Supplementary Material for "Beyond Low-rankness: Guaranteed Matrix Recovery via Modified Nuclear Norm"

Jiangjun Peng[1] , **Yisi Luo**[2] , **Xiangyong Cao**[3*] , **Shuang Xu**[1] , **Deyu Meng**[2,4*]
[1] School of Mathematics and Statistics, Northwestern Polytechnical University, Xi'an 710129, China.
[2] School of Mathematics and Statistics, Xi'an Jiaotong University, Xi'an 710049, China.
[3] School of Computer Science and Technology, Xi'an Jiaotong University, Xi'an 710049, China.
[4] Macao Institute of Systems Engineering, Macau University of Science and Technology, Taipa, Macao
{pengjj, xs}@nwpu.edu.cn, yisiluo1221@foxmail.com, caoxiangyong45@gmail.com,
dymeng@mail.xjtu.edu.cn

Abstract

In this document, we first verify whether MNN satisfies the definition of a norm in Section 1. Next, we provide proofs for the exact recoverability theories of MNN-RPCA (i.e., Theorem 1 in the manuscript) and MNN-MC (i.e., Theorem 2 in the manuscript) in Sections 2 and 3, respectively. Section 4 presents the proof of the optimality objective function as presented in the manuscript. Finally, in Section 5, we provide detailed descriptions of the data used in the manuscript and conduct more granular comparative experiments to further validate the effectiveness of our proposed MNN framework. Code and Supplementary Material are available at https://github.com/andrew-pengjj/modified_nuclear_norm.

Contents

*Corresponding authors

1 Verification of the MNN

Determining whether a definition constitutes a norm, mainly involves verifying its non-negativity, homogeneity, and triangle inequality properties.

It is known that the nuclear norm satisfies the aforementioned three properties of a norm, then we have

1. Non-negativity: $\|\mathbf{X}\|_* \geq 0$, and $\|\mathbf{X}\|_* = 0$ if and only if $\mathbf{X} = \mathbf{0}$.
2. Homogeneity: $\|\alpha\mathbf{X}\|_* = |\alpha| \cdot \|\mathbf{X}\|_*$ for any α.
3. Triangle Inequality: $\|\mathbf{X}+\mathbf{Y}\|_* \leq \|\mathbf{X}\|_* + \|\mathbf{Y}\|_*$.

Since MNN of $\mathbf{X}$ is $\|\mathcal{D}(\mathbf{X})\|_*$ with the fixed normalized linear transformation operator, we can easily obtain the following facts:

1. Non-negativity: $\|\mathcal{D}(\mathbf{X})\|_* \geq 0$, and $\|\mathbf{X}\|_* = 0$ if and only if $\mathbf{X} = \mathbf{0}$.
2. Homogeneity: $\|\alpha\mathcal{D}(\mathbf{X})\|_* = |\alpha| \cdot \|\mathcal{D}(\mathbf{X})\|_*$ for any α.
3. Triangle Inequality: $\|\mathcal{D}(\mathbf{X}+\mathbf{Y})\|_* = \|\mathcal{D}(\mathbf{X}) + \mathcal{D}(\mathbf{Y})\|_* \leq \|\mathcal{D}(\mathbf{X})\|_* + \|\mathcal{D}(\mathbf{Y})\|_*$.

Therefore, the MNN definition is a well-defined norm.

2 The Proof of MNN-RPCA Theorem

For readability, here we list the three mild assumptions in the manuscript as follows:

Assumption 1 (Incoherence Condition)**.** *For the low-rank matrix $\mathbf{X}_0 \in \mathbb{R}^{n_1\times n_2}$ with rank r, it follows the incoherence condition with parameter μ, i.e.,*

$$\begin{aligned}&\max_k \|\mathbf{U}^* e_k\| \leq \frac{\mu r}{n_1}, \max_k \|\mathbf{V}^* e_k\| \leq \frac{\mu r}{n_2},\\ &\|\mathbf{U}\mathbf{V}^*\|_\infty \leq \sqrt{\frac{\mu r}{n_1 n_2}},\end{aligned} \tag{1}$$

where $\mathbf{U} \in \mathbb{R}^{n_1 \times r}$ and $\mathbf{V} \in \mathbb{R}^{n_2 \times r}$ are obtained from the compact singular vector decomposition of $\mathbf{X}_0$ and e_k is the unit orthogonal vector.

Assumption 2 (Random Distribution)**.** *For the sparse term $\mathbf{S}_0$, its support Ω is chosen uniformly among all sets of cardinality m, and the signs of supports are random, i.e.*

$$\mathbb{P}[(\mathbf{S}_0)_{i,j} > 0 | (i,j) \in \Omega] = \mathbb{P}[(\mathbf{S}_0)_{i,j} \leq 0 | (i,j) \in \Omega] = 0.5. \tag{2}$$

Assumption 3 (Normalization)**.** *The Frobenius norm of the matrix corresponding to the linear transformation operator $\mathcal{D}(\cdot)$ is one.*

The MNN-induced RPCA model in the manuscript is:

$$\min_{\mathbf{X},\mathbf{S}} \|\mathcal{D}(\mathbf{X})\|_* + \lambda \|\mathbf{S}\|_1, s.t.\ \mathbf{M} = \mathbf{X} + \mathbf{S}. \tag{3}$$

The exact recoverable guarantee for MNN-RPCA model is:

Theorem 1 (MNN-RPCA Theorem)**.** *Suppose that $\mathcal{D}(\mathbf{X}_0) \in \mathbb{R}^{n_1 \times n_2}$ and $\mathbf{S}_0$ obey Assumptions 1-3, respectively. Without loss of generality, suppose $n_1 \geq n_2$. Then, there is a numerical constant $c > 0$ such that with probability at least $1 - cn_1^{-10}$(over the choice of support of $\mathbf{S}_0$), the MNN-RPCA model (3) with $\lambda = 1/(\sqrt{n_1})$ is exact, i.e., the solution $(\hat{\mathbf{X}}, \hat{\mathbf{S}}) = (\mathbf{X}_0, \mathbf{S}_0)$, provided that*

$$\textit{rank}(\mathbf{X}_0) \leq \rho_r n_2 \mu^{-1} (\log n_1)^{-2}, \text{ and } m \leq \rho_s n_1 n_2, \tag{4}$$

where ρ_r and ρ_s are some positive numerical constants, and m is the number of the support set of $\mathbf{S}_0$.

For ease of exposition, let us first provide the equivalent model of MNN-RPCA.

2.1 The Equivalent Model of MNN-RPCA Model

Since $\mathcal{D}(\cdot)$ is a linear transformation operator, $\mathcal{D}(\cdot)$ can be rewritten as $\mathbf{AX}$, that is, the role of $\mathcal{D}(\cdot)$ can be expressed into a corresponding matrix $\mathbf{A}$. For example, the difference operator $\nabla(\cdot)$ on row of matrix $\mathbf{X} \in \mathbb{R}^{n_1 \times n_2}$ can be rewritten as:

$$\nabla(\mathbf{X}) = \mathbf{AX}, \mathbf{A} = \mathbf{circ}(-1, 1, \overbrace{0, \ldots, 0}^{n_1 - 2}) \tag{5}$$

where "**circ**" denotes the circulant matrix. Thus, the MNN-RPCA model (3) can be rewritten as

$$\min_{\mathbf{X},\mathbf{S}} \quad \|\mathbf{AX}\|_* + \lambda \|\mathbf{S}\|_1, s.t.\ \mathbf{M} = \mathbf{X} + \mathbf{S}. \tag{6}$$

2.2 Mathematical Prerequisites

Before proving Theorem 1, it is helpful to review some basic concepts and introduce certain symbols.

Assuming we have a large data matrix $\mathbf{M}$ and know that it can be decomposed as:

$$\mathbf{M} = \mathbf{X}_0 + \mathbf{S}_0, \tag{7}$$

where $\mathbf{X}_0$ and $\mathbf{S}_0$ are the low-rank and sparse components, respectively. Theorem 1 asserts that by solving a variant of the MNN-RPCA model (6), we can obtain an exact decomposition $(\mathbf{X}_0, \mathbf{S}_0)$.

For a given scalar x, we use $\text{sgn}(x)$ to denote the sign of x. Extended to the matrix case, $\text{sgn}(\mathbf{S})$ is a matrix whose elements represent the signs of the elements of $\mathbf{S}$. Recalling the subdifferential of the ℓ_1 norm at $\mathbf{S_0}$, it takes the form on the support set Ω:

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

where $\mathbf{F}$ is zero on Ω, i.e., $\mathcal{P}_\Omega \mathbf{F} = 0$, and satisfies $||\mathbf{F}||_\infty \leq 1$.

Next, we assume that $\mathbf{X}_0$ with rank r has a singular value decomposition $\mathbf{U\Sigma V}^T$, where $\mathbf{U} \in \mathbb{R}^{n_1 \times r}$ and $\mathbf{V} \in \mathbb{R}^{n_2 \times r}$. Then, according to the chain rule of derivatives, we obtain

$$\frac{\partial \|\mathbf{AX}_0\|_*}{\partial \mathbf{X}_0} = \mathbf{A}^T \mathbf{UV}^T + \mathbf{A}^T \mathbf{W}, \tag{8}$$

where $\mathbf{U}^T\mathbf{W} = \mathbf{0}$, $\mathbf{WV} = \mathbf{0}$, and $\|\mathbf{W}\| \leq 1$. We use T to denote the linear space of matrices, i.e.,

$$T := \{\mathbf{UX}^T + \mathbf{YV}^T, \mathbf{X} \in \mathbb{R}^{n_1 \times r}, \mathbf{Y} \in \mathbb{R}^{n_2 \times r}\},$$

and use $T^\perp$ to represent its orthogonal complement. For any matrix $\mathbf{M}$, the projections onto T and $T^\perp$ are given by

$$\begin{aligned} &\mathcal{P}_T \mathbf{M} = \mathbf{UU}^T\mathbf{M} + \mathbf{MVV}^T - \mathbf{UU}^T\mathbf{MVV}^T, \\ &\mathcal{P}_{T^\perp} \mathbf{M} = (\mathbf{I} - \mathbf{UU}^T)\mathbf{M}(\mathbf{I} - \mathbf{VV}^T), \end{aligned} \tag{9}$$

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

Therefore, for any matrix of the form $\hat{\mathbf{e}}_k \hat{\mathbf{e}}_j^T$, it can be easily observed that

$$\begin{aligned} \|\mathcal{P}_{T^\perp} \hat{\mathbf{e}}_k \hat{\mathbf{e}}_j^T\|_F^2 &= \|(\mathbf{I} - \mathbf{UU}^T)\hat{\mathbf{e}}_k\|^2 \|(\mathbf{I} - \mathbf{VV}^T)\hat{\mathbf{e}}_j\|^2 \\ &\geq \left(1 - \frac{\mu r}{n_1}\right)\left(1 - \frac{\mu r}{n_2}\right). \end{aligned}$$

Assuming $\mu r / n_{(1)} \leq 1$, where $n_{(1)} = \max\{n_1, n_2\}$ and $n_{(2)} = \min\{n_1, n_2\}$. Because $\|\mathcal{P}_{T^\perp} \hat{\mathbf{e}}_k \hat{\mathbf{e}}_j^T\|_F^2 + \|\mathcal{P}_T \hat{\mathbf{e}}_k \hat{\mathbf{e}}_j^T\|_F^2 = 1$, we can derive

$$\|\mathcal{P}_T \hat{\mathbf{e}}_k \hat{\mathbf{e}}_j^T\|_F \leq \sqrt{\frac{2\mu r}{n_{(2)}}}. \tag{10}$$

2.3 Elimination Theorem

We begin with a useful definition.

Definition 1. *If for any $\mathbf{S}'_{ij} \neq 0$, $\mathbf{S}'$ satisfies supp$(\mathbf{S}') \subset$ supp$(\mathbf{S})$ and $\mathbf{S}'_{ij} = \mathbf{S}_{ij}$, we call $\mathbf{S}'$ a trimmed version of $\mathbf{S}$.*

In other words, a trimmed version of $\mathbf{S}$ is obtained by setting some elements of $\mathbf{S}$ to zero. Next, the following elimination theorem asserts that if the solution in Eq. (6) recovers the low-rank and sparse components of $\mathbf{M}_0 = \mathbf{X}_0 + \mathbf{S}_0$, then it can also correctly recover 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. *Assume that the solution of Eq. (6) is exact when the input is $\mathbf{M}_0 = \mathbf{X}_0 + \mathbf{S}_0$, 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 of Eq. (6) is also exact when the input is $\mathbf{M}'_0 = \mathbf{X}_0 + \mathbf{S}'_0$.*

Proof. First, for some $\Omega_0 \subset [n]\times[n]$, we express $\mathbf{S}_0^{'}$ as $\mathbf{S}_0^{'} = \mathbf{P}_{\Omega_0}\mathbf{S}_0$, and let $(\widehat{\mathbf{X}}, \widehat{\mathbf{S}})$ be the solution of Equation (6) when the input is $\mathbf{M}_0^{'} = \mathbf{X}_0 + \mathbf{S}_0^{'}$. Then, we have

$$\|\mathbf{A}\widehat{\mathbf{X}}\|_* + \lambda\|\widehat{\mathbf{S}}\|_1 \leq \|\mathbf{A}\mathbf{X}_0\|_* + \lambda\|\mathbf{P}_{\Omega_0}\mathbf{S}_0\|_1.$$

Further, we can obtain

$$\|\mathbf{A}\widehat{\mathbf{X}}\|_* + \lambda\left(\|\widehat{\mathbf{S}}\|_1 + \|\mathbf{P}_{\Omega_0^\perp}\mathbf{S}_0\|_1\right) \leq \|\mathbf{A}\mathbf{X}_0\|_* + \lambda\|\mathbf{S}_0\|_1.$$

Noting that $(\widehat{\mathbf{X}}, \widehat{\mathbf{S}} + \mathbf{P}_{\Omega_0^\perp}\mathbf{S}_0)$ is a feasible solution of Eq. (6) when the measurement matrix satisfies the condition $\mathbf{M}_0 = \mathbf{X}_0 + \mathbf{S}_0$, and $\|\widehat{\mathbf{S}} + \mathbf{P}_{\Omega_0^\perp}\mathbf{S}_0\|_1 \leq \|\widehat{\mathbf{S}}\|_1 + \|\mathbf{P}_{\Omega_0^\perp}\mathbf{S}_0\|_1$, we obtain

$$\begin{aligned}
&\|\mathbf{A}\widehat{\mathbf{X}}\|_* + \lambda\left(\|\widehat{\mathbf{S}} + \mathbf{P}_{\Omega_0^\perp}\mathbf{S}_0\|_1\right) \\
&\leq \|\mathbf{A}\widehat{\mathbf{X}}\|_* + \lambda\left(\|\widehat{\mathbf{S}}\|_1 + \|\mathbf{P}_{\Omega_0^\perp}\mathbf{S}_0\|_1\right) \\
&\leq \|\mathbf{A}\mathbf{X}_0\|_* + \lambda\|\mathbf{S}_0\|_1.
\end{aligned}$$

However, the above right-hand side is the optimal value, and by the uniqueness of the optimal solution, we must have $\mathbf{A}\widehat{\mathbf{X}} = \mathbf{A}\mathbf{X}_0$ and $\widehat{\mathbf{S}} + \mathbf{P}_{\Omega_0^\perp}\mathbf{S}_0 = \mathbf{S}_0$. Therefore, we further obtain $\widehat{\mathbf{X}} = \mathbf{X}_0$ and $\widehat{\mathbf{S}} = \mathbf{P}_{\Omega_0}\mathbf{S}_0 = \mathbf{S}^{'}$. This proves the lemma. □

2.4 Dual Verification

Lemma 2. *Suppose $\|\mathcal{P}_\Omega\mathcal{P}_T\| \leq 1$. If there exists a pair $(\mathbf{W}, \mathbf{F})$ satisfying the following conditions:*

$$\mathbf{A}^T\mathbf{U}\mathbf{V}^T + \mathbf{A}^T\mathbf{W} = \lambda(sgn(\mathbf{S}_0) + \mathbf{F}) \tag{11}$$

where $\mathcal{P}_T(\mathbf{W}) = 0$, $\|\mathbf{W}\| < 1$, and $\mathcal{P}_\Omega(\mathbf{F}) = 0$, $\|\mathbf{F}\|_\infty < 1$, then $(\mathbf{X}_0, \mathbf{S}_0)$ is the unique solution of Eq. (6).

Lemma 3. *Assuming $\|\mathcal{P}_\Omega\mathcal{P}_T\| \leq 1/2$. If there exists $(\mathbf{W}, \mathbf{F})$ satisfying*

$$\mathbf{A}^T\mathbf{U}\mathbf{V}^T + \mathbf{A}^T\mathbf{W} = \lambda(sgn(\mathbf{S}_0) + \mathbf{F} + \mathcal{P}_\Omega\mathbf{D}) \tag{12}$$

where $\mathcal{P}_T(\mathbf{W}) = 0$, $\|\mathbf{W}\| < \frac{1}{2}$, $\|\mathbf{F}\|_\infty < \frac{1}{2}$, and $\|\mathcal{P}_\Omega\mathbf{D}\|_\infty \leq 1/4$, then $(\mathbf{X}_0, \mathbf{S}_0)$ is the optimal solution of Eq. (6).

The above lemma indicates that, to prove our final exact recovery result, it suffices to generate dual variables $\mathbf{W}$ satisfying the following conditions:

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

2.5 Construct dual variables based on the Golfing Scheme

The remaining task is to construct dual variables that satisfy the corresponding conditions mentioned above. Before introducing our construction method, we first assume that $\Omega \sim \mathrm{Ber}(\rho)$, or equivalently, $\Omega^c \sim \mathrm{Ber}(1-\rho)$. Naturally, the distribution of Ω^c is identical to the distribution of $\Omega^c = \Omega_1 \cup \Omega_2 \cup \cdots \cup \Omega_{j_0}$, where each Ω_j follows a Bernoulli model with parameter q, i.e.,

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

Thus, if the following condition is satisfied:

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

the two models are equivalent. We decompose $\mathbf{W}$ as

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

where each component can be constructed as follows:

Constructing $\mathbf{W}^L$ based on the Golfing scheme. Suppose $j_0 \geqslant 1$, and let $\Omega_j, 1 \leq j \leq j_0$ be such that $\Omega^c = \cup_{1\leq j\leq j_0}\Omega_j$. We then define

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

where

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

Constructing $\mathbf{W}^S$ via Least Squares. Suppose $\|\mathcal{P}_\Omega\mathcal{P}_T\| \leq \frac{1}{2}$. Then, $\|\mathcal{P}_\Omega\mathcal{P}_T\mathcal{P}_\Omega\| < \frac{1}{4}$, and thus, the operator $\mathcal{P}_\Omega - \mathcal{P}_\Omega\mathcal{P}_T\mathcal{P}_\Omega$, which maps Ω to itself, is invertible with the inverse operator denoted as $(\mathcal{P}_\Omega - \mathcal{P}_\Omega\mathcal{P}_T\mathcal{P}_\Omega)^{-1}$. We then set

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

Eq. (16) is equivalent to

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

Since both $\mathbf{W}^L$ and $\mathbf{W}^S$ belong to $T^\perp$, and $\mathcal{P}_\Omega\mathbf{W}^S = \lambda\mathcal{P}_\Omega(\mathcal{I} - \mathcal{P}_T)\left(\mathcal{P}_\Omega - \mathcal{P}_\Omega\mathcal{P}_T\mathcal{P}_\Omega\right)^{-1}(\mathrm{sgn}(\mathbf{S_0})) = \lambda\mathrm{sgn}(\mathbf{S}_0)$, we only need to show that if $\mathbf{W}^L + \mathbf{W}^S$ satisfies the following conditions:

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

Then it is an effective dual certificate. The proof of Eq. (18) can be ensured through the following two lemmas.

Lemma 4. *Assuming $\Omega \sim Ber(\rho)$, where $\rho \leq \rho_s$ for some $\rho_s > 0$. Set $j_0 = 2\lceil \mathrm{log}n\rceil$ (for rectangular matrices, use $\mathrm{log}n_{(1)}$). Then, the $\mathbf{W}^L$ in Eq. (14) satisfies the following conditions:*

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

(b) $\|\mathcal{P}_\Omega\left(\mathbf{A}^T\mathbf{U}\mathbf{V}^T + \mathbf{A}^T\mathbf{W}^L\right)\|_F \leq \lambda\eta$,

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

Lemma 5. *Assume $\Omega \sim Ber(\rho_s)$, and the signs of $\mathbf{S}_0$ are independent and identically distributed symmetric random variables (independent of Ω). Then, the matrix $\mathbf{W}^S$ in Eq. (16) satisfies the following conditions:*

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

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

2.6 Proof of Dual Verification

Before proving Lemma 4, we need to invoke the following three lemmas.

Lemma 6. *[Lemma 4.1 in [Candes and Recht, 2012]]: Assuming $\Omega_0 \sim Ber(\rho_0)$. For some $C_0 > 0$, when $\rho_0 \geq C_0\epsilon^{-2}\beta\mu r \log n_{(1)}/n_{(2)}$, the following equation*

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

hold with a high probability.

Lemma 7. *[Lemma 3.1 in [Candès et al., 2011]]: Suppose $\mathbf{Z} \in T$ is a fixed matrix, and $\Omega_0 \sim Ber(\rho_0)$, when $\rho_0 \geq C_0\epsilon^{-2}\mu r \log n_{(1)}/n_{(2)}$,*

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

holds with a high probability.

Lemma 8. *[Lemma 3.2 in [Candès et al., 2011] and Lemma 6.3 in [Candes and Recht, 2012]]: Suppose $\mathbf{Z}$ is a fixed matrix, and $\Omega_0 \sim Ber(\rho_0)$. For some constant $C_1 > 0$, when $\rho_0 \geq C_1\mu \log n_{(1)}/n_{(2)}$,*

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

holds with a high probability.

Lemma 9. *[Lemma 6.3 in [Candes and Recht, 2012] and Lemma 3.2 in [Candès et al., 2011]]: Suppose $\mathbf{Z}$ is a fixed matrix, and $\Omega_0 \sim Ber(\rho_0)$. For some small constant $C_1 > 0$, when $\rho_0 \geq C_1\mu \log n_{(1)}/n_{(2)}$,*

$$\|\left(\mathbf{I} - \rho_0^{-1}\mathcal{P}_{\Omega_0}\right)\mathbf{Z}\| \leq C_1\sqrt{\frac{\beta n_{(1)}\log n_{(1)}}{\rho_0}}\|\mathbf{Z}\|_\infty, \tag{22}$$

holds with a high probability.

Proof of Lemma 4

Proof. We introduce some symbols first. Setting

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

therefore, for all $j \geq 0$, we have $\mathbf{Z}_j \in T$. Considering the definition of $\mathbf{Y}_j$ in Eq. (15), it follows that $\mathbf{Y}_j \in \Omega^\perp$. Then, we have

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

Therefore, when

$$q \geq C_0\epsilon^{-2}\mu r \log n_{(1)}/n_{(2)}, \tag{23}$$

by employing Lemma 7, we obtain

$$\|\mathbf{Z}_j\|_\infty \leq \epsilon\|\mathbf{Z}_{j-1}\|_\infty. \tag{24}$$

Specifically, this ensures that the following equation holds with high probability, i.e.,

$$\|\mathbf{Z}_j\|_\infty \leq \epsilon^j\|\mathbf{U}\mathbf{V}^T\|_\infty. \tag{25}$$

When q satisfies Eq. (23), by Lemma 6, we have

$$\|\mathbf{Z}_j\|_F \leq \epsilon\|\mathbf{Z}_{j-1}\|_F \tag{26}$$

This further ensures the high-probability satisfaction of the following inequality,

$$\|\mathbf{Z}_j\|_F \leq \epsilon^j\|\mathbf{U}\mathbf{V}^T\|_F \leq \epsilon^j\sqrt{r}. \tag{27}$$

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

$$\begin{aligned}
\|\mathbf{W}^L\| = \|\mathbf{P}_{T^\perp}\mathbf{Y}_{j_0}\|_\infty &\leq \sum_j \|q^{-1}\mathcal{P}_{T^\perp}\mathcal{P}_{\Omega_j}\mathbf{Z}_{j-1}\| \\
&\leq \sum_j \|\mathcal{P}_{T_1^\perp}(q^{-1}\mathcal{P}_{\Omega_j}\mathbf{Z}_{j-1} - \mathbf{Z}_{j-1})\| \\
&\leq \sum_j \|q^{-1}\mathcal{P}_{\Omega_j}\mathbf{Z}_{j-1} - \mathbf{Z}_{j-1}\| \\
&\leq C_1\sqrt{\frac{\beta n_{(1)}\log n_{(1)}}{q}}\sum_j \|\mathbf{Z}_{j-1}\|_\infty \\
&\leq C_1\sqrt{\frac{\beta n_{(1)}\log n_{(1)}}{q}}\sum_j \epsilon^j\|\mathbf{U}\mathbf{V}^T\|_\infty \\
&\leq \frac{C_1}{(1-\epsilon)}\sqrt{\frac{\beta n_{(1)}\log n_{(1)}}{q}}\|\mathbf{U}\mathbf{V}^T\|_\infty.
\end{aligned} \tag{28}$$

where the fourth step follows from Lemma 9, and the fifth step can be directly obtained from Eq. (24). By using the conditions in Eq. (23) and Eq. (1), we can derive, for some constant C_2, we have:

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

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

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

By using Eq. (23) and Eq. (27), we can obtain:

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

Since$\epsilon \leq e^{-1}$, $j_0 \geq 2\log n_{(1)}$, we have $\epsilon^{j_0} \leq 1/n_{(1)}^2$. Since

$$\begin{aligned}
\|\mathcal{P}_\Omega(\mathbf{A}^T(\mathbf{U}\mathbf{V}^T + \mathbf{W}^L))\|_F &\leq \|\mathbf{A}^T(\mathbf{U}\mathbf{V}^T + \mathbf{W}^L)\|_F \\
&\leq \|\mathbf{A}^T\|\|(\mathbf{U}\mathbf{V}^T + \mathbf{W}^L)\|_F \\
&\leq \alpha\|(\mathbf{U}\mathbf{V}^T + \mathbf{W}^L)\|_F \\
&\leq \alpha\|\mathbf{Z}_{j_0}\|_F \leq \frac{\sqrt{r}\alpha}{n_{(1)}^2}.
\end{aligned}$$

holds with at least $1 - n_{(1)}^{-\beta}$ probability, where $\beta > 2$, and $\lambda\gamma \leq 1/4$. Therefore, when $n_{(1)} \geq 2r^{1/4}\alpha^{1/2}$, we can easily establish the validity of condition (b). This completes the proof of this condition.

Proof of (c). Since $\mathbf{A}^T$ satisfies Assumption 3, we have

$$\|\mathbf{A}^T(\mathbf{U}\mathbf{V}^T + \mathbf{W}^L)\|_\infty \leq \|\mathbf{U}\mathbf{V}^T + \mathbf{W}^L\|_\infty. \tag{29}$$

Therefore, we only need to prove $\|\mathbf{U}\mathbf{V}^T + \mathbf{W}^L\|_\infty \leq \lambda/4$. We have $\mathbf{U}\mathbf{V}^T + \mathbf{W}^L = \mathbf{Z}_{j_0} + \mathbf{Y}_{j_0}$ and know that the support set of $\mathbf{Y}_{j_0}$ is Ω^c. Thus, since $\|\mathbf{Z}_{j_0}\|_\infty \leq \|\mathbf{Z}_{j_0}\|_F \leq$

$\lambda/8$, proving condition (c) is satisfied requires demonstrating $\|\mathbf{Y}_{j_0}\|_\infty \le \frac{\lambda}{8}$. To achieve this, we derive

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

Since $\|\mathbf{U}\mathbf{V}^T\|_\infty \le \sqrt{\frac{\mu r}{n_1 n_2}}$ for some C', we can deduce that

$$\|\mathbf{Y}_{j_0}\|_\infty \le \frac{C'\epsilon^2}{\sqrt{\mu r n_{(2)}^{-1} n_{(1)} (\log n_{(1)})^2}}, \tag{30}$$

holds when q satisfies the Eq. (23). Setting $\lambda = 1/\sqrt{n_{(1)}}$, we can obtain that when

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

holds, then $\|\mathbf{Y}_{j_0}\|_\infty \le \lambda/8$.

We have seen that if ϵ is small enough and $j_0 \ge 2\log n_{(1)}$, then (a)-(b) are satisfied. For (c), we can choose ϵ to be $\mathcal{O}\left((\mu r(\log n_{(1)})^2/n_{(2)})^{1/4}\right)$, as long as ρ_r in Eq. (4) is small enough, ensuring that ϵ is also small enough. Note that everything is consistent, as $C_0\epsilon^{-2}\mu r \log n_{(1)}/n_{(2)} < 1$. □

Proof of Lemma 5

Proof. Following the proof in Lemma 2.9 in [Candès *et al.*, 2011], we have

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

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

According to Assumption 3 about the transformation operator in the manuscript, we can further obtain

$$\begin{aligned}\|\mathcal{P}_{\Omega^\perp}(\mathbf{A}^T\mathbf{W}^S)\|_\infty &\le \|\mathbf{A}^T\mathbf{W}^S\|_\infty \\ &\le \|\mathbf{W}^S\|_\infty < \lambda/4.\end{aligned} \tag{31}$$

Thus, this lemma holds. The proof is completed. □

2.7 Proofs of Some Key Lemmas

Proof of Lemma 2

Proof. It is easy to see that for any non-zero matrix $\mathbf{H}$, the pair $(\mathbf{X}_0+\mathbf{H}, \mathbf{S}_0-\mathbf{H})$ is also a feasible solution. Next, we show that the objective function at $(\mathbf{X}_0, \mathbf{S}_0)$ is greater than the objective function at $(\mathbf{X}_0+\mathbf{H}, \mathbf{S}_0-\mathbf{H})$, thus proving that $(\mathbf{X}_0, \mathbf{S}_0)$ is the unique solution. To achieve this, let $\mathbf{U}\mathbf{V}^T + \mathbf{W}^0$ be an arbitrary subgradient of the nuclear norm at $\mathbf{X}_0$, and $\mathrm{sgn}(\mathbf{S}_0)+\mathbf{F}_0$ be an arbitrary subgradient of the ℓ_1 norm at $\mathbf{S}_0$. Then we have

$$\begin{aligned}&\|\mathbf{A}\mathbf{X}_0+\mathbf{A}\mathbf{H}\|_* + \lambda\|\mathbf{S}_0-\mathbf{H}\|_1 \ge \|\mathbf{A}\mathbf{X}_0\|_* + \lambda\|\mathbf{S}_0\|_1 \\ &\quad + \left\langle \mathbf{A}^T\mathbf{U}\mathbf{V}^T + \mathbf{A}^T\mathbf{W}^0, \mathbf{H}\right\rangle - \lambda\left\langle \mathrm{sgn}(\mathbf{S}_0)+\mathbf{F}_0, \mathbf{H}\right\rangle.\end{aligned}$$

Through the duality relationship between the nuclear norm and the operator norm, we can choose $\mathbf{W}^0$ such that $\left\langle \mathbf{W}^0, \mathbf{A}\mathbf{H}\right\rangle = \|\mathcal{P}_{T^\perp}(\mathbf{A}\mathbf{H})\|_*$. Similarly, through the duality relationship between the ℓ_1 norm and the operator norm, we can choose $\mathbf{F}_0$ such that $\langle \mathbf{F}_0, \mathbf{H}\rangle = \|\mathcal{P}_{\Omega^\perp}(\mathbf{H})\|_1$[1]. Therefore, we further have:

$$\begin{aligned}\|\mathbf{A}\mathbf{X}_0+\mathbf{A}\mathbf{H}\|_* &+ \lambda\|\mathbf{S}_0-\mathbf{H}\|_1 \geqslant \|\mathbf{A}\mathbf{X}_0\|_* + \lambda\|\mathbf{S}_0\|_1 \\ &+ \|\mathcal{P}_{T^\perp}(\mathbf{A}\mathbf{H})\|_* + \lambda\|\mathcal{P}_{\Omega^\perp}\mathbf{H}\|_1 \\ &+ \left\langle \mathbf{A}^T\mathbf{U}\mathbf{V}^T - \lambda\mathrm{sgn}(\mathbf{S}_0), \mathbf{H}\right\rangle.\end{aligned}$$

Given the assumption (11), for $\beta = \max(\|\mathbf{W}\|, \|\mathbf{F}\|) < 1$, we have

$$\begin{aligned}\left|\left\langle \mathbf{A}^T\mathbf{U}\mathbf{V}^T - \lambda\mathrm{sgn}(\mathbf{S}_0), \mathbf{H}\right\rangle\right| &\le |\langle \mathbf{W}, \mathbf{A}\mathbf{H}\rangle| + \lambda\,|\langle \mathbf{F}, \mathbf{H}\rangle| \\ &\le \beta\left(\|\mathcal{P}_{T^\perp}(\mathbf{A}\mathbf{H})\|_* + \lambda\|\mathcal{P}_{\Omega^\perp}\mathbf{H}\|_1\right)\end{aligned}$$

Therefore, we have

$$\begin{aligned}&\|\mathbf{A}\mathbf{X}_0+\mathbf{A}\mathbf{H}\|_* + \lambda\|\mathbf{S}_0-\mathbf{H}\|_1 \\ &\qquad \ge \|\mathbf{A}\mathbf{X}_0\|_* + \lambda\|\mathbf{S}_0\|_1 \\ &\qquad + (1-\beta)\left(\|\mathcal{P}_{T^\perp}(\mathbf{A}\mathbf{H})\|_* + \lambda\|\mathcal{P}_{\Omega^\perp}\mathbf{H}\|_1\right).\end{aligned}$$

Note that $\Omega\bigcap T = \{0\}$, unless $\mathbf{H} = 0$, in which case $\|\mathcal{P}_{T^\perp}(\mathbf{A}\mathbf{H})\|_* + \lambda\|\mathcal{P}_{\Omega^\perp}\mathbf{H}\|_1 > 0$. The proof is completed. □

Proof of Lemma 3

Proof. Following the proof process of Lemma 2, we can first get

$$\begin{aligned}&\|\mathbf{A}\mathbf{X}_0+\mathbf{A}\mathbf{H}\|_* + \lambda\|\mathbf{S}_0-\mathbf{H}\|_1 \ge \|\mathbf{A}\mathbf{X}_0\|_* + \lambda\|\mathbf{S}_0\|_1 \\ &\quad + \frac{1}{2}\left(\|\mathcal{P}_{T^\perp}(\mathbf{A}\mathbf{H})\|_* + \lambda\|\mathcal{P}_{\Omega^\perp}\mathbf{H}\|_1\right) - \lambda/4\left\langle \mathcal{P}_\Omega\mathbf{D}, \mathbf{A}\mathbf{H}\right\rangle \\ &\quad \ge \|\mathbf{A}\mathbf{X}_0\|_* + \lambda\|\mathbf{S}_0\|_1 + \frac{1}{2}\left(\|\mathcal{P}_{T^\perp}(\mathbf{A}\mathbf{H})\|_* + \lambda\|\mathcal{P}_{\Omega^\perp}\mathbf{H}\|_1\right) \\ &\quad - \lambda/4\|\mathcal{P}_\Omega(\mathbf{A}\mathbf{H})\|_F.\end{aligned}$$

For any matrices $\mathbf{A}$ and $\mathbf{B}$, according to the normed triangle inequality $\|\mathbf{A}\mathbf{B}\|_F \le \|\mathbf{A}\|\|\mathbf{B}\|_F$, and the support set Ω is randomly distributed, we can get the following inequality:

$$\begin{aligned}\|\mathcal{P}_{\Omega^\perp}(\mathbf{A}\mathbf{H})\|_F &\le \|\mathbf{A}\|_F\|\mathcal{P}_{\Omega^\perp}(\mathbf{H})\|_F \\ &= \|\mathcal{P}_{\Omega^\perp}(\mathbf{H})\|_F,\end{aligned} \tag{32}$$

where the last equality holds because of Assumption 3 about the transformation operator in the manuscript.

Based on the following facts,

$$\begin{aligned}\|\mathcal{P}_\Omega(\mathbf{A}\mathbf{H})\|_F &\le \|\mathcal{P}_\Omega(\mathcal{P}_T + \mathcal{P}_{T^\perp})(\mathbf{A}\mathbf{H})\|_F \\ &\le \|\mathcal{P}_\Omega\mathcal{P}_T(\mathbf{A}\mathbf{H})\|_F + \|\mathcal{P}_\Omega\mathcal{P}_{T^\perp}(\mathbf{A}\mathbf{H})\|_F \\ &\le \frac{1}{2}\|\mathbf{A}\mathbf{H}\|_F + \|\mathcal{P}_{T^\perp}(\mathbf{A}\mathbf{H})\|_F \\ &\le \frac{1}{2}\|\mathcal{P}_\Omega(\mathbf{A}\mathbf{H})\|_F + \frac{1}{2}\|\mathcal{P}_{\Omega^\perp}(\mathbf{A}\mathbf{H})\|_F + \|\mathcal{P}_{T^\perp}(\mathbf{A}\mathbf{H})\|_F \\ &\le \frac{1}{2}\|\mathcal{P}_\Omega(\mathbf{A}\mathbf{H})\|_F + \frac{1}{2}\|\mathcal{P}_{\Omega^\perp}(\mathbf{H})\|_F + \|\mathcal{P}_{T^\perp}(\mathbf{A}\mathbf{H})\|_F,\end{aligned}$$

we can obtain

$$\|\mathcal{P}_\Omega(\mathbf{A}\mathbf{H})\|_F \le \|\mathcal{P}_{\Omega^\perp}(\mathbf{H})\|_F + 2\|\mathcal{P}_{T^\perp}(\mathbf{A}\mathbf{H})\|_F.$$

[1]For example, $\mathbf{F}_0 = -\mathrm{sgn}(\mathcal{P}_{\Omega^\perp}\mathbf{H})$ is such a matrix. Also, through the duality relationship between the nuclear norm and the operator norm, there exists a matrix $\mathbf{W}$ with $\|\mathbf{W}\| = 1$ such that $\langle \mathbf{W}, \mathcal{P}_{T^\perp}(\mathbf{A}\mathbf{H})\rangle = \|\mathcal{P}_{T^\perp}(\mathbf{A}\mathbf{H})\|_*$. We take $\mathbf{W}^0 = \mathcal{P}_{T^\perp}(\mathbf{W})$.

Based on the above formula, $\|\mathbf{X}\|_F \leq \|\mathbf{X}\|_1$ and $\|\mathbf{X}\|_F \leq \|\mathbf{X}\|_*$ for any matrix $\mathbf{X}$, we can get

$$\begin{aligned}\|\mathbf{AX}_0 + \mathbf{AH}\|_* + \lambda\|\mathbf{S}_0 - \mathbf{H}\|_1 \geq \|\mathbf{AX}_0\|_* + \lambda\|\mathbf{S}_0\|_1 \\ + \frac{1-\lambda}{2}\|\mathcal{P}_{T^\perp}(\mathbf{AH})\|_* + \frac{\lambda}{4}\|\mathcal{P}_{\Omega^\perp}\mathbf{H}\|_1.\end{aligned}$$

Therefore, we have $\frac{1-\lambda}{2}\|\mathcal{P}_{T^\perp}(\mathbf{AH})\|_* + \frac{\lambda}{4}\|\mathcal{P}_{\Omega^\perp}\mathbf{H}\|_1$ is strictly positive, Unless $\mathbf{H} \neq \mathbf{0}$. □

3 The Proof of MNN-MC Theorem

The MNN-MC model is:

$$\min_{\mathbf{X}} \|\mathcal{D}(\mathbf{X})\|_*, s.t.\ \mathcal{P}_\Omega(\mathbf{M}) = \mathcal{P}_\Omega(\mathbf{X}), \tag{33}$$

The exact recoverable guarantee for MNN-MC model (33) is:

Theorem 2 (MNN-MC Theorem)**.** *Suppose that $\mathcal{D}(\mathbf{X}_0) \in \mathbb{R}^{n_1\times n_2}$ and $\mathcal{D}(\cdot)$ obey Assumptions 1 and 3, $\Omega \sim Ber(p)$ and m is the number of Ω, where $Ber(p)$ represents the Bernoulli distribution with parameter p. Without loss of generality, suppose $n_1 \geq n_2$. Then, there exist universal constants $c_0, c_1 > 0$ such that $\mathbf{X}_0$ is the unique solution to MNN-MC model (33) with probability at least $1 - c_1 n_1^{-3}\log n_1$, provided that*

$$m \geq c_0\mu r n_1^{5/4}\log(n_1), \tag{34}$$

3.1 The Equivalent Model of MNN-MC Model

The variant of the MNN-MC model (33) can be rewritten as

$$\min_{\mathbf{X}} \quad \|\mathbf{AX}\|_*, \text{ s.t. } \mathcal{P}_\Omega\mathbf{M} = \mathcal{P}_\Omega\mathbf{X}. \tag{35}$$

The Lemma 10 gives sufficient conditions for the uniqueness of the minimizer to Eq. (35).

3.2 Dual Verification

Lemma 10. *Consider a matrix $\mathbf{X}_0 = \mathbf{U\Sigma V}^T$ of rank r which is feasible for the problem (35), and suppose that the following two conditions hold:*

1. *There exist a dual variable λ such that $\mathbf{Y} = \mathcal{R}_\Omega^*\lambda$ obeys*

$$\mathcal{P}_T\mathbf{Y} = \mathbf{UV}^T, \|\mathcal{P}_{T^\perp}\mathbf{Y}\| < 1. \tag{36}$$

2. *The sampling operator $\mathcal{R}_\Omega$ restricted to elements in T is injective.*

Then $\mathbf{X}_0$ is the unique minimizer.

The proof of Lemma 10 uses a standard fact which states that the nuclear norm and the spectral norm are dual to one another.

Lemma 11. *For each pair $\mathbf{W}$ and $\mathbf{H}$, we have*

$$\langle\mathbf{W}, \mathbf{H}\rangle \leq \|\mathbf{W}\|\|\mathbf{H}\|_*. \tag{37}$$

In addition, for each $\mathbf{H}$, there is a $\mathbf{W}$ obeying $\|\mathbf{W}\| = 1$ which achieves the quality.

A variety of proofs are available for Lemma 11, and an elementary argument is sketched in [Recht *et al.*, 2010]. We now turn to the proof of Lemma 10.

Proof. Consider any perturbation $\mathbf{X}_0 + \mathbf{H}$, where $\mathcal{R}_\Omega(\mathbf{H}) = 0$. Then for any $\mathbf{W}^0$ obeying $\mathbf{W}^0 \in T^\perp$ and $\|\mathbf{W}\| \leq 1$, $\mathbf{UV}^T + \mathbf{W}^0$ is a subgradient of the nuclear norm at $\mathbf{X}_0$, therefore,

$$\begin{aligned}\|\mathbf{AX}_0 + \mathbf{AH}\|_* \geq \|\mathbf{AX}_0\|_* \\ + \left\langle \mathbf{A}^T\mathbf{UV}^T + \mathbf{A}^T\mathbf{W}^0, \mathbf{H}\right\rangle.\end{aligned} \tag{38}$$

Letting $\mathbf{W} = \mathcal{P}_{T^\perp}(\mathbf{Y})$, we may write $\mathbf{UV}^T = \mathcal{R}_\Omega^*\lambda - \mathbf{W}$. Since $\|\mathbf{W}\| < 1$ and $\mathcal{R}_\Omega(\mathbf{H}) = 0$, it then follows that

$$\begin{aligned}\|\mathbf{AX}_0 + \mathbf{AH}\|_* &\geq \|\mathbf{AX}_0\|_* + \left\langle \mathbf{A}^T\mathbf{W}^0 - \mathbf{A}^T\mathbf{W}, \mathbf{H}\right\rangle \\ &\geq \|\mathbf{AX}_0\|_* + \left\langle \mathbf{W}^0 - \mathbf{W}, \mathbf{AH}\right\rangle.\end{aligned} \tag{39}$$

Now, by construction,

$$\begin{aligned}\left\langle \mathbf{W}^0 - \mathbf{W}, \mathbf{AH}\right\rangle &\geq \left\langle \mathcal{P}_{T^\perp}\left(\mathbf{W}^0 - \mathbf{W}\right), \mathbf{AH}\right\rangle \\ &= \left\langle \mathbf{W}^0 - \mathbf{W}, \mathcal{P}_{T^\perp}(\mathbf{AH})\right\rangle.\end{aligned} \tag{40}$$

We use Lemma 11 and set $\mathbf{W}^0 = \mathcal{P}_{T^\perp}(\mathbf{Z})$ where $\mathbf{Z}$ is any matrix obeying $\|\mathbf{Z}\| \leq 1$ and $\langle \mathbf{Z}, \mathcal{P}_{T^\perp}(\mathbf{AH})\rangle = \|\mathcal{P}_{T^\perp}(\mathbf{AH})\|_*$. Then $\mathbf{W}^0 \in \perp^T$, $\|\mathbf{W}^0\| \leq 1$, and

$$\left\langle \mathbf{W}^0 - \mathbf{W}, \mathbf{AH}\right\rangle \geq (1 - \|\mathbf{W}\|)\,\|\mathcal{P}_{T^\perp}(\mathbf{AH})\|_*, \tag{41}$$

which by assumption is strictly positive unless $\mathcal{P}_{T^\perp}(\mathbf{H}) = 0$. In other words, $\|\mathbf{AX}_0 + \mathbf{AH}\|_* \geq \|\mathbf{AX}_0\|_*$ unless $\mathcal{P}_{T^\perp}(\mathbf{H}) = 0$. Assume then that $\mathcal{P}_{T^\perp}(\mathbf{H}) = 0$ or equivalently that $\mathbf{H} \in T$. Then $\mathcal{R}_\Omega(\mathbf{H}) = 0$ implies that $\mathbf{H} = 0$ by the injectivity assumption. In conclusion, $\|\mathbf{AX}_0 + \mathbf{AH}\|_* \geq \|\mathbf{AX}_0\|_*$ unless $\mathbf{H} = 0$. □

3.3 Construct dual variables

The remaining task is to construct dual variable that satisfy the corresponding conditions mentioned in Lemma 10 (and show the injectivity of the sampling operator restricted to matrices in T along the way). Set $\mathcal{P}_\Omega$ to be the orthogonal projector onto the indices in Ω so that the (i, j)-th component of $\mathcal{P}_\Omega(\mathbf{X})$ is equal to X_{ij} if $(i, j) \in \Omega$ and zero otherwise.

Following the settings in [Candes and Recht, 2012], we introduce the operator $\mathcal{A}_{\Omega T}$ defined by

$$\mathcal{A}_{\Omega T}(\mathbf{X}) = \mathcal{P}_\Omega\mathcal{P}_T(\mathbf{X}). \tag{42}$$

Then if $\mathcal{A}_{\Omega T}^*\mathcal{A}_{\Omega T} = \mathcal{P}_T\mathcal{P}_\Omega\mathcal{P}_T$ has full rank when restricted to T, the dual variable $\mathbf{Y}$ can be given by

$$\mathbf{Y} = \mathcal{A}_{\Omega T}(\mathcal{A}_{\Omega T}^*\mathcal{A}_{\Omega T})^{-1}(\mathbf{UV}^T) \tag{43}$$

We clarify the meaning of (43) to avoid any confusion. $(\mathcal{A}_{\Omega T}^*\mathcal{A}_{\Omega T})^{-1}(\mathbf{UV}^T)$ is meant to be that element $\mathbf{F}$ in T obeying $(\mathcal{A}_{\Omega T}^*\mathcal{A}_{\Omega T})(\mathbf{F}) = \mathbf{UV}^T$. Then, according to the setup of $\mathbf{Y}$ in Eq. (43), we can obtain

$$\begin{aligned}\mathcal{P}_T(\mathbf{Y}) &= \mathcal{P}_T\mathcal{A}_{\Omega T}(\mathcal{A}_{\Omega T}^*\mathcal{A}_{\Omega T})^{-1}(\mathbf{UV}^T) \\ &= (\mathcal{P}_T\mathcal{P}_\Omega\mathcal{P}_T)(\mathcal{A}_{\Omega T}^*\mathcal{A}_{\Omega T})^{-1}(\mathbf{UV}^T) \\ &= (\mathcal{P}_T\mathcal{P}_\Omega\mathcal{P}_\Omega\mathcal{P}_T)(\mathcal{A}_{\Omega T}^*\mathcal{A}_{\Omega T})^{-1}(\mathbf{UV}^T) \\ &= (\mathcal{A}_{\Omega T}^*\mathcal{A}_{\Omega T})(\mathcal{A}_{\Omega T}^*\mathcal{A}_{\Omega T})^{-1}(\mathbf{UV}^T) \\ &= \mathbf{UV}^T.\end{aligned} \tag{44}$$

To summarize the aims of our proof strategy,

1) We must first show that $\mathcal{A}_{\Omega T}^*\mathcal{A}_{\Omega T} = \mathcal{P}_T\mathcal{P}_\Omega\mathcal{P}_T$ is a one-to-one linear mapping from T onto itself. In this case, $\mathcal{A}_{\Omega T} = \mathcal{P}_\Omega\mathcal{P}_T$ as mapping from T to $R^{n_1\times n_2}$ is injective. This is the second sufficient condition of Lemma 10. Moreover, our anstatz for $\mathbf{Y}$ given by Eq. (43) is well defined.

2) Having established that $\mathbf{Y}$ is well defined, we will show that

$$\|\mathcal{P}_{T^\perp}(\mathbf{Y})\| < 1, \tag{45}$$

thus proving the first sufficient condition.

3.4 Proof of Dual Verification

Next, we need to prove the injectivity of $\mathcal{A}_{\Omega T}$, and $\|\mathcal{P}_{T^\perp}(\mathbf{Y})\| < 1$.

The Injectivity Property

To prove this, we will show that the linear operator $p^{-1}\mathcal{P}_T(\mathcal{P}_\Omega - p\mathcal{I})\mathcal{P}_T$ has small operator norm, which we recall is $\sup_{\|\mathbf{X}\|_F\le 1} p^{-1}\|\mathcal{P}_T(\mathcal{P}_\Omega - p\mathcal{I})\mathcal{P}_T(\mathbf{X})\|_F$. Proving the above conclusion requires the following two lemmas.

Lemma 12 (Theorem 4.1 in [Candes and Recht, 2012]). *Suppose Ω is sampled from Ber(p) and put $n = \max\{n_1, n_2\}$. Suppose that the low-rank matrix $\mathbf{X}_0$ satisfies the incoherence condition with μ. Then there is a numerical constant C_R such that for all $\beta > 1$,*

$$p^{-1}\|\mathcal{P}_T\mathcal{P}_\Omega\mathcal{P}_T - p\mathcal{P}_T\| \le C_R\sqrt{\frac{\mu nr(\beta\log n)}{m}} \tag{46}$$

with probability at least $1 - 3n^{-\beta}$ provided that $C_R\sqrt{\frac{\mu nr(\beta\log n)}{m}} < 1$.

Lemma 13 (Theorem 4.2 in [Candes and Recht, 2012]). *Let $\{\delta_{ab}\}$ be independent $0/1$ Bernoulli variables with $\mathbb{P}(\delta_{ab} = 1) = p = \frac{m}{n_1 n_2}$ and put $n = \max\{n_1, n_2\}$. Suppose that $\|\mathcal{P}_T(\mathbf{e}_a\mathbf{e}_b^*)\|_F^2 \le 2\mu_0 r/n$. Set*

$$\begin{aligned}\mathbf{Z} &\equiv p^{-1}\left\|\sum_{ab}(\delta_{ab} - p)\mathcal{P}_T(\mathbf{e}_a\mathbf{e}_b^*)\otimes\mathcal{P}_T(\mathbf{e}_a\mathbf{e}_b^*)\right\| \\ &= p^{-1}\left\|\mathcal{P}_T\mathcal{P}_\Omega\mathcal{P}_T - p\mathcal{P}_T\right\|.\end{aligned} \tag{47}$$

1. There exists a constant $C_R^{'}$ such that

$$\mathbb{E}(\mathbf{Z}) \le C_R^{'}\sqrt{\frac{\mu nr\log n}{m}} \tag{48}$$

provided that the right-band side is smaller than 1.

2. Suppose $\mathbb{E}(\mathbf{Z}) \le 1$. Then for each $\lambda > 0$, we have

$$\begin{aligned}&\mathbb{P}\left(|\mathbf{Z} - \mathbb{E}(\mathbf{Z})| > \lambda\sqrt{\frac{\mu nr\log n}{m}}\right) \\ &\le 3\exp\left(-\gamma_0^{'}\min\left\{\lambda^2\log n, \lambda\frac{m\log n}{\mu nr}\right\}\right)\end{aligned} \tag{49}$$

for some positive constant $\gamma_0^{'}$.

Take m large enough so that $C_R\sqrt{\mu(nr/m)\log n} \le 1/2$. Then it follows from (46) that

$$\begin{aligned}\frac{p}{2}\|\mathcal{P}_T(\mathbf{X})\|_F &\le \|(\mathcal{P}_T\mathcal{P}_\Omega\mathcal{P}_T)(\mathbf{X})\|_F \\ &\le \frac{3p}{2}\|\mathcal{P}_T(\mathbf{X})\|_F\end{aligned} \tag{50}$$

for all $\mathbf{X}$ with large probability. In particular, the operator $\mathcal{A}_{\Omega T}^*\mathcal{A}_{\Omega T} = \mathcal{P}_T\mathcal{P}_\Omega\mathcal{P}_T$ mapping T onto itself is well conditioned, and hence invertible. An immediate consequence is the following corollary.

Corollary. *Assume that $C_R\sqrt{\mu(nr/m)\log n} \le 1/2$. With the same probability as in Lemma 12, we have*

$$\|\mathcal{P}_\Omega\mathcal{P}_T(\mathbf{X})\|_F \le \sqrt{3p/2}\|\mathcal{P}_T(\mathbf{X})\|_F. \tag{51}$$

Proof. We have $\|\mathcal{P}_\Omega\mathcal{P}_T(\mathbf{X})\|_F^2 = \langle\mathbf{X}, (\mathcal{P}_\Omega\mathcal{P}_T)^*\mathcal{P}_\Omega\mathcal{P}_T\mathbf{X}\rangle = \langle\mathbf{X}, (\mathcal{P}_T\mathcal{P}_\Omega\mathcal{P}_T)\mathbf{X}\rangle$, and thus

$$\begin{aligned}\|\mathcal{P}_\Omega\mathcal{P}_T(\mathbf{X})\|_F^2 &= \langle\mathcal{P}_T(\mathbf{X}), (\mathcal{P}_T\mathcal{P}_\Omega\mathcal{P}_T)\mathbf{X}\rangle \\ &\le \|\mathcal{P}_T(\mathbf{X})\|_F\|(\mathcal{P}_T\mathcal{P}_\Omega\mathcal{P}_T)\mathbf{X}\|_F,\end{aligned} \tag{52}$$

where the inequality is due to Cauchy-Schwarz. The conclusion (51) follows from (50). This completes the proof. □

The Size Property

Next, we prove that $\|\mathcal{P}_{T^\perp}(\mathbf{Y})\| < 1$. Before proving this conclusion, we need to introduce five lemmas.

Lemma 14 (Lemma 4.4 in [Candes and Recht, 2012]). *Fix $\beta \ge 2$ and $\lambda \ge 1$. There is a numerical constant C_0 such that if $m \ge \lambda\mu^2 nr\beta\log n$, then*

$$p^{-1}\|(\mathcal{P}_{T^\perp}\mathcal{P}_\Omega\mathcal{P}_T)(\mathbf{U}\mathbf{V}^T)\| \le C_0\lambda^{-1/2} \tag{53}$$

with probability at least $1 - n^{-\beta}$.

Lemma 15 (Lemma 4.5 in [Candes and Recht, 2012]). *Fix $\beta \ge 2$ and $\lambda \ge 1$. There are numerical constants C_1 and c_1 such that if $m \ge \lambda\mu\max(\sqrt{\mu}, \mu)nr\beta\log n$, then*

$$p^{-1}\|(\mathcal{P}_{T^\perp}\mathcal{P}_\Omega\mathcal{P}_T)\mathcal{H}(\mathbf{U}\mathbf{V}^T)\| \le C_1\lambda^{-1} \tag{54}$$

with probability at least $1 - c_1 n^{-\beta}$, where $\mathcal{H} \equiv \mathcal{P}_\Omega - p^{-1}\mathcal{P}_T\mathcal{P}_\Omega\mathcal{P}_T$.

Lemma 16 (Lemma 4.6 in [Candes and Recht, 2012]). *Fix $\beta \ge 2$ and $\lambda \ge 1$. There are numerical constants C_2 and c_2 such that if $m \ge \lambda\mu^{4/3}nr^{4/3}\beta\log n$, then*

$$p^{-1}\|(\mathcal{P}_{T^\perp}\mathcal{P}_\Omega\mathcal{P}_T)\mathcal{H}^2(\mathbf{U}\mathbf{V}^T)\| \le C_2\lambda^{-3/2} \tag{55}$$

with probability at least $1 - c_2 n^{-\beta}$.

Lemma 17 (Lemma 4.7 in [Candes and Recht, 2012]). *Fix $\beta \ge 2$ and $\lambda \ge 1$. There are numerical constants C_3 and c_3 such that if $m \ge \lambda\mu^2 nr^2\beta\log n$, then*

$$p^{-1}\|(\mathcal{P}_{T^\perp}\mathcal{P}_\Omega\mathcal{P}_T)\mathcal{H}^3(\mathbf{U}\mathbf{V}^T)\| \le C_3\lambda^{-1/2} \tag{56}$$

with probability at least $1 - c_3 n^{-\beta}$.

Lemma 18 (Lemma 4.8 in [Candes and Recht, 2012]). *Under the assumptions of Lemma 12, there is a numerical constant C_{k_0} such that if $m \geq (2C_R)^2 \mu n r \beta \log n$, then*

$$p^{-1} \left\| (\mathcal{P}_{T^\perp} \mathcal{P}_\Omega) \mathcal{P}_T \sum_{k \geq k_0} \mathcal{H}^k (\mathbf{U}\mathbf{V}^T) \right\| \leq C_{k_0} (\frac{n^2 r}{m})^{1/2} \times (\frac{munr\beta \log n}{m})^{k_0/2} \tag{57}$$

with probability at least $1 - 3n^{-\beta}$.

Next, we give the proof of $\|\mathcal{P}_{T^\perp}(\mathbf{Y})\| < 1$.

Proof. Introducing

$$\mathcal{H} \equiv \mathcal{P}_T - p^{-1} \mathcal{P}_T \mathcal{P}_\Omega \mathcal{P}_T, \tag{58}$$

where $\|\mathcal{H}(\mathbf{X})\|_F \leq C_R \sqrt{\mu_0 (nr/m) \beta \log n} \|\mathcal{P}_T(\mathbf{X})\|_F$ holds with large probability because of Lemma 12. For any matrix $\mathbf{X} \in T$, $(\mathcal{P}_T \mathcal{P}_\Omega \mathcal{P}_T)^{-1}(\mathbf{X})$ can be expressed in terms of power series

$$(\mathcal{P}_T \mathcal{P}_\Omega \mathcal{P}_T)^{-1}(\mathbf{X}) = p^{-1}(\mathbf{X} + \mathcal{H}(\mathbf{X}) + \mathcal{H}^2(\mathbf{X}) + \cdots) \tag{59}$$

for $\mathcal{H}$ is a contraction when m is sufficiently large. Since $\mathbf{Y} = \mathcal{P}_\Omega \mathcal{P}_T (\mathcal{P}_T \mathcal{P}_\Omega \mathcal{P}_T)^{-1} \times (\mathbf{U}\mathbf{V}^T)$, $\mathcal{P}_{T^\perp}(\mathbf{Y})$ may be decomposed as

$$\begin{aligned} \mathcal{P}_{T^\perp}(\mathbf{Y}) = p^{-1} (\mathcal{P}_{T^\perp} \mathcal{P}_\Omega \mathcal{P}_T)(\mathbf{U}\mathbf{V}^T + \mathcal{H}(\mathbf{U}\mathbf{V}^T) \\ + \mathcal{H}^2(\mathbf{U}\mathbf{V}^T) + \cdots). \end{aligned} \tag{60}$$

Therefore, to bound the norm of the left-hand side, it is of course sufficient to bound the norm of the summands in the right-hand side.

Under all of the assumptions of Theorem 2 in the manuscript, with the guidance of Lemmas 14, 15, 16, and 18, the latter applied with $k_0 = 3$. Together they imply that there are numerical constants c and c_0 such that $\|\mathcal{P}_{T^\perp}(\mathbf{Y})\| < 1$ with probability at least $1 - cn^{-\beta}$ provided that the number of samples obeys

$$m \geq c_0 \max\left(\mu^2, \mu^{4/3} r, \mu n^{1/4}\right) n r \beta \log(n). \tag{61}$$

This completes the proof. □

4 Proof of the Optimal Solution to the Objective Function

Based on Theorems 1 and 2 in the manuscript, we deduce the following corollary:

Corollary. *Suppose $\mathbf{X}_0$ and $\mathbf{S}_0$ satisfy Assumptions 1 and 2, and transformation operator $\mathcal{D}(\cdot)$ satisfy Assumption 3. Denote the objective functions of RPCA and MC models as*

$$\begin{aligned} \mathcal{J}_1^{\mathcal{D}}(\mathbf{X}) &:= \|\mathcal{D}(\mathbf{X})\|_* + \lambda \|\mathbf{M} - \mathbf{X}\|_1, \\ \mathcal{J}_2^{\mathcal{D}}(\mathbf{X}) &:= \|\mathcal{D}(\mathbf{X})\|_* + \mu \|\mathcal{P}_\Omega(\mathbf{M} - \mathbf{X})\|_F^2, \end{aligned} \tag{62}$$

respectively, where $\lambda = 1/\sqrt{n_1}$ and $\mu = (\sqrt{n_1} + \sqrt{n_2})\sqrt{p}\sigma$ according to [Candes and Plan, 2010], and n_1, n_2, σ, p are the sizes of matrix, noise standard variance, and missing ratio. Then, for any feasible solution $\mathbf{X}$, we have:

$$\mathcal{J}_1^{\mathcal{D}}(\mathbf{X}) \geq \mathcal{J}_1^{\mathcal{D}}(\mathbf{X}_0), \mathcal{J}_2^{\mathcal{D}}(\mathbf{X}) \geq \mathcal{J}_2^{\mathcal{D}}(\mathbf{X}_0). \tag{63}$$

Next, we give the proof of this corollary.

Proof. Since Theorem 1 holds under Assumptions 1-3, then we can conclude that the point $(\mathbf{X}_0, \mathbf{S}_0)$ minimizes the objective function value of the MNN-RPCA model (3), i.e., $\|\mathcal{D}(\mathbf{X}_0)\|_* + \lambda \|\mathbf{S}_0\|_* = \|\mathcal{D}(\mathbf{X}_0)\|_* + \lambda \|\mathbf{M} - \mathbf{X}_0\|_1$ is the minimum value among all feasible solutions. Therefore, we can obtain:

$$\mathcal{J}_1^{\mathcal{D}}(\mathbf{X}) \geq \mathcal{J}_1^{\mathcal{D}}(\mathbf{X}_0). \tag{64}$$

Similarly, for the MNN-MC model (33), we have:

$$\mathcal{J}_2^{\mathcal{D}}(\mathbf{X}) \geq \mathcal{J}_2^{\mathcal{D}}(\mathbf{X}_0). \tag{65}$$

This completes the proof. □

5 More Experiments

5.1 Data Description

In this section, we provide additional explanations for the data used in the manuscript.

Hyperspectral image datasets description. Hyperspectral images (HSI) differ from traditional images in that they capture radiance over numerous continuous discrete bands, resulting in three-dimensional data comprising a series of images. Due to the similarity across adjacent spectral bands captured at the same moment for the same object, HSIs exhibit strong global low-rank properties. Additionally, each spectral band contains individual images, enriching the dataset with local information. Hence, HSIs are suitable for testing in this study. We select five widely used HSIs, including Cuprite, DCMall, KSC, Pavia, and Pavia University (PaviaU), each preprocessed to a size of $200 \times 200 \times 50$ in the experiment.

Multispectral image datasets description. Multispectral images (MSI), akin to HSIs, capture radiance across a range of bands but typically have a smaller spectrum in the visible range. As such, MSIs also exhibit global low-rankness and contain local information. For this experiment, we utilize the CAVE database [2], a benchmark database for MSI data processing tasks, comprising 11 datasets sized $512 \times 512 \times 31$, including watercolors, toys, apples, paint, hairs, feathers, strawberries, lemons, face, clay, and bread.

Color video description. Color video data comprises streaming video data collected continuously over a short duration, showing correlation across adjacent time points and containing both global low-rank properties and local information. We test all methods on 10 color video sequences from the YUV database [3], each sized $176 \times 144 \times 3 \times 200$. To simplify processing, we select data from the first 100 time points and combine RGB channels and time dimensions, resulting in a final size of $176 \times 144 \times 300$ for the test data.

MRI-CT dataset description. MRI and CT image data are sequences of images acquired over time from patients, demonstrating continuity between adjacent time points due to the slow movement of the imaging device. Hence, MRI and CT images exhibit global low-rankness and contain local information, making them suitable for testing. We evaluate

[2] https://www.cs.columbia.edu/CAVE/databases/multispectral/
[3] http://trace.eas.asu.edu/yuv/

all methods on four datasets sized $256 \times 256 \times 127$ from the Cancer Imaging Archive [4].

5.2 More Results

In the manuscript, we only provide the mean values of all methods on four categories of data. From the table in the manuscript, we have seen that the MNN-based models achieve the best restoration performance in the vast majority of cases. Here, taking PSNR as an example, we further present the performance of all methods on each dataset.

Figure 1 to Figure 3 sequentially display histograms of restoration metric distributions for all methods on multiple datasets for denoising tasks under three categories: HSI, MSI, and Color Video. From these histograms, it can be observed that the PSNR values of our four MNN-based models consistently remain high, especially MNN-Sobel and MNN-L2. Additionally, we also notice that for a few datasets and noise levels, the PSNR values of TCTV are higher than those of the four MNN-based models and significantly higher than CTV. This indicates the necessity of employing tensor tools for modeling tensor data. Therefore, in the future, we will consider further extending the MNN framework to tensor structures.

Figures 4 to 7 sequentially display histograms of restoration metric distributions for all methods on multiple datasets for completing tasks under four categories: HSI, MSI, Color Video, and CT. Similar to the denoising results, the four MNN-based models perform excellently on the vast majority of datasets and missing rates, consistent with the mean PSNR results presented in Table 2 of the manuscript. Additionally, we also observe that when the sampling rate is relatively high, such as when the sampling rate (sr) is set to 0.2, KBR based on tensor decomposition exhibits outstanding PSNR performance, even surpassing MNN-Sobel, MNN-L1, and MNN-L2 on a few datasets. This indicates that tensor modeling still holds certain advantages.

References

[Candes and Plan, 2010] Emmanuel J Candes and Yaniv Plan. Matrix completion with noise. *Proceedings of the IEEE*, 98(6):925–936, 2010.

[Candes and Recht, 2012] Emmanuel Candes and Benjamin Recht. Exact matrix completion via convex optimization. *Communications of the ACM*, 55(6):111–119, 2012.

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

[Recht *et al.*, 2010] Benjamin Recht, Maryam Fazel, and Pablo A Parrilo. Guaranteed minimum-rank solutions of linear matrix equations via nuclear norm minimization. *SIAM review*, 52(3):471–501, 2010.

[4] https://www.cancerimagingarchive.net/

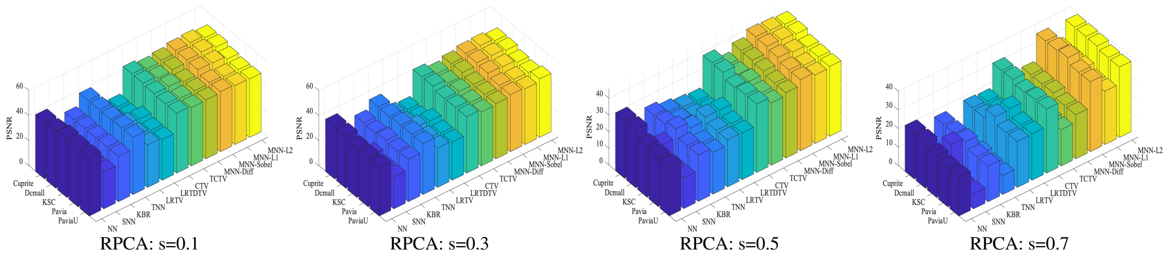


Figure 1: Performance comparison in terms of PSNR of recovered hyperspectral images obtained by all competing methods under RPCA tasks.

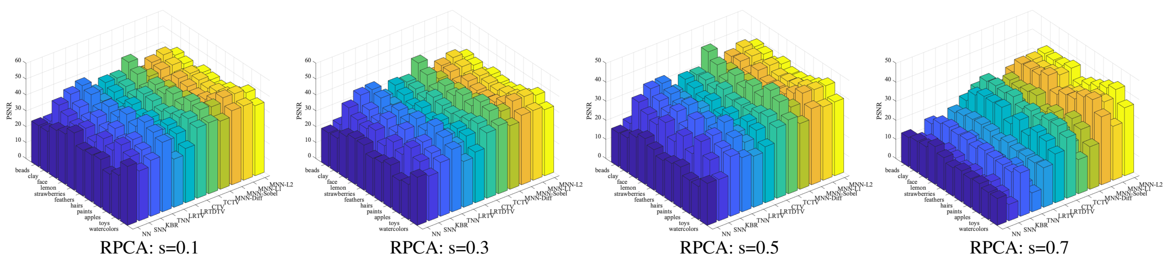


Figure 2: Performance comparison in terms of PSNR of recovered multispectral images obtained by all competing method under RPCA tasks.

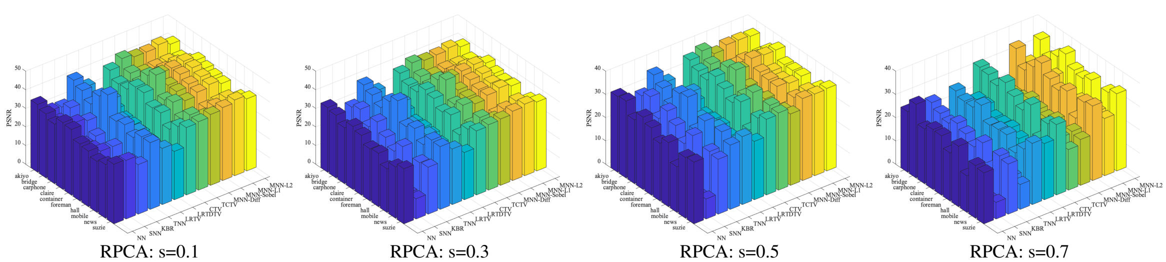


Figure 3: Performance comparison in terms of PSNR of recovered color videos obtained by all competing method under RPCA tasks.

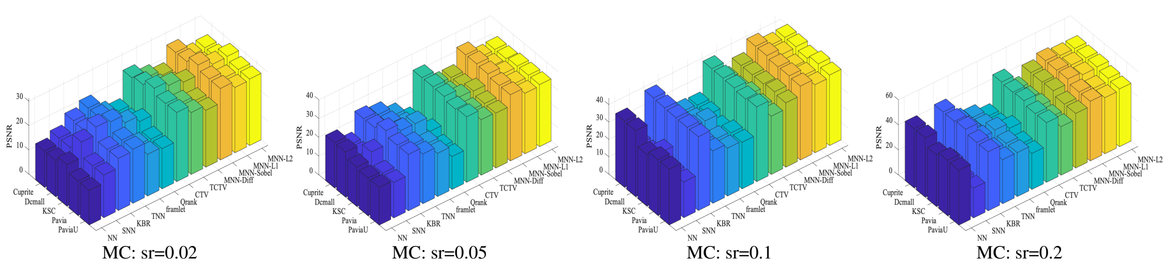


Figure 4: Performance comparison in terms of PSNR of recovered hyperspectral images obtained by all competing methods under MC tasks.

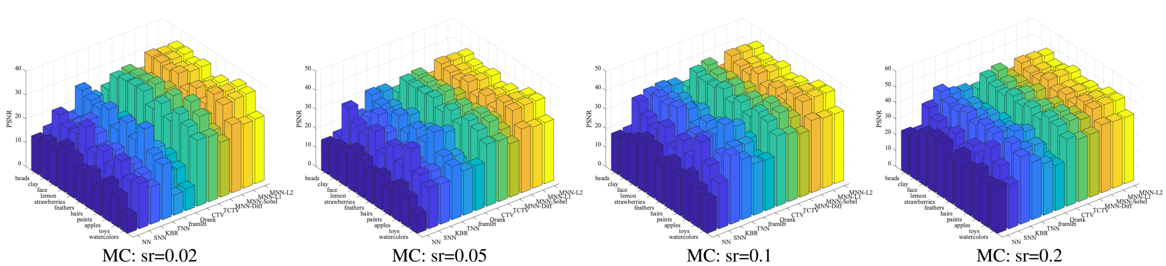
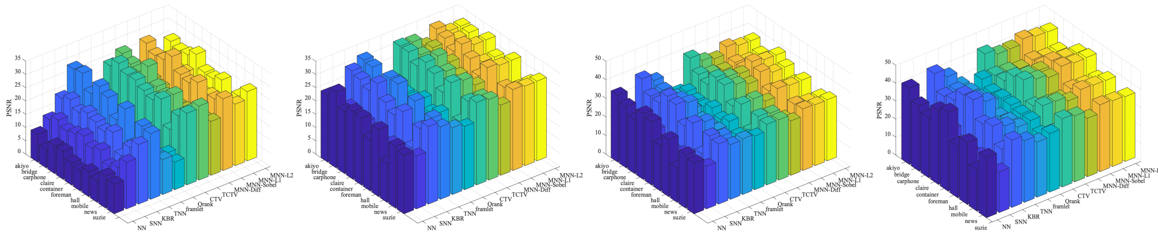
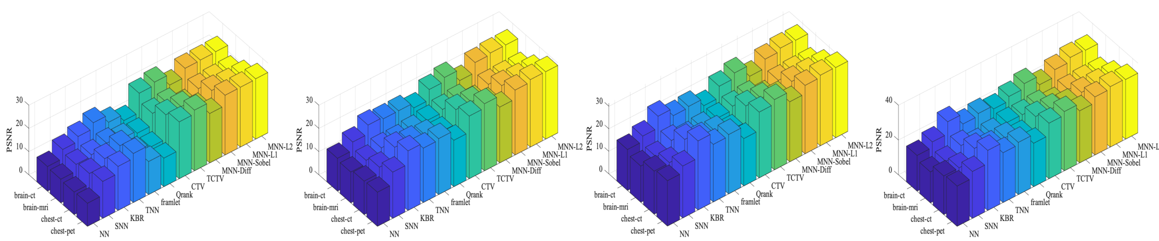


Figure 5: Performance comparison in terms of PSNR of recovered multispectral images obtained by all competing method under MC tasks.



MC: sr=0.02 MC: sr=0.05 MC: sr=0.1 MC: sr=0.2

Figure 6: Performance comparison in terms of PSNR of recovered color videos obtained by all competing method under MC tasks.



MC: sr=0.02 MC: sr=0.05 MC: sr=0.1 MC: sr=0.2

Figure 7: Performance comparison in terms of PSNR of recovered MRI and CT data obtained by all competing method under MC tasks.