{H}\|_* + \|\mathbf{X}_0\mathbf{B}_1 + \mathbf{H}\mathbf{B}_1\|_* + 3\lambda\|\mathbf{S} - \mathbf{H}\|_1 \\
&\geqslant \sum_{i=1}^2 \|\mathbf{A}_i\mathbf{X}_0\|_* + \|\mathbf{X}_0\mathbf{B}_1\|_* + 3\lambda\|\mathbf{S}_0\|_1 \\
&+ \sum_{i=1}^2 \left\langle \mathbf{U}_i\mathbf{V}_i^T + \mathbf{W}_i^0, \mathbf{A}_i\mathbf{H}\right\rangle + \left\langle \mathbf{U}_3\mathbf{V}_3^T + \mathbf{W}_3^0, \mathbf{H}\mathbf{B}_1\right\rangle \\
&- 3\lambda\left\langle \mathrm{sgn}(\mathbf{S}) + \mathbf{F}_0, \mathbf{H}\right\rangle.
\end{aligned}$$

Now pick $\{\mathbf{W}_i^0\}_{i=1}^3$ such that $\langle \mathbf{W}_i^0, \mathbf{A}_i\mathbf{H}\rangle = \|\mathcal{P}_{T_i^\perp}(\mathbf{A}_i\mathbf{H})\|_*, (i=1,2)$, $\langle \mathbf{W}_3^0, \mathbf{HB}_1\rangle = \|\mathcal{P}_{T_3^\perp}(\mathbf{HB}_1)\|_*$ and $\mathbf{F}_0$ such that $\langle \mathbf{F}_0, \mathbf{H}\rangle = -\|\mathcal{P}_{\Omega^\perp}\mathbf{H}\|_1$. We further have

$$\begin{aligned}
&\sum_{i=1}^{2}\|\mathbf{A}_i\mathbf{X}_0+\mathbf{A}_i\mathbf{H}\|_*+\|\mathbf{X}_0\mathbf{B}_1+\mathbf{HB}_1\|_*+3\lambda\|\mathbf{S}_0-\mathbf{H}\|_1\\
&\geqslant \sum_{i=1}^{2}\|\mathbf{A}_i\mathbf{X}_0\|_*+\|\mathbf{X}_0\mathbf{B}_1\|_*+3\lambda\|\mathbf{S}_0\|_1\\
&+\sum_{i=1}^{2}\|\mathcal{P}_{T_i^\perp}(\mathbf{A}_i\mathbf{H})\|_*+\|\mathcal{P}_{T_3^\perp}(\mathbf{HB}_1)\|_*+3\|\mathcal{P}_{\Omega^\perp}\mathbf{H}\|_1\\
&+\sum_{i=1}^{2}\left\langle \mathbf{U}_i\mathbf{V}_i^T, \mathbf{A}_i\mathbf{H}\right\rangle+\left\langle \mathbf{U}_3\mathbf{V}_3^T, \mathbf{HB}_1\right\rangle-3\lambda\left\langle \mathrm{sgn}(\mathbf{S}_0), \mathbf{H}\right\rangle.
\end{aligned}$$

Considering

$$\begin{aligned}
&\sum_{i=1}^{2}\left\langle \mathbf{U}_i\mathbf{V}_i^T, \mathbf{A}_i\mathbf{H}\right\rangle\\
&+\left\langle \mathbf{U}_3\mathbf{V}_3^T, \mathbf{HB}_1\right\rangle-3\lambda\left\langle \mathrm{sgn}(\mathbf{S}_0), \mathbf{H}\right\rangle \qquad (31)\\
&=\left\langle \sum_{i=1}^{2}\mathbf{A}_i^T\mathbf{U}_i\mathbf{V}_i^T+\mathbf{U}_3\mathbf{V}_3^T\mathbf{B}_1^T-3\lambda\mathrm{sgn}(\mathbf{S}_0), \mathbf{H}\right\rangle
\end{aligned}$$

and by assumption (11),

$$\begin{aligned}
&\left|\left\langle \sum_{i=1}^{2}\mathbf{A}_i^T\mathbf{U}_i\mathbf{V}_i^T+\mathbf{U}_3\mathbf{V}_3^T\mathbf{B}_1^T-3\lambda\mathrm{sgn}(\mathbf{S}_0), \mathbf{H}\right\rangle\right|\\
&\leq \sum_{i=1}^{2}|\langle \mathbf{W}_i, \mathbf{A}_i\mathbf{H}\rangle|+|\langle \mathbf{W}_3, \mathbf{HB}_1\rangle|+3\lambda|\langle \mathbf{F}, \mathbf{H}\rangle|\\
&\leq \beta\left(\sum_{i=1}^{2}\|\mathcal{P}_{T_i^\perp}(\mathbf{A}_i\mathbf{H})\|_*+\|\mathcal{P}_{T_3^\perp}(\mathbf{HB}_1)\|_*+3\lambda\|\mathcal{P}_{\Omega^\perp}\mathbf{H}\|_1\right)
\end{aligned}$$

for $\beta=\max\left(\{\|\mathbf{W}_i\|\}_{i=1}^3, \|\mathbf{F}\|_\infty\right)<1$. Thus,

$$\begin{aligned}
&\sum_{i=1}^{2}\|\mathbf{A}_i\mathbf{X}_0+\mathbf{A}_i\mathbf{H}\|_*+\|\mathbf{X}_0\mathbf{B}_1+\mathbf{HB}_1\|_*+3\lambda\|\mathbf{S}_0-\mathbf{H}\|_1\\
&\geq \sum_{i=1}^{2}\|\mathbf{A}_i\mathbf{X}_0\|_*+\|\mathbf{X}_0\mathbf{B}_1\|_*+3\lambda\|\mathbf{S}_0\|_1+(1-\beta)\\
&\times\left(\sum_{i=1}^{2}\|\mathcal{P}_{T_i^\perp}(\mathbf{A}_i\mathbf{H})\|_*+\|\mathcal{P}_{T_3^\perp}(\mathbf{HB}_1)\|_*+3\lambda\|\mathcal{P}_{\Omega^\perp}\mathbf{H}\|_1\right).
\end{aligned}$$

Note that $\Omega\bigcap T_i=\{0\}$, we have $\sum_{i=1}^{2}\|\mathcal{P}_{T_i^\perp}(\mathbf{A}_i\mathbf{H})\|_*+\|\mathcal{P}_{T_3^\perp}(\mathbf{HB}_1)\|_*+3\lambda\|\mathcal{P}_{\Omega^\perp}\mathbf{H}\|_1>0$ unless $\mathbf{H}=0$. ■

### 1.7.2 *Proof of Lemma 3*

**Proof:** Following the proof ideas of Lemma 2, we first get

$$\begin{aligned}
&\sum_{i=1}^{2}\|\mathbf{A}_i\mathbf{X}_0+\mathbf{A}_i\mathbf{H}\|_*+\|\mathbf{X}_0\mathbf{B}_1+\mathbf{HB}_1\|_*+3\lambda\|\mathbf{S}_0-\mathbf{H}\|_1\\
&\geq \sum_{i=1}^{2}\|\mathbf{A}_i\mathbf{X}_0\|_*+\|\mathbf{X}_0\mathbf{B}_1\|_*+3\lambda\|\mathbf{S}_0\|_1+\frac{1}{2}\\
&\left(\sum_{i=1}^{2}\|\mathcal{P}_{T_i^\perp}(\mathbf{A}_i\mathbf{H})\|_*+\|\mathcal{P}_{T_3^\perp}(\mathbf{HB}_1)\|_*+3\lambda\|\mathcal{P}_{\Omega^\perp}\mathbf{H}\|_1\right)\\
&-\lambda\left\langle \mathcal{P}_\Omega\mathbf{D}, \mathbf{A}_1\mathbf{H}+\mathbf{A}_2\mathbf{H}+\mathbf{HB}_1\right\rangle\\
&\geq \sum_{i=1}^{2}\|\mathbf{A}_i\mathbf{X}_0\|_*+\|\mathbf{X}_0\mathbf{B}_1\|_*+3\lambda\|\mathbf{S}_0\|_1+\frac{1}{2}\\
&\left(\sum_{i=1}^{2}\|\mathcal{P}_{T_i^\perp}(\mathbf{A}_i\mathbf{H})\|_*+\|\mathcal{P}_{T_3^\perp}(\mathbf{HB}_1)\|_*+3\lambda\|\mathcal{P}_{\Omega^\perp}\mathbf{H}\|_1\right)\\
&-\frac{\lambda}{4}\left(\|\mathcal{P}_\Omega(\mathbf{A}_1\mathbf{H})\|_F+\|\mathcal{P}_\Omega(\mathbf{A}_2\mathbf{H})\|_F+\|\mathcal{P}_\Omega(\mathbf{HB}_1)\|_F\right).
\end{aligned}$$

Due to the fact that

$$\begin{aligned}
\|\mathcal{P}_\Omega(\mathbf{A}_i\mathbf{H})\|_F&\leq\|\mathcal{P}_\Omega\left(\mathcal{P}_{T_i}+\mathcal{P}_{T_i^\perp}\right)(\mathbf{A}_i\mathbf{H})\|_F\\
&\leq\|\mathcal{P}_\Omega\mathcal{P}_{T_i}(\mathbf{A}_i\mathbf{H})\|_F+\|\mathcal{P}_\Omega\mathcal{P}_{T_i^\perp}(\mathbf{A}_i\mathbf{H})\|_F\\
&\leq\frac{1}{2}\|\mathbf{A}_i\mathbf{H}\|_F+\|\mathcal{P}_{T_i^\perp}(\mathbf{A}_i\mathbf{H})\|_F\\
&\leq\frac{1}{2}\|\mathcal{P}_\Omega(\mathbf{A}_i\mathbf{H})\|_F+\frac{1}{2}\|\mathcal{P}_{\Omega^\perp}(\mathbf{A}_i\mathbf{H})\|_F\\
&+\|\mathcal{P}_{T_i^\perp}(\mathbf{A}_i\mathbf{H})\|_F,(i=1,2),
\end{aligned}$$

$$\begin{aligned}
\|\mathcal{P}_\Omega(\mathbf{HB}_1)\|_F&\leq\|\mathcal{P}_\Omega\left(\mathcal{P}_{T_3}+\mathcal{P}_{T_3^\perp}\right)(\mathbf{HB}_1)\|_F\\
&\leq\|\mathcal{P}_\Omega\mathcal{P}_{T_3}(\mathbf{HB}_1)\|_F+\|\mathcal{P}_\Omega\mathcal{P}_{T_3^\perp}(\mathbf{HB}_1)\|_F\\
&\leq\frac{1}{2}\|\mathbf{HB}_1\|_F+\|\mathcal{P}_{T_3^\perp}(\mathbf{HB}_1)\|_F\\
&\leq\frac{1}{2}\|\mathcal{P}_\Omega(\mathbf{HB}_1)\|_F+\frac{1}{2}\|\mathcal{P}_{\Omega^\perp}(\mathbf{HB}_1)\|_F\\
&+\|\mathcal{P}_{T_3^\perp}(\mathbf{HB}_1)\|_F,
\end{aligned}$$

we can get

$$\begin{aligned}
&\|\mathcal{P}_\Omega(\mathbf{A}_i\mathbf{H})\|_F\leq\|\mathcal{P}_{\Omega^\perp}(\mathbf{A}_i\mathbf{H})\|_F+2\|\mathcal{P}_{T_i^\perp}(\mathbf{A}_i\mathbf{H})\|_F,(i=1,2),\\
&\|\mathcal{P}_\Omega(\mathbf{HB}_1)\|_F\leq\|\mathcal{P}_{\Omega^\perp}(\mathbf{HB}_1)\|_F+2\|\mathcal{P}_{T_3^\perp}(\mathbf{HB}_1)\|_F.
\end{aligned}$$

In conclusion,

$$\begin{aligned}
&\sum_{i=1}^{2}\|\mathbf{A}_i\mathbf{X}_0+\mathbf{A}_i\mathbf{H}\|_*+\|\mathbf{X}_0\mathbf{B}_1+\mathbf{HB}_1\|_*+3\lambda\|\mathbf{S}_0-\mathbf{H}\|_1\\
&\geq\sum_{i=1}^{2}\|\mathbf{A}_i\mathbf{X}_0\|_*+\|\mathbf{X}_0\mathbf{B}_1\|_*+3\lambda\|\mathbf{S}_0\|_1\\
&+\left(\sum_{i=1}^{2}\|\mathcal{P}_{T_i^\perp}(\mathbf{A}_i\mathbf{H})\|_*+\|\mathcal{P}_{T_3^\perp}(\mathbf{HB}_1)\|_*\right)\\
&+\frac{\lambda}{4}\left(\sum_{i=1}^{2}\|\mathcal{P}_{\Omega^\perp}(\mathbf{A}_i\mathbf{H})\|_1+\|\mathcal{P}_{\Omega^\perp}(\mathbf{HB}_1)\|_1\right),
\end{aligned}$$

and the last two terms are strictly positive when $\mathbf{H}\neq\mathbf{0}$. ■

TABLE 1
The quantitative comparison of the 3DCTV-G, 3DCTV-RPCA and RPCA model on DC mall dataset under different levels of Gaussian noise. The best and second results on each line are highlighted in bold italics and underline, respectively. The value in the table is the average result of ten independent trails.

| Noise | Metric | Noisy | 3DCTV-G | 3DCTV-RPCA | RPCA |
|---|---|---|---|---|---|
| $\sigma = 0.10$ | psnr | 20.00 | **34.84** | 34.56 | 32.06 |
| | ssim | 0.5147 | **0.9707** | 0.9629 | 0.9586 |
| | ergas | 375.89 | **67.55** | 68.71 | 90.69 |
| $\sigma = 0.20$ | psnr | 13.98 | **31.05** | 30.34 | 27.97 |
| | ssim | 0.2563 | **0.9320** | 0.9060 | 0.9001 |
| | ergas | 751.51 | **103.78** | 110.62 | 144.80 |
| $\sigma = 0.30$ | psnr | 10.46 | **29.02** | 27.93 | 25.59 |
| | ssim | 0.1437 | **0.8933** | 0.8457 | 0.8398 |
| | ergas | 1127.1 | **130.60** | 145.22 | 190.93 |
| $\sigma = 0.40$ | psnr | 7.95 | **27.72** | 26.22 | 23.97 |
| | ssim | 0.887 | **0.8578** | 0.7852 | 0.7838 |
| | ergas | 1504.5 | **151.54** | 176.34 | 230.71 |

## 2 Proof of Theorem 2

Now we shall give the detailed proof of Theorem 2 presented in the main text. To this end, we need the following two lemmas.

**Lemma 9:** The sequence of dual variables $\{\mathbf{\Gamma}_i^k\}_{i=1}^4$ in Algorithm 1 are bounded.

**Proof:** According to the optimality principle, we have

$$\begin{aligned}&\mathbf{0} \in \partial(\|\mathbf{G}_i^{k+1}\|_*) - \mathbf{\Gamma}_i^k - \mu_k(\nabla_i(\mathbf{X}^{k+1}) - \mathbf{G}_i^{k+1}),\\ &\mathbf{0} \in \partial(3\lambda\|\mathbf{S}^{k+1}\|_1) - \mathbf{\Gamma}_4^k - \mu_k(\mathbf{M} - \mathbf{X}^{k+1} - \mathbf{S}^{k+1}).\end{aligned} \tag{32}$$

Combining this with the update criterion of the $\{\mathbf{\Gamma}_i^k\}_{i=1}^4$ in Algorithm 1, we have

$$\begin{aligned}&\mathbf{\Gamma}_i^{k+1} \in \partial(\|\mathbf{G}_i^{k+1}\|_*), i = 1,2,3,\\ &\mathbf{\Gamma}_4^{k+1} \in \partial(3\lambda\|\mathbf{S}^{k+1}\|_1),\\ &\mathbf{\Gamma}_4^k = \sum_{i=1}^3 \nabla_i^{\mathrm{T}}(\mathbf{\Gamma}_i^k).\end{aligned} \tag{33}$$

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 [3], [4]. Thus, using Theorem 4 in [3], we get that $\{\mathbf{\Gamma}_i^k\}_{i=1}^4$ are bounded. ■

**Lemma 10:** The accumulation point $(\{\mathbf{G}_i^k\}_{i=1}^3, \mathbf{X}^k, \mathbf{S}^k)$ generated by Algorithm 1 is a feasible solution of 3DCTV-RPCA model (1).

**Proof:** Based on the general ADMM principle, we have

$$\begin{cases}\|\mathbf{\Gamma}_i^{k+1} - \mathbf{\Gamma}_i^k\|_F = \mu^k\|\nabla_i\mathbf{X}^{k+1} - \mathbf{G}_i^{k+1}\|_F, n = 1,2,3,\\ \|\mathbf{\Gamma}_4^{k+1} - \mathbf{\Gamma}_4^k\|_F = \mu^k\|\mathbf{M} - \mathbf{X}^{k+1} - \mathbf{S}^{k+1}\|_F.\end{cases} \tag{34}$$

Since $\{\mu^k\}$ is an increasing sequence and $\lim_{k\to+\infty}\mu^k = +\infty$, and according to Lemma 9, we have

$$\begin{cases}\lim\limits_{k\to+\infty}\|\nabla_i\mathbf{X}^{k+1} - \mathbf{G}_i^{k+1}\|_F = 0, n = 1,2,3,\\ \lim\limits_{k\to+\infty}\|\mathbf{M} - \mathbf{X}^{k+1} - \mathbf{S}^{k+1}\|_F = 0.\end{cases} \tag{35}$$

This completes the proof. ■

With the above lemmas, we next give the proof of Theorem 2 in the main text.

**Proof:** Suppose $(\mathbf{X}^*, \mathbf{S}^*, \{\mathbf{G}_i^*\}_{i=1}^3)$ is an optimal solution of the 3DCTV-RPCA model (1), and $(\{\mathbf{\Gamma}_i^*\}_{i=1}^4)$ is the optimal solution of its dual model, it thus get that $(\mathbf{X}^*, \mathbf{S}^*, \{\mathbf{G}_i^*\}_{i=1}^3, \{\mathbf{\Gamma}_i^*\}_{i=1}^4)$ forms the saddle point of the Lagrangian function (i.e., Eq. (14) in the main text).

Recall Eq. (33), we have

$$\begin{aligned}\sum_{i=1}^3\|\mathbf{G}_i^k\|_* + 3\lambda\|\mathbf{S}^k\|_1 &\le \sum_{i=1}^3(\|\mathbf{G}_i^*\|_* - \langle\mathbf{\Gamma}_i^k, \mathbf{G}_i^* - \mathbf{G}_i^k\rangle)\\ &+ 3\lambda\|\mathbf{S}^*\|_1 - \langle\mathbf{\Gamma}_4^k, \mathbf{S}^* - \mathbf{S}^k\rangle\\ &= \sum_{i=1}^3\|\mathbf{G}_i^*\|_* + 3\lambda\|\mathbf{S}^*\|_1\\ &+ \sum_{i=1}^3\langle\mathbf{\Gamma}_i^k, \nabla_i(\mathbf{X}^k - \mathbf{X}^*)\rangle + \langle\mathbf{\Gamma}_4^k, \mathbf{S}^k - \mathbf{S}^*\rangle\\ &= \sum_{i=1}^3\|\mathbf{G}_i^*\|_* + 3\lambda\|\mathbf{S}^*\|_1 + \langle\mathbf{\Gamma}_4^k, \mathbf{S}^k + \mathbf{X}^k - \mathbf{M}\rangle\end{aligned} \tag{36}$$

Combining Eq. (36) with Lemma 10, we further have

$$\lim_{k\to+\infty}\sum_{i=1}^3\|\mathbf{G}_i^k\|_* + 3\lambda\|\mathbf{S}^k\|_1 = \sum_{i=1}^3\|\mathbf{G}_i^*\|_* + 3\lambda\|\mathbf{S}^*\|_1. \tag{37}$$

This completes the proof. ■

## 3 The Extended Models Based on 3DCTV Regularizer

### 3.1 The 3DCTV-G model

The first extended model is:

$$\begin{aligned}\min_{\mathbf{X},\mathbf{E}}\ &\sum_{i=1}^3\|\mathbf{G}_i\|_* + \lambda\|\mathbf{S}\|_2\\ \text{s.t.}\ &\mathbf{M} = \mathbf{X} + \mathbf{S},\\ &\mathbf{G}_i = \nabla_i(\mathbf{X}), i = 1,2,3,\end{aligned} \tag{38}$$

where $\nabla_i(\cdot), i = 1,2,3$ is a differential operator. This model (38) named **3DCTV-G** model can be used to separate pure Gaussian noise. We can use Algorithm 1 in the manuscript to solve this model. And according to the setting of [5], the trade-off coefficient $\lambda$ can be set as $\lambda = 1/(6(\sqrt{n_1} + \sqrt{n_2})\sigma)$, where $\sigma$ is estimated variance of the Gaussian noise.

Tables 1 and 2 list the specific performance of 3DCTV-G model under Gaussian noise on different datasets. In these two tables, it can be observed that the 3DCTV-G model gains a 1dB higher value of PSNR than the 3DCTV-RPCA model on average. The restoration images are provided in Figs. 1-3 for visualize comparison. From these figures, we can see that 3DCTV-G model can better remove the noise and preserve the local texture details.

### 3.2 The 3DWCTV-RPCA model

To remove the stripe noise, we propose another extended model named 3DWCTV-RPCA model, which is of the form as

$$\begin{aligned}\min_{\mathbf{X},\mathbf{E}}\ &2\|\mathbf{G}_1\|_* + \sum_{i=2}^3\|\mathbf{G}_i\|_* + 3\lambda\|\mathbf{S}\|_1\\ \text{s.t.}\ &\mathbf{M} = \mathbf{X} + \mathbf{S},\\ &\mathbf{G}_i = \nabla_i(\mathbf{X}), i = 1,2,3,\end{aligned} \tag{39}$$

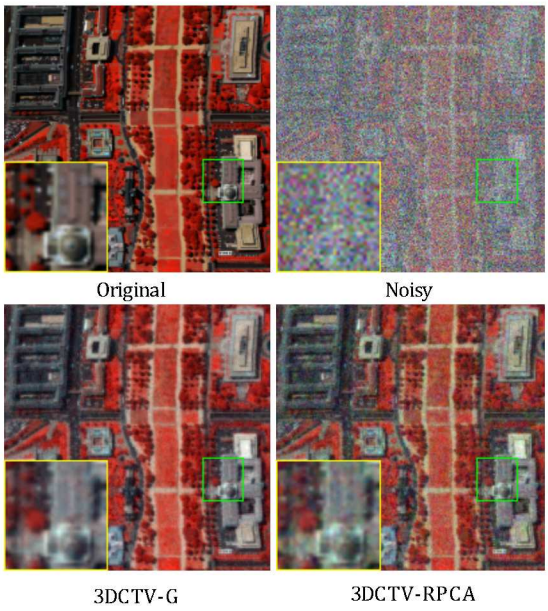



Fig. 1. Recovered images of all competing methods with bands 58-27-17 as R-G-B. (a) The simulated DC mall image. (b) The noise image with Gaussian noise variance is 0.4. (c-n)Restoration results obtained by 13 comparison methods with a demarcated zoomed in 3 times.

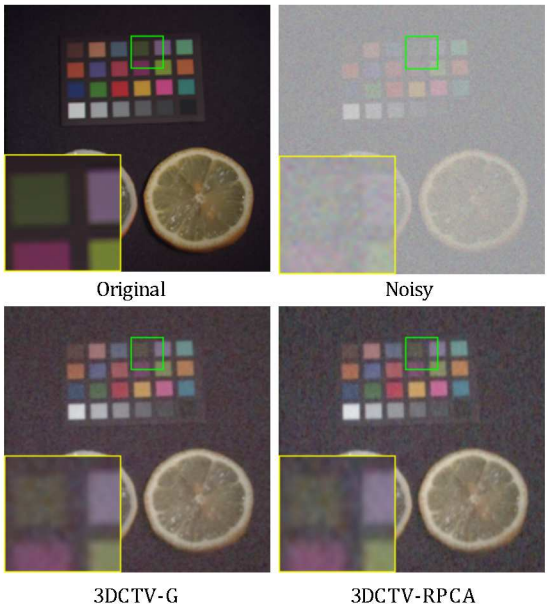



Fig. 2. Recovered images of all competing methods with bands 58-27-17 as R-G-B. (a) An image selected from one data set from the CAVE datasets. (b) The noise image with Gaussian noise variance is 0.4. (c-n) Restoration results obtained by 3 comparison methods with a demarcated zoomed in three times.

where the trade-off parameter $\lambda$ is also set as $\lambda = 1/\sqrt{n_{(1)}}$.

In Eq. (39), the weight on the gradient map $\mathbf{G}_1$ is set twice of that imposed on $\mathbf{G}_2$ and $\mathbf{G}_3$, which makes the model have a better removal effect for horizontal stripe noise. It is easy to observe in Fig. 4 that the 3DWCTV-RPCA model can effectively remove horizontal stripes.

## 4 Hyper-parameter Robustness

As suggested by the theoretical assertion, $\lambda = 1/\sqrt{n_{(1)}}$ is a good choice for the proposed model. To better clarify the issue of hyper-parameter setting, we provide a series of

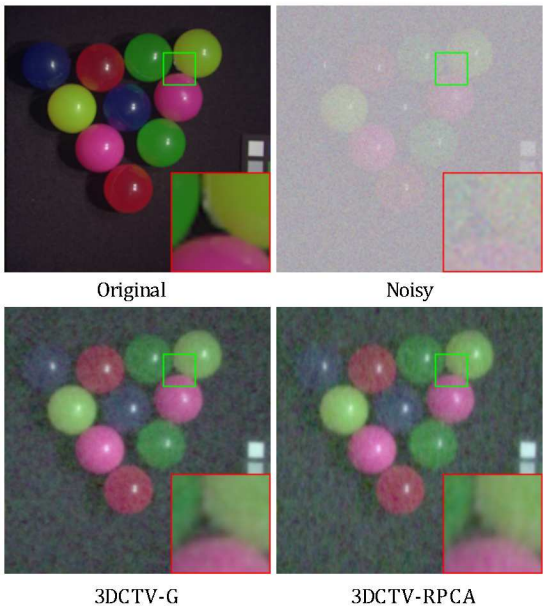



Fig. 3. Recovered images of all competing methods with bands 58-27-17 as R-G-B. (a) An image selected from one data set from the CAVE datasets. (b) The noise image with Gaussian noise variance is 0.4. (c-n) Restoration results obtained by 3 comparison methods with a demarcated zoomed in three times.

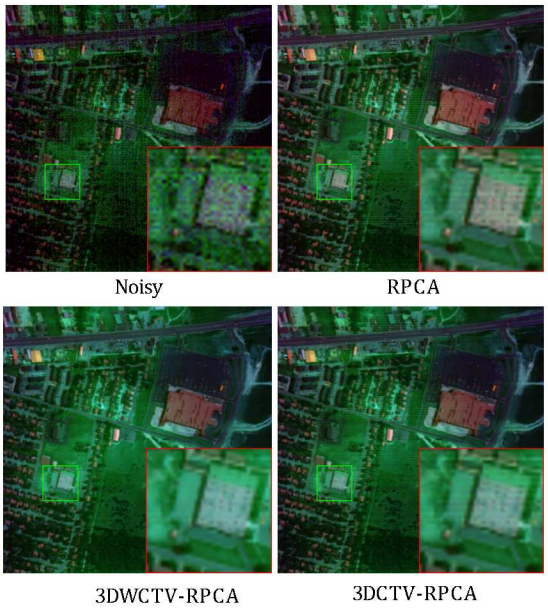



Fig. 4. Recovered images of all competing methods with bands 6-104-36 as R-G-B. (a) The original urban part image. (b-d) Restoration results obtained by 3 comparison methods with a demarcated zoomed in three times.

experiments for the selection of parameter $\lambda$. Specifically, we plot the relative error with varying $\lambda$ under different cases in Fig. 5 using the same settings of our simulated experiments in the main text. From the figure, one can easily see that the optimal performance of all cases are approximately obtained on $\lambda$ with order of $O(1/\sqrt{n_{(1)}})$. It can also be observed that although the best performance is not always exactly attained at $\lambda = 1/\sqrt{n_{(1)}}$, it still achieves comparable performance with the optimal one. Combining this observation with our theoretical assertions (e.g., Theorem 1), we still prefer to suggest users to directly adopt $\lambda = 1/\sqrt{n_{(1)}}$ in practical applications for more convenience.

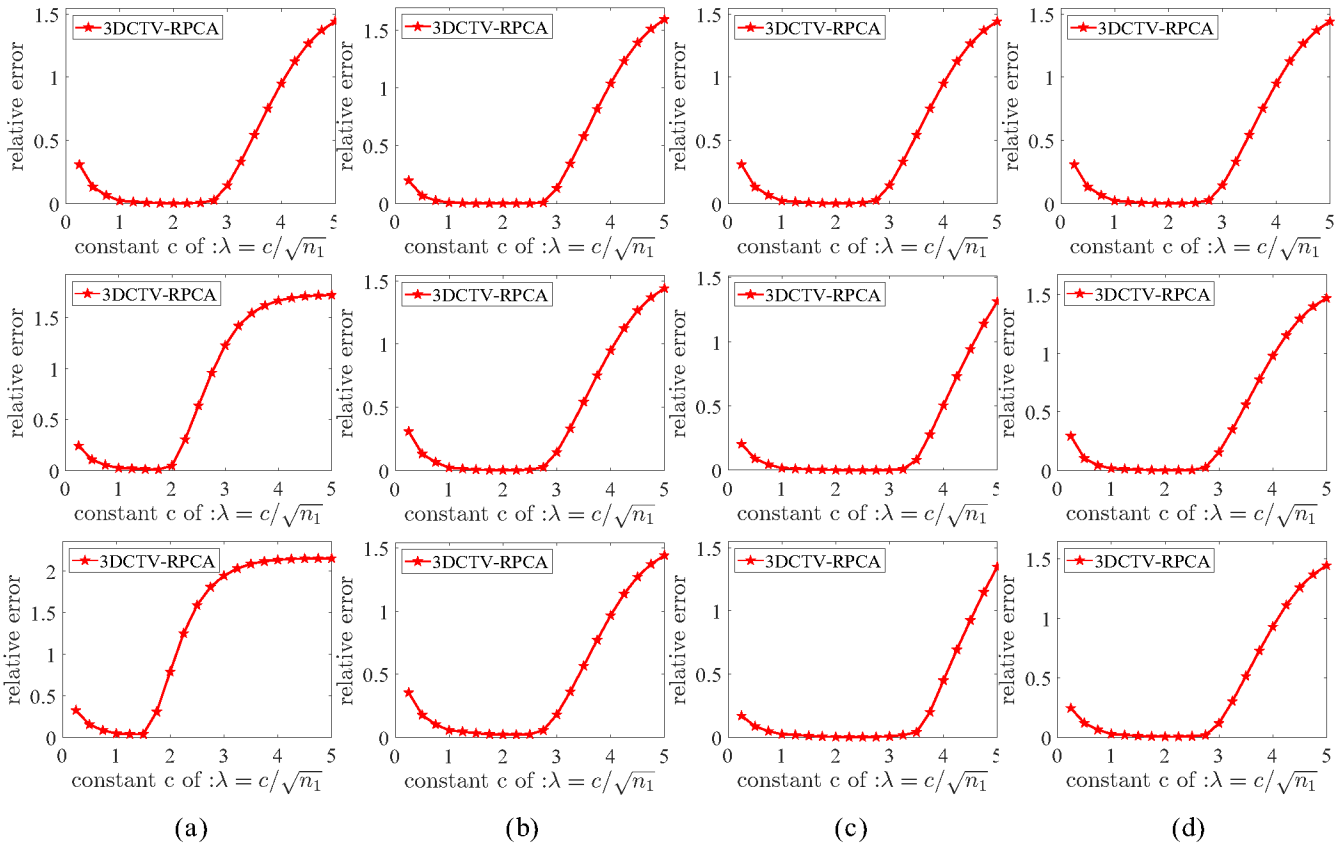



Fig. 5. The performance in terms of relative error with varying $\lambda$, where $\lambda = c/\sqrt{n_1}$. (a) $\mathbf{X}$ with fixed size $400 \times 200$, composed by low-rank term $\mathbf{X}_0$ with rank 20 and sparse term $\mathbf{S}_0$ with sparsity 0.1, 0.2 and 0.3 from upper to lower; (b) $\mathbf{X}$ with fixed size $400 \times 200$, composed by $\mathbf{S}_0$ with sparsity 0.1 and $\mathbf{X}_0$ with rank $10, 20, 40$ from upper to lower; (c) $\mathbf{X}$ with sizes $400 \times 200$, $2500 \times 200$ and $10000 \times 200$, composed by $\mathbf{X}_0$ with rank 20 and $\mathbf{S}_0$ with sparsity 0.1 from upper to lower; (d) $\mathbf{X}$ with fixed size $400 \times 200$, composed by $\mathbf{X}_0$ with rank 20, $\mathbf{S}_0$ with sparsity 0.1 and Gaussian noise with standard variances $0, 0.05, 0.1$ from upper to lower.

TABLE 2
The quantitative comparison of all competing methods under different levels of noises on 32 scenes in the CAVE datasets. Each value is the mean of all data performance. The best and second results on each line are highlighted in bold italics and underline, respectively. The value in the table is the average of ten times.

| Noise | Metric | Noisy | 3DCTV-G | 3DCTV-RPCA | RPCA |
|---|---|---|---|---|---|
| $\sigma$=0.10 | psnr | 20.00 | **34.18** | 33.47 | 25.59 |
| | ssim | 0.4180 | ***0.9434*** | 0.9006 | 0.8450 |
| | ergas | 520.58 | ***104.15*** | 111.16 | 282.86 |
| $\sigma$=0.20 | psnr | 13.98 | **31.30** | 30.27 | 22.44 |
| | ssim | 0.2033 | ***0.8895*** | 0.7803 | 0.7185 |
| | ergas | 1041.18 | ***144.80*** | 158.94 | 385.78 |
| $\sigma$=0.30 | psnr | 10.46 | **29.73** | 28.24 | 20.54 |
| | ssim | 0.1246 | ***0.8449*** | 0.6712 | 0.6248 |
| | ergas | 1561.74 | ***172.57*** | 199.46 | 471.91 |
| $\sigma$=0.40 | psnr | 7.96 | **28.69** | 26.69 | 19.25 |
| | ssim | 0.0847 | ***0.8079*** | 0.5778 | 0.5533 |
| | ergas | 2082.38 | ***193.94*** | 237.36 | 543.23 |

## 5 Experiments on Tensor RPCA

Because the gradient map obtained by imposing the difference operation to the original data can inherit the low rank of the original data, it is not hard to construct a tensor version of CTV to deal with higher-order data. Specifically, by using the tensor nuclear norm via t-SVD framework defined in [6], [7], we can readily generalize the proposed CTV regularizer from the matrix form to the tensor form (we shortly call it t-CTV for convenience), that is,

$$\|\mathcal{X}\|_{\text{t-CTV}} := \frac{1}{3}\sum_{i=1}^{3}\|\mathcal{G}_i\|_*, \tag{40}$$

where $\mathcal{G}_i (i = 1, 2, 3)$ represents the tensor form of the gradient map obtained by applying the difference operator along the $i-$th mode on original data $\mathcal{X}$, and $*$ is the tensor nuclear norm as defined in [6], [7].

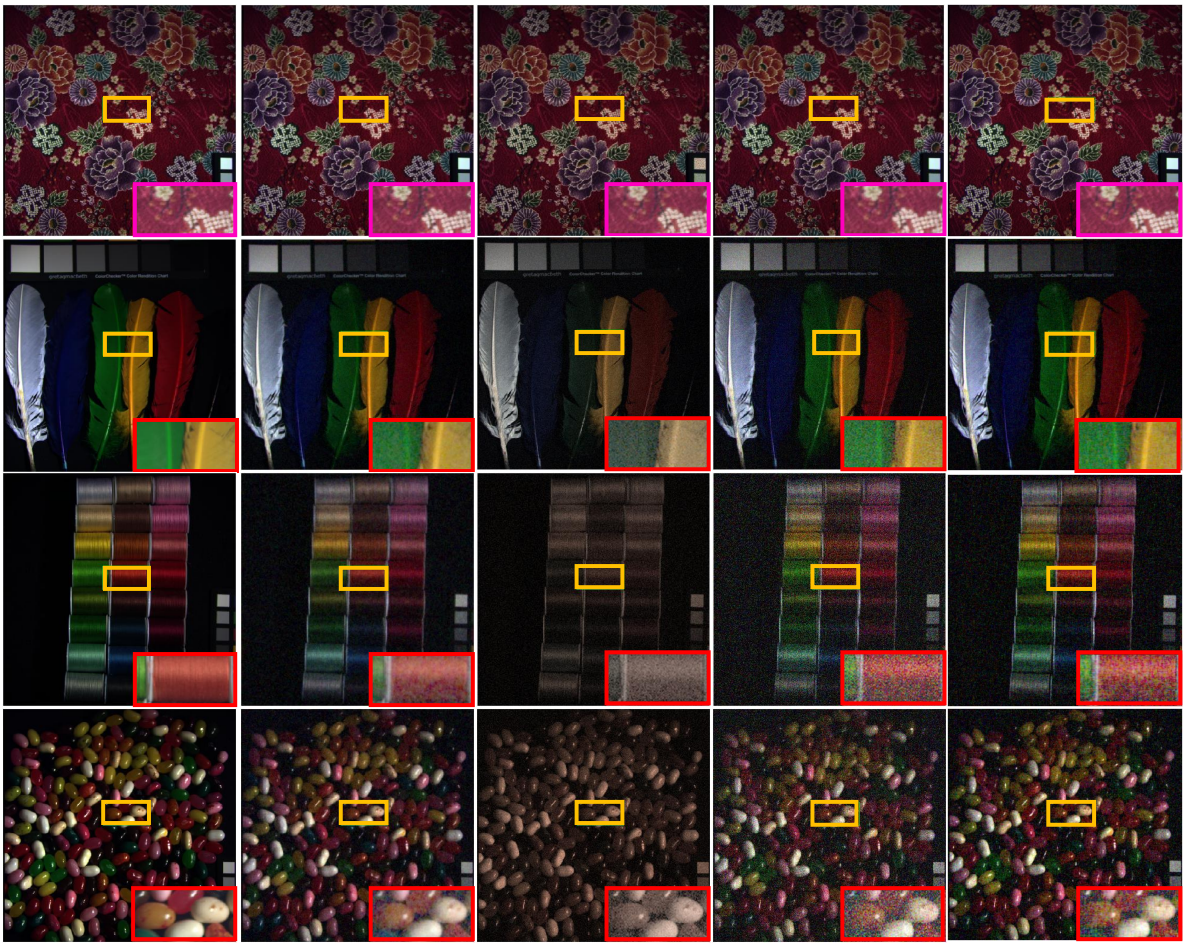
Therefore, by replacing the tensor nuclear norm with the above t-CTV regularizer in traditional tensor RPCA model, we can then get the following new model:

$$\|\mathcal{X}\|_{\text{t-CTV}} + \lambda\|\mathcal{E}\|_1, \ s.t. \ \mathcal{Y} = \mathcal{X} + \mathcal{E}. \tag{41}$$

To illustrate the performance of this new model, we have chosen the typical TRPCA model [6], [7] with solid theoretical guarantee, together with the aforementioned tensor formulation of CTV (t-CTV shortely) based TRPCA model, as the representative of the tensor RPCA model for performance comparison. As the experimental settings in Section 7.2 of the manuscript, we have used four datasets in the CAVE database in experiments, and the results are shown in Table 3. It can be easily seen that the performance of the CTV-RPCA and t-CTV are evidently higher than that of the RPCA and TRPCA models. This validates the superiority of this proposed new regularizer. Specifically, compared with t-CTV, the performance of CTV-RPCA is slightly worse than t-CTV in removing sparse noise, but is obviously better in removing mixed noise and Gaussian noise, which shows that the performance gain is mainly attained from the CTV formulation, but not essentially on its matrix or tensor forms in these cases. The advantage of our methods can also be observed from Fig. 6, which plots the recovered images of all competing methods. It can be seen that our method can

TABLE 3
Performance comparison of all competing methods under different noisy cases for the CAVE database. The best and second best results on each dataset are highlighted in bold italics and underline, respectively.

| Data / Noise | Metric Index | cloth | | | | feathers | | | | jelly beans | | | | thread spools | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | RPCA | TRPCA | t-CTV | CTV | RPCA | TRPCA | t-CTV | CTV | RPCA | TRPCA | t-CTV | CTV | RPCA | TRPCA | t-CTV | CTV |
| S=0.20 | MPSNR | 31.94 | 38.53 | 41.33 | **42.33** | 26.22 | 39.70 | **44.43** | 41.20 | 28.98 | 34.65 | **37.72** | 36.25 | 28.73 | 47.55 | **50.47** | 44.95 |
| | MSSIM | 0.980 | 0.988 | 0.993 | **0.997** | 0.936 | 0.991 | 0.996 | **0.997** | 0.963 | 0.987 | **0.994** | 0.993 | 0.954 | 0.998 | **0.990** | 0.998 |
| | ERGAS | 102.9 | 50.3 | 38.05 | **35.4** | 226.7 | 56.12 | **34.11** | 46.7 | 173.9 | 82.9 | **59.14** | 74.0 | 205.6 | 28.71 | **18.72** | 35.0 |
| | TIMES | **13.16** | 172.6 | 342.5 | 122.5 | **13.21** | 174.2 | 345.2 | 123.8 | **13.45** | 178.2 | 348.2 | 124.5 | **13.43** | 175.2 | 345.6 | 131.4 |
| S=0.40 | MPSNR | 24.64 | 30.55 | 36.42 | **37.36** | 21.21 | 36.31 | **41.98** | 37.97 | 23.20 | 31.96 | **36.00** | 34.07 | 24.24 | 43.42 | **48.05** | 41.67 |
| | MSSIM | 0.914 | 0.954 | 0.987 | **0.992** | 0.837 | 0.982 | **0.995** | 0.992 | 0.895 | 0.972 | **0.991** | 0.987 | 0.887 | 0.995 | **0.998** | 0.995 |
| | ERGAS | 210.3 | 106.4 | 55.24 | **54.4** | 396.0 | 75.5 | **41.76** | 63.9 | 314.4 | 111.1 | **70.88** | 92.8 | 337.7 | 42.9 | **23.85** | 51.7 |
| | TIMES | **12.16** | 175.6 | 380.8 | 121.5 | **12.21** | 174.2 | 385.8 | 120.8 | **11.45** | 168.2 | 387.6 | 120.5 | **11.43** | 175.2 | 385.6 | 123.4 |
| G=0.20 | MPSNR | 22.17 | 19.04 | 23.67 | **27.40** | 21.56 | 19.26 | 24.33 | **29.23** | 22.46 | 19.13 | 24.07 | **28.65** | 23.23 | 19.34 | 24.49 | **30.59** |
| | MSSIM | 0.817 | 0.648 | 0.745 | **0.878** | 0.666 | 0.382 | 0.478 | **0.759** | 0.779 | 0.587 | 0.676 | **0.859** | 0.726 | 0.373 | 0.462 | **0.772** |
| | ERGAS | 275.4 | 404.5 | 236.9 | **152.6** | 374.5 | 486.2 | 270.8 | **158.3** | 328.6 | 489.9 | 277.2 | **163.5** | 374.7 | 594.2 | 327.9 | **164.6** |
| | TIMES | **11.78** | 163.5 | 349.6 | 119.7 | **13.21** | 174.2 | 351.8 | 123.8 | **13.45** | 178.2 | 352.6 | 124.5 | **13.43** | 175.2 | 345.4 | 131.4 |
| G=0.40 | MPSNR | 18.46 | 13.24 | 18.34 | **24.26** | 18.21 | 13.34 | 18.63 | **25.81** | 19.09 | 13.29 | 19.23 | **25.44** | 20.13 | 13.38 | 19.28 | **26.88** |
| | MSSIM | 0.642 | 0.375 | 0.509 | **0.755** | 0.499 | 0.183 | 0.248 | **0.562** | 0.627 | 0.346 | 0.447 | **0.729** | 0.560 | 0.164 | 0.227 | **0.568** |
| | ERGAS | 420.1 | 790.0 | 438.7 | **219.2** | 547.5 | 961.6 | 522.6 | **231.4** | 479.0 | 961.7 | 526.4 | **235.5** | 534.1 | 1181 | 640.7 | **249.3** |
| | TIMES | **13.94** | 165.9 | 345.4 | 119.2 | **12.21** | 163.2 | 350.6 | 117.8 | **13.15** | 168.2 | 354.4 | 119.5 | **13.43** | 165.2 | 348.4 | 118.4 |
| S=0.10 S=0.10 | MPSNR | 25.21 | 23.59 | 27.83 | **30.12** | 23.87 | 24.21 | 29.13 | **31.88** | 24.89 | 23.83 | 28.54 | **30.93** | 25.47 | 24.44 | 29.53 | **33.31** |
| | MSSIM | 0.898 | 0.812 | 0.875 | **0.936** | 0.766 | 0.615 | 0.719 | **0.866** | 0.849 | 0.764 | 0.838 | **0.920** | 0.787 | 0.611 | 0.711 | **0.863** |
| | ERGAS | 196.1 | 238.6 | 146.1 | **111.8** | 288.9 | 274.4 | 155.6 | **117.3** | 253.6 | 284.1 | 165.3 | **126.2** | 290.9 | 329.6 | 183.2 | **120.4** |
| | TIMES | **11.82** | 162.3 | 369.8 | 121.1 | **11.38** | 160.8 | 391.0 | 124.1 | **12.34** | 171.1 | 390.8 | 122.6 | **13.78** | 172.8 | 380.1 | 125.6 |
| S=0.20 S=0.20 | MPSNR | 20.87 | 17.15 | 22.15 | **25.86** | 20.22 | 17.56 | 22.69 | **26.82** | 21.29 | 17.35 | 22.49 | **26.53** | 22.45 | 17.70 | 22.83 | **27.67** |
| | MSSIM | 0.751 | 0.544 | 0.679 | **0.833** | 0.578 | 0.298 | 0.409 | **0.646** | 0.708 | 0.488 | 0.612 | **0.776** | 0.599 | 0.281 | 0.387 | **0.613** |
| | ERGAS | 320.0 | 503.4 | 282.9 | **183.5** | 435.5 | 591.7 | 327.4 | **205.4** | 374.8 | 601.5 | 333.4 | **208.7** | 408.8 | 718.3 | 397.8 | **227.1** |
| | TIMES | **11.84** | 162.1 | 342.1 | 121.5 | **11.34** | 160.1 | 351.8 | 124.5 | **13.34** | 172.1 | 312.6 | 123.6 | **13.43** | 175.2 | 349.2 | 131.4 |



(a) Original (b) CTV-RPCA (c) RPCA (d) TRPCA (e) t-CTV

Fig. 6. From upper to lower: Restored images of the cloth data under sparse noise S=0.2, the feather data under mixture noise with S=0.1 and G=0.1, thread spools data under mixture noise with S=0.2 and G=0.2, and jelly beans data under mixture noise with S=0.2 and G=0.2 by different competing methods.

TABLE 4
Hyperspectral image inpainting quantitative results of all competing methods under different sampling rates. Each value is the mean of ten experiments. The best result on each line is highlighted in bold. ↑ / ↓ indicates that better results are with larger/smaller metric values.

| Datasets | SR | 1% | | | 3% | | | 5% | | | 10% | | | 20% | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Methods | NN | TNN | Ours | NN | TNN | Ours | NN | TNN | Ours | NN | TNN | Ours | NN | TNN | Ours |
| DC mall | PSNR↑ | 13.08 | 20.62 | **22.76** | 17.59 | 25.42 | **29.12** | 24.67 | 27.65 | **32.91** | 38.13 | 31.25 | **44.67** | 47.46 | 36.26 | **51.34** |
| | SSIM ↑ | 0.088 | 0.503 | **0.602** | 0.494 | 0.784 | **0.883** | 0.804 | 0.859 | **0.951** | 0.981 | 0.930 | **0.996** | 0.998 | 0.973 | **0.999** |
| | FSIM ↑ | 0.555 | 0.765 | **0.785** | 0.781 | 0.888 | **0.929** | 0.905 | 0.923 | **0.969** | 0.990 | 0.959 | **0.997** | 0.998 | 0.983 | **0.999** |
| | ERGAS ↓ | 826.7 | 341.4 | **283.1** | 525.7 | 199.6 | **141.7** | 258.6 | 156.6 | **90.45** | 54.19 | 106.2 | **24.37** | 17.35 | 63.21 | **11.29** |
| | MSAM ↓ | 1.011 | 0.290 | **0.230** | 0.420 | 0.208 | **0.143** | 0.216 | 0.175 | **0.102** | 0.045 | 0.132 | **0.029** | 0.022 | 0.087 | **0.016** |
| Indian Pines | PSNR↑ | 6.71 | 18.63 | **24.75** | 11.96 | 25.15 | **31.78** | 21.40 | 27.59 | **37.19** | 35.75 | 31.49 | **48.16** | 54.49 | 36.76 | **66.68** |
| | SSIM ↑ | 0.051 | 0.408 | **0.814** | 0.210 | 0.697 | **0.952** | 0.523 | 0.783 | **0.981** | 0.941 | 0.878 | **0.997** | 0.998 | 0.949 | **1.000** |
| | FSIM ↑ | 0.397 | 0.533 | **0.757** | 0.413 | 0.715 | **0.912** | 0.641 | 0.782 | **0.966** | 0.940 | 0.866 | **0.996** | 0.997 | 0.939 | **1.000** |
| | ERGAS ↓ | 1075 | 303.2 | **145.7** | 590.4 | 150.1 | **67.16** | 202.8 | 116.8 | **38.40** | 44.18 | 79.79 | **12.60** | 7.31 | 47.73 | **1.84** |
| | MSAM ↓ | 0.890 | 0.218 | **0.085** | 0.303 | 0.107 | **0.036** | 0.122 | 0.081 | **0.020** | 0.024 | 0.054 | **0.006** | 0.003 | 0.031 | **0.001** |

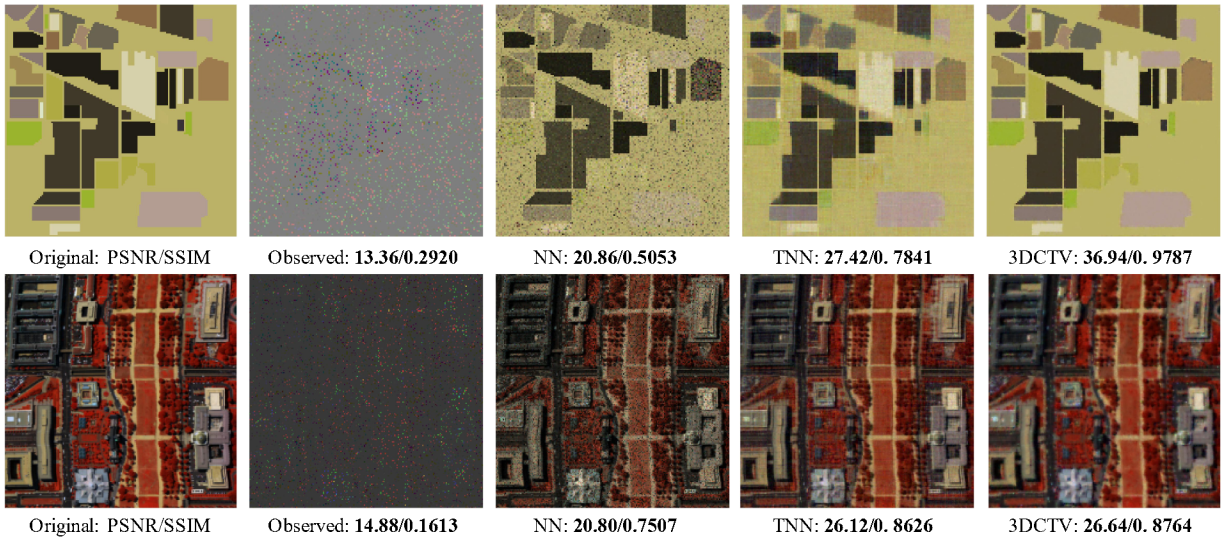



Fig. 7. Recovered images of all competing methods with bands 49-29-7 as R-G-B under Indian Pines and DC mall datasets.

better remove noise and preserve the texture details of the noisy image, which is consistent with the analysis of Table 3.

## 6 Image Inpainting Experiments via 3DCTV

Similar to the nuclear norm, which inclines to be suitable for dealing with general low-rank data, 3DCTV is also suitable for dealing with general joint low-rank and locally smooth data, regardless of the specific application. Since this paper studies the RPCA problem, it does not involve other applications of 3DCTV. Here, we have selected two commonly used hyperspectral datasets, i.e., DC mall and Indian Pines, to illustrate the effectiveness 3DCTV on the image inpainting task. The experimental results under different sampling ratios (SR) are listed in Table 4. Furthermore, we also provide the visual results under sampling ratio 5% in Fig. 7. It can be easily observed that the proposed 3DCTV regularized model can achieve a compelling performance as compared with two other popular regularizers, namely nuclear norm (NN) and tensor nuclear norm (TNN). We think such experiments could show a wider range of potential usefulness of the proposed regularizer to other data recovery tasks.

## References

[1] E. J. Candès and B. Recht, "Exact matrix completion via convex optimization," *Foundations of Computational mathematics*, vol. 9, no. 6, p. 717, 2009.
[2] E. J. Candès, X. Li, Y. Ma, and J. Wright, "Robust principal component analysis?" *Journal of the ACM (JACM)*, vol. 58, no. 3, p. 11, 2011.
[3] Z. Lin, M. Chen, and Y. Ma, "The augmented lagrange multiplier method for exact recovery of corrupted low-rank matrices," *arXiv preprint arXiv:1009.5055*, 2010.
[4] J.-F. Cai, E. J. Candès, and Z. Shen, "A singular value thresholding algorithm for matrix completion," *SIAM Journal on optimization*, vol. 20, no. 4, pp. 1956–1982, 2010.
[5] Z. Zhou, X. Li, J. Wright, E. Candes, and Y. Ma, "Stable principal component pursuit," in *2010 IEEE international symposium on information theory*. IEEE, 2010, pp. 1518–1522.
[6] C. Lu, J. Feng, Y. Chen, W. Liu, Z. Lin, and S. Yan, "Tensor robust principal component analysis: Exact recovery of corrupted low-rank tensors via convex optimization," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 5249–5257.
[7] C. Lu, J. Feng, Y. Chen, W. Liu, Z. Lin, and S. Yan, "Tensor robust principal component analysis with a new tensor nuclear norm," *IEEE transactions on pattern analysis and machine intelligence*, vol. 42, no. 4, pp. 925–938, 2019.