



A robust low-rank matrix completion based on truncated nuclear norm and Lp-norm

Hao Liang¹ · Li Kang^{1,2} · Jianjun Huang¹

Accepted: 17 February 2022 / Published online: 13 March 2022

© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2022

Abstract

The low-rank matrix completion problem has aroused notable attention in various fields, such as engineering and applied sciences. The classical methods approximate the rank minimization problem by minimizing the nuclear norm, therefore obtaining unsatisfactory results, which may deviate from the true solution. In addition, most methods minimize the square error directly, which may be sensitive to the outliers. This paper presents a robust matrix completion model, which is suitable for a low sampling rate. First, the truncated nuclear norm is introduced, which is a more accurate and robust approximation to the rank function. Then, the Lp-norm may be employed as an error function, which provides a robust estimation. Finally, several optimization algorithms are employed to solve the model. Numerical simulations and experimental data analysis show the effectiveness and advantages of the proposed method. Notably, the algorithm can better approximate rank minimization problems and enhance robustness to outliers, especially when the sampling rate is very low. The method's practical potential is illustrated on the MovieLens-1M dataset.

Keywords Sparse representation · Low-rank matrix completion · Lp-norm · Truncated nuclear norm · Alternating direction multiplier method

✉ Li Kang
kangli@szu.edu.cn

Hao Liang
haoliang@stu.xmu.edu.cn

¹ College of Electronics and Information Engineering, Shenzhen University, Shenzhen 518060, China

² The Guangdong Key Laboratory of Intelligent Information Processing, Shenzhen 518060, China

1 Introduction

With the improvement of mathematical theory, the linear representation method has formed a nice theoretical framework, which has attracted widespread attention [1, 2]. Sparse representation method is one of the most representative method, which has been widely employed in various fields, such as signal processing, image processing, machine learning, and computer vision [3–5]. For many problems, such as image restoration, super-resolution, visual tracking, image classification and image segmentation, sparse representation methods also show its practical potential [6–9]. And low-rank matrix completion plays an importance role in sparse low-rank representation.

In recent years, estimating the missing values via the very limited information in an unknown matrix has aroused notable interest [10, 11]. Obviously, the completion of any matrix is an inappropriate problem, since the number of samples is limited, leading to being infinitely many cases for restorable matrices. Therefore, we need some other qualifying information to determine which of these candidates is more appropriate. In many cases, we hope that the restored matrix has low-rank or approximately low-rank features, which means it may be solved by a low-rank matrix completion method.

A typical low-rank matrix completion problem is the Netflix problem. From 2006 to 2009, Netflix held a data science competition to improve its proprietary movie recommendation system. The training set is a rating matrix containing approximately 480,000 users (i.e., rows) and 18,000 movies (i.e., columns), with a total of 8.6 billion potential users. Among them, only 1.2% of the scores can be observed, and the goal was to predict the remaining 98.8% of the blank items based on the existing data. How to use a customer's ratings of a limited number of movies to predict their preferences, and then recommend movies that match their preferences, which is equivalent to inferring unknown elements through the known elements in the matrix. Due to the limited factors which may affect customers' preferences, the matrix is essentially a low-rank matrix. Nowadays, matrix completion plays an important role in collaborative filtering, recommendation systems, data clustering, video denoising, network coding, medical imaging, and other fields, which has become one of the most significant and challenging tasks in computer vision and machine learning.

Candès and Recht pointed out that the matrix recovery is not as difficult to solve as people think [12]. They proved that even if the cardinality of the sample set is surprisingly small, most low-rank matrices can be accurately recovered from a small number of sample entries. More importantly, they proved that a low-rank matrix completion problem can be solved by minimizing the rank function.

However, due to the non-convexity and discontinuity of the rank function, the above-mentioned rank minimization problem is usually NP-hard, and existing algorithms cannot solve it directly and effectively. As is well known, the nuclear norm is the most common alternative to the rank function. Simultaneously, the value of the observed item may not have to be strictly accurate due to noise interference, therefore we can restore it by minimizing the prediction square error. But unfortunately, this kind of method may be very sensitive to outliers.

In this article, we propose a low-rank matrix completion model based on truncated nuclear norm and L_p -norm. Different from minimizing the sum of all singular values based on the nuclear norm method, a truncated nuclear norm is introduced, which only

minimizes $\min(m, n) - r$ singular values. In this way, a more accurate and robust approximation function to the rank function can be obtained. In addition, the L_p -norm is employed as the error function to improve the robustness. Therefore, our new goal is to minimize the joint truncated nuclear norm and L_p -norm. When p approaches 0, the proposed method is more robust and effective than the standard matrix completion method. Although our objective function is not a convex problem, an efficient solution based on the alternating direction multiplier method has been derived.

In the numerical simulation, experiments are carried out on both synthetic and real image data, and evaluations are performed by using relative error (RE) and peak signal-to-noise ratio (PSNR) indicators. All empirical results show that the presented model is better than several state-of-the-art matrix completion methods, especially at low sampling rates. The method efficiency is illustrated on the MovieLens-1M dataset.

All in all, the main contributions of this article are as follows:

1. A new objective function based on the truncated nuclear norm and L_p -norm is proposed for the low-rank matrix completion problem;
2. Though the optimization objective function is a non-convex and non-trivial problem, an efficient optimization scheme has been derived to solve this problem.

1.1 Outlines

The remainder of this paper is organized as follows: In the next section, the background of the matrix completion and some related definitions are provided. In Sect. 3, a robust matrix completion model is proposed based on the truncated nuclear norm and L_p -norm. An effective optimization scheme is designed in Sects. 4. Section 5 conducts the numerical experiments to compare the proposed method with existing methods. Finally, Sect. 6 concludes this paper.

1.2 Notations

The set R^n and $R^{m \times n}$ denote the space of n dimensional column vector and the space of $m \times n$ dimensional matrix, respectively. Lowercase bold letters indicate vectors, while uppercase bold letters indicate matrices, with the (i, j) element of the matrix X being expressed as X_{ij} . For a matrix X , $\text{rank}(X)$ and $\text{Tr}(X)$ denote the rank and the trace of X . The Frobenious norm of X is defined as $\|X\|_F = (\sum_i \sum_j X_{ij}^2)^{\frac{1}{2}}$. The nuclear norm of X is denoted as $\|X\|_* = \sum_{i=1}^{\min(m,n)} \sigma_i$, with σ_i denoting the i -th singular value.

2 Preliminary

2.1 Background

We first describe the low-rank matrix completion problem formally. Let

$$X_{\Omega} = \{X_{ij} | (i, j) \in \Omega\},$$

where X denotes a matrix, and Ω denotes the index of the observation matrix elements, i.e., $\Omega \subset \{1, \dots, m\} \times \{1, \dots, n\}$. Assuming that we have obtained some observations, D_{Ω} , in the low-rank matrix, X , the task of matrix completion is to restore the unobserved elements in X . Specifically, the completion problem of X may be described by solving the following rank minimization problem as

$$\begin{aligned} & \min_X \text{rank}(X), \\ & \text{s.t. } X_{\Omega} = D_{\Omega}. \end{aligned} \tag{1}$$

However, in most cases, due to the non-convexity and discontinuity of the rank function, the problem (1) is NP-hard [13], which means it is difficult to be solved directly and effectively. Theoretical studies show that the nuclear norm, $\|X\|_*$, i.e., the sum of singular values, is the tightest convex lower bound of the rank function. Therefore, the nuclear norm is commonly used as the convex replacement of the matrix rank function [14]. Thus, problem (1) may be expressed approximately by the nuclear norm as

$$\begin{aligned} & \min_X \|X\|_*, \\ & \text{s.t. } X_{\Omega} = D_{\Omega}. \end{aligned} \tag{2}$$

Estimating the low-rank matrix will inevitably be disturbed by noise, such that the recovered matrix does not have to perfectly match the corresponding items of the observation items. Thus, we relax the constraints of the equation, and problem (2) may be described as [15]

$$\min_X \|X\|_* + \gamma \|X_{\Omega} - D_{\Omega}\|_2^2, \tag{3}$$

where γ denotes the penalty parameter. Many researchers have strived to find the optimal solution of this problem, deriving a series of related approaches, such as the singular value threshold [16] and singular value projection algorithm [17]. In addition, a variety of objective function have been studied, such as the truncated nuclear norm [18], Schatten p -norm [19], and other variants [20–22], which may lead to a more accurate estimation.

2.2 Definition of Lp-Norm

Definition 1 Given a matrix $X \in R^{m \times n}$, the Lp-norm of X is defined as [19]

$$\|X\|_p = \left(\sum_{i=1}^n \sum_{j=1}^m |X_{ij}|^p \right)^{\frac{1}{p}}. \tag{4}$$

2.3 Definition of truncated nuclear norm

Definition 2 Given a matrix $X \in R^{m \times n}$, its truncated nuclear norm $\|X\|_r$ is defined as the sum of $\min(m, n) - r$ smallest singular values, which may be expressed as [18]

$$\|X\|_r = \sum_{i=r+1}^{\min(m,n)} \sigma_i, \quad (5)$$

where σ_i denotes the i -th singular value of X .

3 The proposed method

Note that L $_p$ -norm can be regarded as an extension of L2-norm. We, therefore, use L $_p$ -norm ($0 < p \leq 1$) to replace L2-norm in problem (3). Employing L $_p$ -norm as an error function can effectively enhance the robustness to the outliers [23]. Then, we can transform the problem (3) into the following form as

$$\min_X \|X\|_* + \gamma \|X_\Omega - D_\Omega\|_p^p. \quad (6)$$

Simultaneously, noting that the value of the largest r non-zero singular values will not affect the rank of a matrix, we do not constrain them in the newly designed truncated nuclear norm, and focus on minimizing the remaining $\min(m, n) - r$ smallest singular values, i.e., convert $\|X\|_*$ into the optimization of $\|X\|_r$, such that problem (6) may be transformed into

$$\min_X \|X\|_r + \gamma \|X_\Omega - D_\Omega\|_p^p. \quad (7)$$

Since $\|X\|_r$ is non-convex, it is obviously difficult to solve problem (7) directly. To handle this issue, the following lemma from [18] is given without proof:

Lemma 1 Given a matrix $X \in R^{m \times n}$, for any matrix $A \in R^{r \times m}$ and $B \in R^{r \times n}$ that satisfy $AA^T = I_{r \times r}$ and $BB^T = I_{r \times r}$, where $I_{r \times r}$ denotes a r -order identity matrix. Thus, for any non-negative integer r satisfying $r \leq \min(m, n)$, it holds [18]

$$\|X\|_r = \|X\|_* - \max_{AA^T=I_{r \times r}, BB^T=I_{r \times r}} \text{Tr}(AXB^T). \quad (8)$$

In summary, combining the truncated nuclear norm and L $_p$ -norm, we may rewrite the optimization problem (7) as

$$\min_X \|X\|_* - \max_{AA^T=I_{r \times r}, BB^T=I_{r \times r}} \text{Tr}(AXB^T) + \gamma \|X_\Omega - D_\Omega\|_p^p, \quad (9)$$

with $A \in R^{r \times m}$ and $B \in R^{r \times n}$. This is our proposed robust model combining L $_p$ -norm and the truncated nuclear norm (L $_p$ -TNN). The solution scheme of model (9) will be detailed in Sect. 4.

4 Efficient implementation

4.1 Three-steps method to solve problem (9)

For problem (9), we have to solve the problem $\max_{AA^T=I_{r \times r}, BB^T=I_{r \times r}} \text{Tr}(AXB^T)$ first. Inspired by [18], a simple and effective three-step iterative method has been established here. Specifically, we first let $X^{(0)} = D_\Omega$ as the initial value. Then, in the k -th iteration, when $X^{(k)}$ is obtained, $A^{(k)}$ and $B^{(k)}$ may be obtained by the singular value decomposition (SVD) procedure, i.e., $X^{(k)} = U^{(k)} \Sigma^{(k)} V^{(k)T}$, where

$$A^{(k)} = (u_1, \dots, u_r)^T, B^{(k)} = (v_1, \dots, v_r)^T, \tag{10}$$

with $U^{(k)} = (u_1, \dots, u_m) \in R^{m \times m}$ and $V^{(k)} = (v_1, \dots, v_n) \in R^{n \times n}$. Then, $X^{(k+1)}$ may be updated by solving the following problem

$$\min_X \|X\|_* - \text{Tr}(A^{(k)} X B^{(k)T}) + \gamma \|X_\Omega - D_\Omega\|_p^p. \tag{11}$$

In this way, the maximization problem in minimization can be converted into a three-steps iterative process. This process is summarized in Algorithm 1, and the stopping criterion will be given in Sect. 4.7.

Algorithm 1 Three-steps to solve the problem (9)

Input: Original incomplete data matrix, D_Ω , stopping threshold, ϵ_0 , iteration number, iter, the number of truncate singular values, r , and $X^{(0)} = D_\Omega$.

repeat

step 1. singular value decomposition of $X^{(k)}$

$$[U^{(k)}, \Sigma^{(k)}, V^{(k)T}] = \text{svd}(X^{(k)}),$$

where

$$U^{(k)} = (u_1, \dots, u_m) \in R^{m \times m}, V^{(k)} = (v_1, \dots, v_n) \in R^{n \times n}.$$

step 2. calculate $A^{(k)}$ and $B^{(k)}$

$$A^{(k)} = (u_1, \dots, u_r)^T, B^{(k)} = (v_1, \dots, v_r)^T.$$

step 3. calculate $X^{(k+1)}$

$$X^{(k+1)} = \arg \min_X \|X\|_* - \text{Tr}(A^{(k)} X B^{(k)T}) + \gamma \|X_\Omega - D_\Omega\|_p^p.$$

until reaches the stopping criterion or the maximum number of iterations.

Output: The recovered matrix X^* .

As can be seen in Algorithm 1, how to effectively solve step 3 is extremely significant. Noting that it is still a complex optimization problem, we restate and simplify it as follows

$$\min_X \|X\|_* - \text{Tr}(AXB^T) + \gamma \|X_\Omega - D_\Omega\|_p^p, \tag{12}$$

where \mathbf{A} and \mathbf{B} are known matrices. It should be noted that $\|\mathbf{X}\|_*$ and $\text{Tr}(\mathbf{A}\mathbf{X}\mathbf{B}^T)$ are convex functions, and $\|\mathbf{X}_\Omega - \mathbf{D}_\Omega\|_p^p$ is a concave function, therefore the Augmented Lagrangian Method (ALM) and Alternating Direction Multiplier Method (ADMM) can be used to optimization. We will describe the solution to problem (12) in Sect. 4.2.

4.2 Augmented lagrangian method to solve problem (12)

Here, we add some equivalent constraints to (12), such that it may be expressed as

$$\min_{\mathbf{X}, \mathbf{E}_\Omega = \mathbf{X}_\Omega - \mathbf{D}_\Omega, \mathbf{X} = \mathbf{W}} \|\mathbf{X}\|_* - \text{Tr}(\mathbf{A}\mathbf{W}\mathbf{B}^T) + \gamma \|\mathbf{E}_\Omega\|_p^p. \tag{13}$$

The augmented lagrangian function corresponding to the above problem (13) can be expressed as

$$\begin{aligned} L(\mathbf{X}, \mu, \mathbf{A}, \mathbf{\Sigma}, \mathbf{W}, \mathbf{E}_\Omega) = & \|\mathbf{X}\|_* - \text{Tr}(\mathbf{A}\mathbf{W}\mathbf{B}^T) + \gamma \|\mathbf{E}_\Omega\|_p^p \\ & + \frac{\mu}{2} \left\| \mathbf{E}_\Omega - (\mathbf{X}_\Omega - \mathbf{D}_\Omega) + \frac{1}{\mu} \mathbf{A} \right\|_F^2 \\ & + \frac{\mu}{2} \left\| \mathbf{X} - \mathbf{W} + \frac{1}{\mu} \mathbf{\Sigma} \right\|_F^2, \end{aligned} \tag{14}$$

where $\mu > 0$ denotes the penalty parameter, with \mathbf{A} and $\mathbf{\Sigma}$ denoting the scaled dual variables.

In the interest of brevity, we denote the objective function as $f(\mathbf{X})$, and the two constraints as $h(\mathbf{X}) = 0$ (i.e., $h(\mathbf{X}) = \mathbf{E}_\Omega - (\mathbf{X}_\Omega - \mathbf{D}_\Omega)$) and $g(\mathbf{X}) = 0$ (i.e., $g(\mathbf{X}) = \mathbf{X} - \mathbf{W}$). Then, ALM can be used to solve the problem (13), and we summarize the mainly procedure in Algorithm 2.

Algorithm 2 ALM to solve the problem (13)

Input: Constant $\rho > 0$, $\mu > 0$, \mathbf{A} , and $\mathbf{\Sigma}$.

repeat

step 1. update \mathbf{X}

$$\mathbf{X} = \arg \min_{\mathbf{X}} f(\mathbf{X}) + \frac{\mu}{2} \left\| h(\mathbf{X}) + \frac{1}{\mu} \mathbf{A} \right\|_F^2 + \frac{\mu}{2} \left\| g(\mathbf{X}) + \frac{1}{\mu} \mathbf{\Sigma} \right\|_F^2.$$

step 2. update \mathbf{A}

$$\mathbf{A} = \mathbf{A} + \mu h(\mathbf{X}).$$

step 3. update $\mathbf{\Sigma}$

$$\mathbf{\Sigma} = \mathbf{\Sigma} + \mu g(\mathbf{X}).$$

step 4. update μ

$$\mu = \rho \mu.$$

until algorithm convergence.

Output: The optimal solution \mathbf{X}^* .

As may be seen in Algorithm 2, step 2, step 3, and step 4 are relatively easy to solve. The main difficulty lies in the solution in step 1. The optimization problem that needs to be solved in step 1 can be expressed as

$$\begin{aligned}
 & \min_{X, E_\Omega, W} \|X\|_* - \text{Tr}(A W B^T) + \gamma \|E_\Omega\|_p^p \\
 & + \frac{\mu}{2} \left\| E_\Omega - (X_\Omega - D_\Omega) + \frac{1}{\mu} A \right\|_F^2 \\
 & + \frac{\mu}{2} \left\| X - W + \frac{1}{\mu} \Sigma \right\|_F^2.
 \end{aligned} \tag{15}$$

It is extremely difficult to accurately minimize X, E_Ω, W in problem (15) simultaneously, and the required algorithm’s complexity may be very high. Thus, the alternating direction multiplier method (ADMM) [24] may be employed to simplify the problem (15), and we will explain it in detail in Sect. 4.3.

4.3 Alternating direction multiplier method to solve problem (15)

As mentioned above, for the problem (15), it is very difficult to optimize three variables at the same time, so we employ ADMM to convert this complex problem into several sub-problems. Specifically, we may optimize the three variables separately. When optimizing for a certain variable, the other two variables will be fixed, leading to the following three sub-problems.

- Sub-problem 1: When optimizing X , we fix E_Ω and W . Then, problem (15) may be simplified as

$$\min_X \|X\|_* + \frac{\mu}{2} \left\| E_\Omega - (X_\Omega - D_\Omega) + \frac{1}{\mu} A \right\|_F^2 + \frac{\mu}{2} \left\| X - W + \frac{1}{\mu} \Sigma \right\|_F^2. \tag{16}$$

- Sub-problem 2: When optimizing E_Ω , we fix X and W . Then, problem (15) may be simplified as

$$\min_{E_\Omega} \gamma \|E_\Omega\|_p^p + \frac{\mu}{2} \left\| E_\Omega - (X_\Omega - D_\Omega) + \frac{1}{\mu} A \right\|_F^2. \tag{17}$$

- Sub-problems 3: When optimizing W , we fix E_Ω and X . Then, problem (15) may be simplified as

$$\min_W -\text{Tr}(A W B^T) + \frac{\mu}{2} \left\| X - W + \frac{1}{\mu} \Sigma \right\|_F^2. \tag{18}$$

In this way, we can transform the complex problem (15) into three simpler sub-problems. Specific solutions to these sub-problems are given in Sects. 4.4, 4.5, and 4.6, respectively.

4.4 Solution of sub-problem (16)

We first introduce the definition of singular value threshold operator.

Definition 3 Consider the SVD of a matrix $X \in R^{m \times n}$ of rank r

$$X = U \Sigma V^T, \Sigma = \text{diag}(\{\sigma_i\}_{1 \leq i \leq r}), \quad (19)$$

where $U \in R^{m \times r}$ and $V \in R^{n \times r}$ denote unit orthogonal matrices, and the i -th singular values σ_i are all positive numbers. For given $\tau \geq 0$, we define the singular value threshold operator \mathcal{D}_τ as [16]

$$\mathcal{D}_\tau(X) = U \mathcal{D}_\tau(\Sigma) V^T, \mathcal{D}_\tau(\Sigma) = \text{diag}(\max\{\sigma_i - \tau, 0\}).$$

Immediately, the following lemma from [16] is given without proof:

Lemma 2 For every $\tau \geq 0$ and a given matrix $Y \in R^{m \times n}$, the singular value threshold operator defined by Definition 3 always satisfies [16]

$$\mathcal{D}_\tau(Y) = \arg \min_X \left\{ \frac{1}{2} \|X - Y\|_F^2 + \tau \|X\|_* \right\}. \quad (20)$$

Then, the problem (16) may be simplified as

$$\min_X \|X\|_* + \frac{\mu}{2} \|X_\Omega - M_\Omega\|_F^2 + \frac{\mu}{2} \|X - N\|_F^2, \quad (21)$$

where $M_\Omega = E_\Omega + D_\Omega + \frac{1}{\mu} A$, $N = W + \frac{1}{\mu} \Sigma$.

Here, we consider another optimization problem first

$$\min_X \|X_\Omega - M_\Omega\|_F^2 + \|X - N\|_F^2. \quad (22)$$

Let $X_{\bar{\Omega}} = \{X_{ij} | (i, j) \notin \Omega\}$. Then, the optimal solution to problem (22), i.e., \hat{X} , can be expressed as

$$\hat{X}_\Omega = \frac{M_\Omega + N_\Omega}{2}, \hat{X}_{\bar{\Omega}} = N_{\bar{\Omega}}. \quad (23)$$

Exploiting the relationship between problems (21) and (22), the optimal solution of problem (21) may be obtained as $\mathcal{D}_{\frac{1}{\mu}}(\hat{X})$ by using Lemma 2. In this way, the iterative process of sub-problem (16) may be solved. The k -th iteration process of X can be expressed as

$$X^{(k+1)} = \mathcal{D}_{\frac{1}{\mu}}(\hat{X}^{(k)}), \quad (24)$$

where $\hat{X}^{(k)}$ can be obtained from $E_\Omega^{(k)}$, $D_\Omega^{(k)}$, $A^{(k)}$, and $\Sigma^{(k)}$ (see equation (23)).

4.5 Solution to sub-problem (17)

Problem (17) can be simplified as

$$\min_{E_{\Omega}} \frac{\gamma}{\mu} \|E_{\Omega}\|_p^p + \frac{1}{2} \|E_{\Omega} - C_{\Omega}\|_F^2, \tag{25}$$

where $C_{\Omega} = X_{\Omega} - D_{\Omega} - \frac{1}{\mu} \Lambda$. Since each element $\{E_{ij}|(i,j) \in \Omega\}$ of the matrix E_{Ω} in the above problem can be separated and solved separately, one may, reminiscent to [19], get the optimized problem for each element

$$\min_{E_{ij}} \lambda |E_{ij}|^p + \frac{1}{2} |E_{ij} - C_{ij}|^2, \tag{26}$$

where $\{C_{ij}|(i,j) \in \Omega\}$ denotes the element of the matrix C_{Ω} and $\lambda = \frac{\gamma}{\mu} (\lambda 0)$.

In the interest of brevity, the problem may be rewritten as the following equivalent form

$$\min_x \psi(x) = \lambda |x|^p + \frac{1}{2} |x - a|^2. \tag{27}$$

It is easy to know that the objective function $\psi(x)$ is differentiable when $x \neq 0$, and this point is a discontinuous point of the derivative. The symbolic function $\text{sgn}(x)$ is introduced to represent the first derivative of $\psi(x)$ when $x \neq 0$, being expressed as

$$\text{sgn}(x) = \begin{cases} -1, & x < 0, \\ 1, & x > 0. \end{cases} \tag{28}$$

Then, when $x \neq 0$, the first order derivative of $\psi(x)$ is

$$\psi'(x) = x - a + \lambda p |x|^{p-1} \text{sgn}(x). \tag{29}$$

Similarly, we can obtain the second and third order derivative of $\psi(x)$ when $x \neq 0$ as

$$\psi''(x) = 1 - \lambda p(1 - p) |x|^{p-2}, \tag{30}$$

$$\psi^{(3)}(x) = \lambda p(1 - p)(2 - p) |x|^{p-3} \text{sgn}(x). \tag{31}$$

According to the expression of $\psi^{(3)}(x)$, we can know that when $x > 0$, $\psi^{(3)}(x) > 0$, which means $\psi''(x)$ strictly increases monotonically. Also, when $x < 0$, $\psi^{(3)}(x) < 0$, which means $\psi''(x)$ is strictly monotonously decreasing. Let $\psi''(x) = 0$, we can find a constant x_0 as follows

$$x_0 = (\lambda p(1 - p))^{1/(2-p)}. \tag{32}$$

We summarize the nature of this function in Table 1. Note that $\psi(0) = 0$ holds, we may further discuss the objective function $\psi(x)$ in different situations.

– Case 1: $\psi'(-x_0) \leq 0, \psi'(x_0) \geq 0$.

In this case, when $x < 0$, $\psi'(x) < \psi'(-x_0) \leq 0$, which means $\psi(x)$ is strictly monotonously decreasing. And when $x > 0$, $\psi'(x) > \psi'(x_0) \geq 0$, meaning $\psi(x)$ is

strictly monotonously increasing. Thus, the optimal solution of problem (27) in Case 1 is $x^* = 0$.

- Case 2: $\psi'(-x_0) \leq 0, \psi'(x_0) < 0$.

In this case, when $x < 0$, $\psi'(x) < \psi'(-x_0) \leq 0$, which means $\psi(x)$ is strictly monotonously decreasing. Since $\psi'(x_0) < 0$, we suppose that when $x > 0$, $\psi'(x) = 0$ has two roots x_1, x_2 ($x_1 < x_2$). We use Table 2 to more intuitively show the situation of $\psi(x)$ when $x > 0$.

As can be seen, when $x > 0$, $\psi(x) \geq \psi(x_2)$, that is, $\psi(x_2)$ is the minimum value when $x > 0$. And x_2 is the root of $\psi'(x) = 0$ greater than x_0 , which becomes a suspected optimal solution of problem (27). Newton's method [25] can be used to iteratively solve the equation $\psi'(x) = 0$ to obtain x_2 . We can set the initial value to $2x_0$ to ensure that the algorithm accurately converge to x_2 .

Considering all situation comprehensively, the optimal solution to problem (27) in Case 2 may be expressed as

$$x^* = \arg \min_{x \in \{0, x_2\}} \psi(x). \quad (33)$$

- Case 3: $\psi'(-x_0) > 0, \psi'(x_0) \geq 0$.

We follow the ideas of Case 2 to analyze. In this case, $\psi(x)$ is strictly monotonously increasing when $x > 0$, $\psi'(x) \geq 0$. Since $\psi'(-x_0) > 0$, we suppose that $\psi'(x) = 0$ has two roots x_3, x_4 ($x_3 < x_4$) when $x < 0$. We use Table 3 to more intuitively show the situation of $\psi(x)$ when $x < 0$.

As can be seen, when $x < 0$, $\psi(x) \geq \psi(x_3)$, that is, $\psi(x_3)$ is the minimum value when $x < 0$. And x_3 is the root of $\psi'(x) = 0$ less than $-x_0$, which is a suspected optimal solution of problem (27). We can also use Newton's method to iteratively solve the equation $\psi'(x) = 0$ to obtain x_3 . The initial value is recommended to be set to $-2x_0$, ensuring that the algorithm accurately converge to x_3 .

Considering all situation comprehensively, the optimal solution of problem (27) under this condition may be expressed as

$$x^* = \arg \min_{x \in \{0, x_3\}} \psi(x). \quad (34)$$

- Case 4: $\psi'(-x_0) > 0, \psi'(x_0) < 0$.

This situation is the combination of Case 2 and Case 3. When $\psi'(x_0) < 0$, according to the analysis result of Case 2, we obtain the minimum value $\psi(x_2)$ when $x > 0$. When $\psi'(-x_0) > 0$, according to the analysis result of Case 3, we obtained the minimum value $\psi(x_3)$ when $x < 0$. Among them, the method of finding x_2 and x_3 is consistent with the above two cases. Then, all suspected optimal solutions are comprehensively considered, the optimal solution to an optimization problem (27) under this condition can be expressed as

Table 1 Analysis of $\psi(x)$

	$< -x_0$	$-x_0$	$(-x_0, 0)$	$(0, x_0)$	x_0	x_0
$\psi^{(3)}(x)$	< 0			0		
$\psi''(x)$	0	0	< 0	< 0	0	0
$\psi'(x)$	\nearrow	\max	\searrow	\searrow	\min	\nearrow

Table 2 Analysis of $\psi(x)$ in Case 2

	$(0, x_1)$	x_1	(x_1, x_0)	x_0	(x_0, x_2)	x_2	x_2
$\psi''(x)$	< 0			0	0		
$\psi'(x)$	0	0	< 0	< 0	< 0	0	0
$\psi(x)$	\nearrow	\max	\searrow	\searrow	\searrow	\min	\nearrow

$$x^* = \arg \min_{x \in \{0, x_2, x_3\}} \psi(x). \tag{35}$$

All in all, we summarize all the cases of the optimization problem (27), such that its solution may be expressed as

$$\begin{cases} \psi'(-x_0) \leq 0, \psi'(x_0) \geq 0, x^* = 0, \\ \psi'(-x_0) \leq 0, \psi'(x_0) < 0, x^* = \arg \min_{x \in \{0, x_2\}} \psi(x), \\ \psi'(-x_0) > 0, \psi'(x_0) \geq 0, x^* = \arg \min_{x \in \{0, x_3\}} \psi(x), \\ \psi'(-x_0) > 0, \psi'(x_0) < 0, x^* = \arg \min_{x \in \{0, x_2, x_3\}} \psi(x). \end{cases} \tag{36}$$

In this way, sub-problem (17) may also be solved successfully.

4.6 Solution to sub-problem (18)

Exploiting the properties of trace, i.e., $\text{Tr}(A) = \text{Tr}(A^T)$ and $\text{Tr}(ABC) = \text{Tr}(BCA)$, problem (18) may be equivalently converted to the following problem as

$$\min_W -\mu \text{Tr}\left(\frac{1}{\mu} A^T B W^T\right) + \frac{\mu}{2} \left\| W - \left(X + \frac{1}{\mu} \Sigma\right) \right\|_F^2. \tag{37}$$

Adding a constant term, i.e., $\frac{\mu}{2} \left\| \frac{1}{\mu} A^T B \right\|_F^2$, the problem (37) may be rewritten as

$$\min_W -\mu \text{Tr}\left(\frac{1}{\mu} A^T B W^T\right) + \frac{\mu}{2} \left\| W - \left(X + \frac{1}{\mu} \Sigma\right) \right\|_F^2 + \frac{\mu}{2} \left\| \frac{1}{\mu} A^T B \right\|_F^2. \tag{38}$$

Then, combining the relationship between the F-norm and trace, the above formula may be simplified as

$$\min_{\mathbf{W}} \frac{\mu}{2} \left\| \mathbf{W} - \left(\mathbf{X} + \frac{1}{\mu} \boldsymbol{\Sigma} + \frac{1}{\mu} \mathbf{A}^T \mathbf{B} \right) \right\|_F^2, \tag{39}$$

leading to a very straightforward optimal solution, i.e.,

$$\mathbf{W}^* = \mathbf{X} + \frac{1}{\mu} \boldsymbol{\Sigma} + \frac{1}{\mu} \mathbf{A}^T \mathbf{B}. \tag{40}$$

Then, we can obtain the iterative solution to problem (18). In the k -th iteration, the corresponding elements of the matrix $\mathbf{W}^{(k+1)}$ in set Ω are kept consistent with $\mathbf{W}^{(k)}$, yielding

$$\widehat{\mathbf{W}}^{(k+1)} = \mathbf{X}^{(k+1)} + \frac{1}{\mu} \boldsymbol{\Sigma}^{(k)} - \frac{1}{\mu} \mathbf{A}^{(k+1)T} \mathbf{B}^{(k+1)}, \tag{41}$$

$$\mathbf{W}^{(k+1)} = (\widehat{\mathbf{W}}^{(k+1)})_{\overline{\Omega}} + (\mathbf{W}^{(k)})_{\Omega}. \tag{42}$$

4.7 Algorithm summary

In our proposed algorithm, the stopping criterion is set to

$$\frac{\left\| \mathbf{X}^{(k+1)} - \mathbf{X}^{(k)} \right\|_F}{\max(\left\| \mathbf{X}^{(k+1)} \right\|_F, 1)} \leq tol, \tag{43}$$

where tol is the stopping threshold. It is suggested that the stopping threshold, tol , is selected from 10^{-4} to 10^{-7} . Based on the analysis in Sect. 4, we summarize the algorithm for the proposed Lp-TNN model in Algorithm 3. The MATLAB code for this algorithm is available online at <https://github.com/HauLiang/Lp-TNN>.

Table 3 Analysis of $\psi(x)$ in Case 3

	$< x_3$	x_3	$(x_3, -x_0)$	$-x_0$	$(-x_0, x_4)$	x_4	$(x_4, 0)$
$\psi''(x)$	0			0	< 0		
$\psi'(x)$	< 0	0	0	0	0	0	< 0
$\psi(x)$	\searrow	min	\nearrow	\nearrow	\nearrow	max	\searrow

Algorithm 3 The algorithm for Lp-TNN

Input: Step constant, $\rho > 0$, original incomplete data matrix, D_Ω , stopping threshold, tol , iteration number, n , the number of truncate singular values, r , constant, $\mu > 0$, constant, γ , p-value, p , matrix, $E_\Omega^{(0)}$, matrix, $A^{(0)}$, matrix, $\Sigma^{(0)}$, matrix, $X^{(0)} = D_\Omega$, and $W^{(0)} = D_\Omega$.

repeat

step 1. SVD of $X^{(k)}$

$$[U^{(k)}, \Sigma^{(k)}, V^{(k)T}] = \text{svd}(X^{(k)}).$$

where

$$U^{(k)} = (u_1^{(k)}, \dots, u_m^{(k)}) \in R^{m \times m}, V^{(k)} = (v_1^{(k)}, \dots, v_n^{(k)}) \in R^{n \times n}.$$

step 2. update $A^{(k+1)}$ and $B^{(k+1)}$

$$A^{(k+1)} = (u_1^{(k)}, \dots, u_r^{(k)})^T, B^{(k+1)} = (v_1^{(k)}, \dots, v_r^{(k)})^T.$$

step 3. update $X^{(k+1)}$

$$X^{(k+1)} = \mathcal{D}_{\frac{1}{\mu}}(\widehat{X}^{(k)}),$$

where

$$\widehat{X}_\Omega^{(k)} = \frac{M_\Omega^{(k)} + N_\Omega^{(k)}}{2}, \widehat{X}_{\Omega^c}^{(k)} = N_{\Omega^c}^{(k)},$$

with

$$M_\Omega^{(k)} = E_\Omega^{(k)} + D_\Omega^{(k)} + \frac{1}{\mu} A^{(k)}, N^{(k)} = W^{(k)} + \frac{1}{\mu} \Sigma^{(k)}.$$

step 4. update $E_\Omega^{(k+1)}$

$$\forall (i, j), E_{ij}^{(k+1)} = \arg \min_{E_{ij}} \frac{\gamma}{\mu} |E_{ij}|^p + \frac{1}{2} |E_{ij} - C_{ij}^{(k)}|^2,$$

where

$$C_\Omega^{(k)} = X_\Omega^{(k+1)} - D_\Omega^{(k)} - \frac{1}{\mu} A^{(k)}.$$

step 5. update $W^{(k+1)}$

$$W^{(k+1)} = (\widehat{W}^{(k+1)})_{\Omega} + (W^{(k)})_{\Omega^c}.$$

step 6. update $A^{(k+1)}$

$$A^{(k+1)} = A^{(k)} + \mu (E_\Omega^{(k+1)} - X^{(k+1)} + D_\Omega).$$

step 7. update $\Sigma^{(k+1)}$

$$\Sigma^{(k+1)} = \Sigma^{(k)} + \mu (X^{(k+1)} - W^{(k+1)}).$$

step 8. update μ

$$\mu = \rho \mu.$$

until

$$\frac{\|X^{(k+1)} - X^{(k)}\|_F}{\max(\|X^{(k+1)}\|_F, 1)} \leq tol \quad \text{or} \quad \text{iteration} \geq n.$$

Output: The optimal solution X^* .

4.8 Convergence analysis and computational cost

In this section, we first discuss the convergence property of the proposed Lp-TNN. In the three-steps procedure, similar to [18], this iterative scheme can be proved to

converge to a local minimum. Then, the ADMM is employed to convert a complex problem into several sub-problems, which ensures the convergence of the algorithm (see e.g., [26]). For the three sub-problems, the convergence of sub-problem 1 is ensured by singular value threshold operator [16], the convergence of sub-problem 2 is ensured by the Newton's method (see e.g., [27, 28]), while sub-problem 3 has a closed-form solution.

Then, we analyze the complexity of the proposed algorithm. From Algorithm 3, we can know that the main per-iteration computation cost lies in the SVD operation. For a large-scale problem, we may employ some existing technologies, such as [29], to accelerate the calculation of SVD and make the algorithm more effective.

5 Numerical results

In this section, we compare the proposed method with the classic matrix completion methods on synthetic data and real-image data. The selection of parameters for the comparison approaches depend on the suggestions in their papers or the default parameters of the published code to obtain the best performance. The proposed Lp-TNN is compared with the following approaches:

- SVT algorithm [16], the singular value threshold algorithm.
- SVP algorithm [17], the singular value projection algorithm.
- TNNR algorithm [18], the truncated nuclear norm regularization algorithm.
- Sp-lp algorithm [19], the Schatten p -norm and ℓ_p -norm algorithm.
- FGSR algorithm [22], the factor group-sparse regularization algorithm.

5.1 Parameter setting

Firstly, we set the parameter p . We find that the performance of matrix recovery may be slightly worse when the value of p increases. This may be because when p tends to 1, the enhancement in robustness to outliers becomes smaller [30]. Simultaneously, we find that a smaller p value may lead to a longer running time. For simplicity, p -value is set to 0.2. For other parameters, the regularization parameter, γ , is set to 1.1, the step size constant, ρ , to 1.3, the update proportional constant, μ , to 0.1, the maximum number of iterations, n , to 1000, and the algorithm threshold, tol , to 10^{-4} , to make the algorithm have better performance. As for the setting of r , we test it from 0 to 20 and choose the best result as the final recovered matrix. If there are no special instructions, the parameters of the subsequent experiments are set to the same.

5.2 Synthetic data

Consider the randomly generated $n \times n$ matrix completion problem with rank r . We first randomly generate two matrices, X_1 and X_2 , where $X_1 \in R^{n \times r}$ and $X_2 \in R^{n \times r}$. Thus, the original matrix X can be generated by $X = X_1 X_2^T$. For the original matrix

X , we set different sampling rates, different ranks, and add different signal-to-noise ratios (SNRs) of noise to evaluate the performance of the algorithms in various situations. The relative error (RE) can be used to evaluate the performance of the algorithm, which may be defined as [31]

$$RE = \frac{\|X^* - X\|_F}{\|X\|_F}, \quad (44)$$

where X and X^* denote the original matrix and recovered matrix, respectively. The smaller the RE , the better the performance.

First, we discuss the impact of different SNRs of noise on model performance. The size of the matrix is set to 300×300 , and the rank is set to 10. The Gaussian white noise is added with different SNRs (dB) under different sampling rates. Note that the experiment is repeated 50 times and takes the average value. The REs and running time of the algorithms are recorded, and the results are shown in Figs. 1 and 2, respectively.

Then, we discuss the impact of different matrix ranks on model performance. The matrix size is also set to 300×300 and 0.1 dB Gaussian white noise is added. We set different ranks of the original matrix and take the average value after repeating 50 times under different sampling rates. We observe the REs and running time of the algorithms, and the results are shown in Figs. 3 and 4, respectively.

As can be seen in Figs. 1 and 2, the proposed model can maintain a small RE at a lower sampling rate and restore the original matrix with less running time and extremely high recovery efficiency. Figs. 3 and 4 also verify this. Simultaneously, under the influence of various SNR, the proposed method still restore the matrix stably and efficiently, which shows that it provides a robust estimation. In addition, when the rank of the matrix varies, the model can recover the matrix efficiently, showing its strong stability and adaptability.

From the above synthetic data experiments, it may be seen that for completely randomly generated matrix data, our model can perform the task of matrix completion well without causing failure to converge, which also verifies its good convergence.

5.3 Real image data

In most cases, an image can be viewed as an approximate low-rank matrix. Sometimes, images may be partially damaged due to encoding and transmission problems. Thus, the matrix completion algorithm may be employed to these images to recover the lost information. Here, we choose an RGB image and sample it under different sampling rates to obtain an image with missing information. The peak signal-to-noise ratio (PSNR) of the restored image is employed to evaluate the performance of different methods, which can be expressed as

$$PSNR = 10 \log_{10} \left(\frac{255^2}{\frac{1}{3m_1m_2} \sum_{i=1}^3 \|X_i^* - X_i\|_F^2} \right), \quad (45)$$

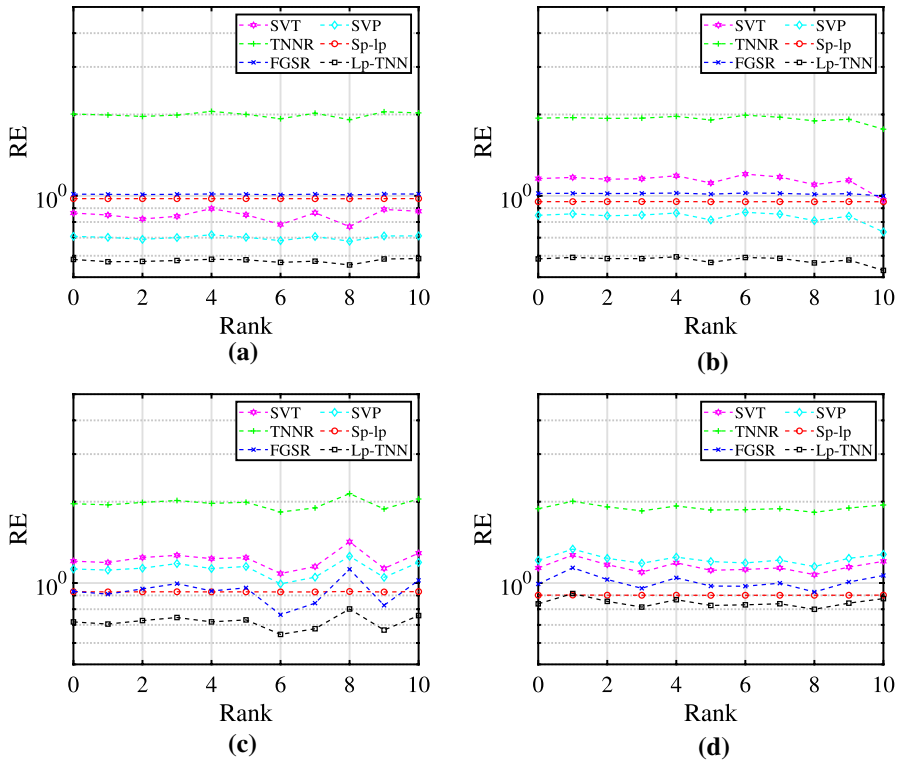


Fig. 1 The REs versus the SNRs using the synthetic data under different sampling rates: a 0.05, b 0.10, c 0.15, and d 0.20

where X_i and X_i^* denote the original and recovered image matrix of the i -th RGB channel (size: (m_1, m_2)). The larger the PSNR, the better the performance.

First, we show the restored pictures by different methods at a sampling rate of 0.4 in Fig. 5. As may be seen, our model may more clearly recover the picture.

Then, we exploit the sampling rate as an independent variable to calculate the PSNRs and running time under different sampling rates. The results are shown in Fig. 6.

Figure 6 shows that the proposed method can achieve the highest PSNR under various sampling rates, and it has a higher operating efficiency when the sampling rate is low. This may be attributed to the introduction of Lp-norm since it is robust to the outliers, and when the sampling rate is low, outlier points are more likely to be generated.

Simultaneously, it can be seen that as the sampling rate increases, the PSNR of each algorithm increases, which shows the rationality of the experiment. Because as the sampling rate increases, more information about the image can be achieved, which will reduce the difficulty of image restoration and produce a higher PSNR.

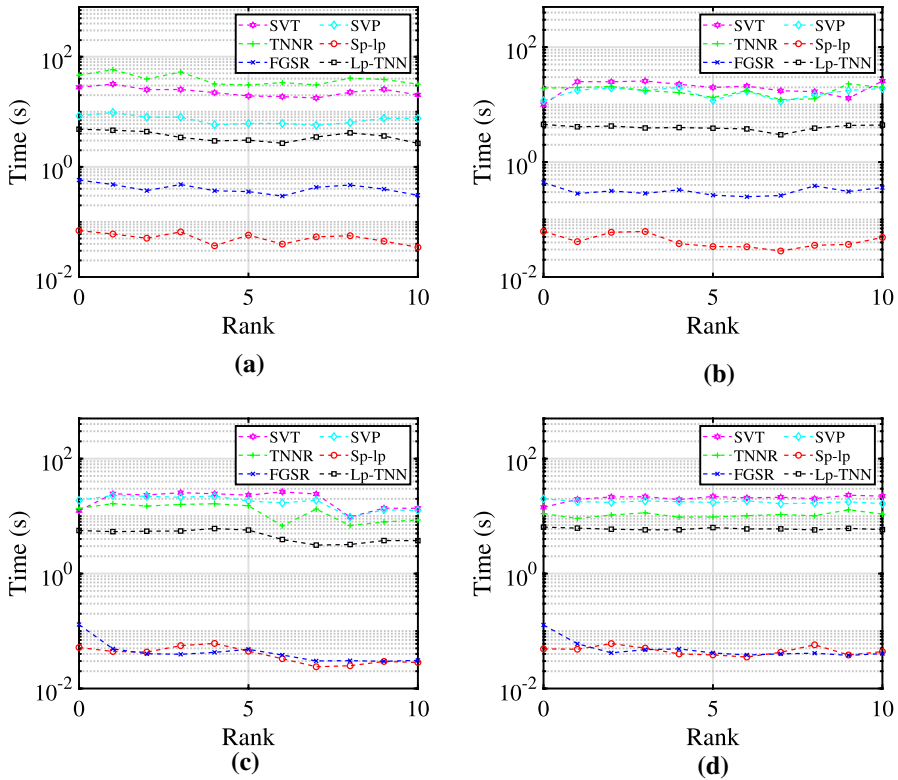


Fig. 2 The running time versus the SNRs using the synthetic data under different sampling rates: a 0.05, b 0.10, c 0.15, and d 0.20

5.4 MovieLens-1M dataset

In this subsection, the proposed method is implemented on the real-world MovieLens-1M dataset, which is available in [32]. This dataset consists of 1 million ratings (1 to 5) for 3900 movies by 6040 users. Here, the stopping threshold, tol , is set to 10^{-2} , and the update proportional constant, μ , to 0.001. And we set the maximum allowable iterations, n , to 100.

We randomly mask 50% of the known ratings and perform the recovery. The performance of the proposed method can be illustrated using the normalized mean absolute error (NMAE) and normalized root-mean-squared-error (RMSE) [33]. The measured NMAEs and RMSEs are shown in Table 4. It may be seen that the Lp-TNN obtains a smaller indicator, again outperforming the alternative methods.

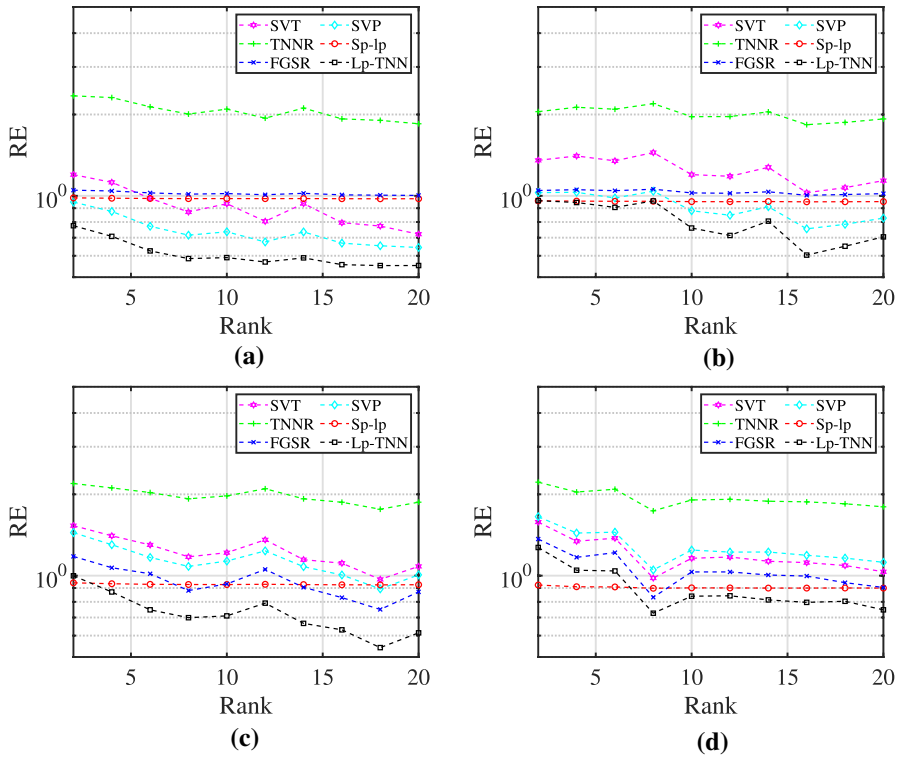


Fig. 3 The REs versus the ranks using the synthetic data under different sampling rates: **a** 0.05, **b** 0.10, **c** 0.15, and **d** 0.20

6 Conclusion

The matrix completion problem has a wide application in various fields. However, many matrix completion methods fail to provide a reliable estimation when the sample size is small. In this article, we present a matrix completion approach by combining truncated nuclear norm and Lp-norm, which can avoid a sharp drop when encountering the outliers. The truncated nuclear norm is a more accurate and robust approximation function of the rank function, while the Lp-norm as an error function can improve the model robustness, both of which provide a foundation for better performance.

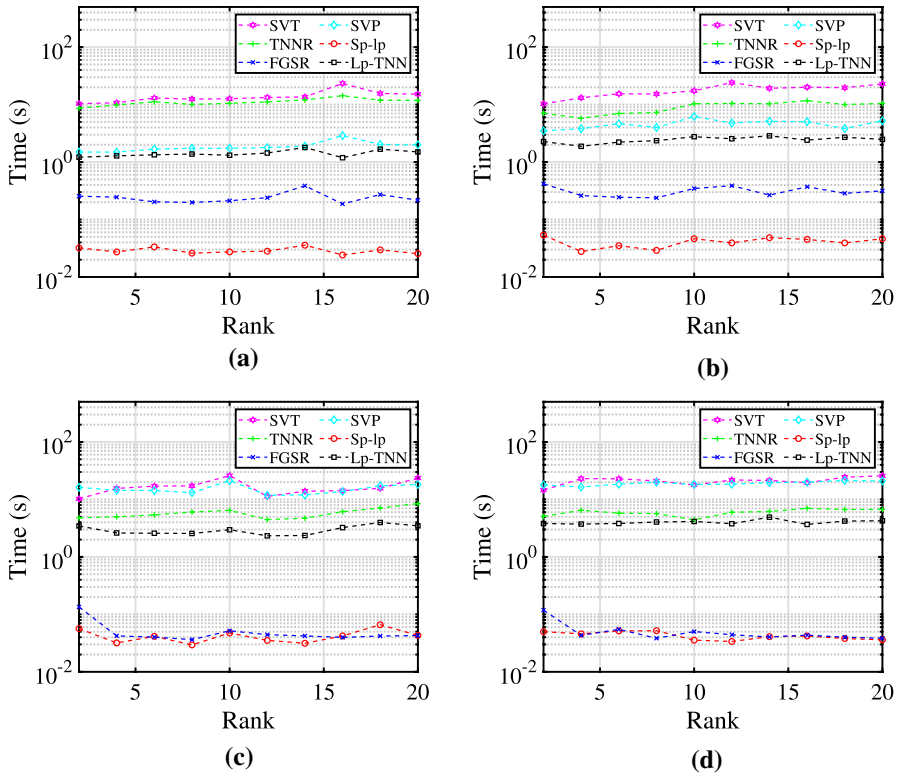


Fig. 4 The running time versus the ranks using the synthetic dataset under different sampling rates: **a** 0.05, **b** 0.10, **c** 0.15, and **d** 0.20

Although the objective function of the proposed model is not convex, a simple and efficient iterative scheme has been studied. We first decomposed the complex original problem into several relatively easy-to-handle sub-problems by the alternating direction multiplier method. Then, for the sub-problems, the function analysis method, the singular value threshold operator, and the augmented Lagrangian method are employed to form an efficient implementation, respectively.

Finally, we conduct a numerical experiment. We perform corresponding matrix completion experiments on both synthetic and real data. The experimental results of synthetic data show that our algorithm may always obtain better performances under arbitrary random initialization and Gaussian noise interference. Additionally, the proposed methods may not introduce too much computational cost, which confirms its effectiveness. The experimental results of real images show that the proposed method may efficiently restore realistic images, which can provide a robust estimation while suffering Gaussian noise. Especially in the case of a low sampling rate, the image can be restored with a very high peak signal-to-noise ratio and a faster rate. The method’s practical potential is illustrated on the MovieLens-1M dataset.

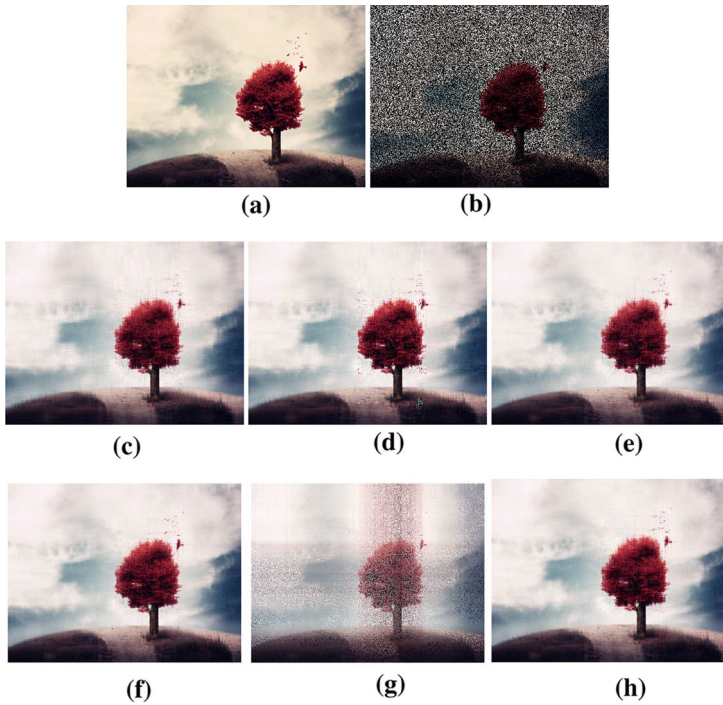


Fig. 5 The recovered images when the sampling rate is 0.4 by different algorithms: **a** original image, **b** sampling image, **c** SVT, **d** SVP, **d** TNNR, **f** Sp-lp, **g** FGSR, and **h** Lp-TNN

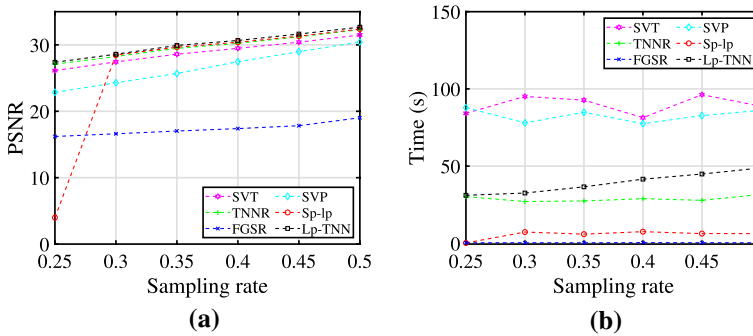


Fig. 6 The reconstruction performance by different algorithms, showing the analysis of the **a** PSNR and **b** running time

Table 4 The NAMEs and RMSEs of the MovieLens-1M dataset by different algorithms

	SVT	SVP	TNNR	Sp-lp	FGSR	Lp-TNN
NAME	0.2490	0.2386	0.8746	0.3525	0.2430	0.2284
RMSE	0.3452	0.3325	0.9779	0.4636	0.3354	0.3219
Time	1593.12 s	499.33 s	304.32 s	268.92 s	17.64 s	616.18 s

The best indicators obtained are shown in bold

Declarations

Conflict of interest The authors declare that they have no conflict of interest.

References

- Li X, Shen C, Dick A et al (2015) Online metric-weighted linear representations for robust visual tracking. *IEEE Transac Pattern Anal Mach Intell* 38(5):931–950. <https://doi.org/10.1109/TPAMI.2015.2469276>
- Tsiligianni E, Deligiannis N (2019) Deep coupled-representation learning for sparse linear inverse problems with side information. *IEEE Signal Process Lett* 26(12):1768–1772. <https://doi.org/10.1109/LSP.2019.2929869>
- Tang H, Sun Y, Yin B et al (2011) 3D face recognition based on sparse representation. *J Supercomput* 58(1):84–95. <https://doi.org/10.1007/s11227-010-0533-9>
- Ding X, Liang H, Jakobsson A et al (2022) High-resolution source localization exploiting the sparsity of the beamforming map. *Signal Process* 192:108377. <https://doi.org/10.1016/j.sigpro.2021.108377>
- Xu JL, Hsu YL (2022) Analysis of agricultural exports based on deep learning and text mining. *J Supercomput* 1–17. <https://doi.org/10.1007/s11227-021-04238-w>
- Zhang M, Desrosiers C (2018) High-quality image restoration using low-rank patch regularization and global structure sparsity. *IEEE Transact Image Process* 28(2):868–879. <https://doi.org/10.1109/TIP.2018.2874284>
- Jia S, Zhang X, Li Q (2015) Spectral-spatial hyperspectral image classification using ℓ_1/ℓ_2 regularized low-rank representation and sparse representation-based graph cuts. *IEEE J Sel Topics Appl Earth Obs Remote Sens* 8(6):2473–2484. <https://doi.org/10.1109/JSTARS.2015.2423278>
- Hassan MM, Ullah S, Hossain MS et al (2021) An end-to-end deep learning model for human activity recognition from highly sparse body sensor data in Internet of Medical Things environment. *J Supercomput* 77(3):2237–2250. <https://doi.org/10.1007/s11227-020-03361-4>
- Jiang J, Kang L, Huang J et al (2020) Deep learning based mild cognitive impairment diagnosis using structure MR images. *Neurosci Lett* 730:134971. <https://doi.org/10.1016/j.neulet.2020.134971>
- Yang L, Fang J, Duan H et al (2018) Fast low-rank Bayesian matrix completion with hierarchical Gaussian prior models. *IEEE Transact Signal Process* 66(11):2804–2817. <https://doi.org/10.1109/TSP.2018.2816575>
- Zhang S, Wang M (2019) Correction of corrupted columns through fast robust Hankel matrix completion. *IEEE Transact Signal Process* 67(10):2580–2594. <https://doi.org/10.1109/TSP.2019.2904021>
- Candès EJ, Recht B (2009) Exact matrix completion via convex optimization. *Found Comput Math* 9(6):717–772. <https://doi.org/10.1007/s10208-009-9045-5>
- Fazel M, Hindi H, Boyd S P (2001) A Rank Minimization Heuristic With Application to Minimum Order System Approximation. In: *Proceedings of the 2001 American Control Conference*. IEEE, 6: 4734–4739. <https://doi.org/10.1109/ACC.2001.945730>

14. Candès EJ, Tao T (2010) The power of convex relaxation: near-optimal matrix completion. *IEEE Transact Inf Theory* 56(5):2053–2080. <https://doi.org/10.1109/TIT.2010.2044061>
15. Tan VYF, Balzano L, Draper SC (2011) Rank minimization over finite fields. In: (2011) IEEE international symposium on information theory proceedings. IEEE 1195–1199. <https://doi.org/10.1109/ISIT.2011.6033722>
16. Cai JF, Candès EJ, Shen Z (2010) A singular value thresholding algorithm for matrix completion. *SIAM J Optim* 20(4):1956–1982. <https://doi.org/10.1137/080738970>
17. Jain P, Meka R, Dhillon I (2010) Guaranteed rank minimization via singular value projection. *Adv in Neural Inf Process Syst*, 23
18. Hu Y, Zhang D, Ye J et al (2012) Fast and accurate matrix completion via truncated nuclear norm regularization. *IEEE Trans on pattern analysis and machine intelligence* 35(9):2117–2130. <https://doi.org/10.1109/TPAMI.2012.271>
19. Nie F, Wang H, Huang H et al (2015) Joint Schatten p -norm and ℓ_p -norm robust matrix completion for missing value recovery. *Knowl Inf Syst* 42(3):525–544. <https://doi.org/10.1007/s10115-013-0713-z>
20. Yang L, Kou KI, Miao J (2021) Weighted truncated nuclear norm regularization for low-rank quaternion matrix completion. *J Vis Commun Image Represent* 81:103335. <https://doi.org/10.1016/j.jvcir.2021.103335>
21. Zhang H, Qian J, Zhang B et al (2019) Low-rank matrix recovery via modified Schatten- p norm minimization with convergence guarantees. *IEEE Trans Image Process* 29:3132–3142. <https://doi.org/10.1109/TIP.2019.2957925>
22. Fan J, Ding L, Chen Y, et al. (2019) Factor group-sparse regularization for efficient low-rank matrix recovery. *Adv Neural Inf Process Syst*, 32
23. Nie F, Huang H, Cai X, et al. (2010) Efficient and Robust Feature Selection Via Joint ℓ_2 , 1-Norms Minimization. In: Proceedings of the 23rd International Conference on Neural Information Processing Systems-Volume 2. 1813–1821
24. Gabay D, Mercier B (1976) A dual algorithm for the solution of nonlinear variational problems via finite element approximation. *Comput Math Appl* 2(1):17–40. [https://doi.org/10.1016/0898-1221\(76\)90003-1](https://doi.org/10.1016/0898-1221(76)90003-1)
25. Gerald C F. Applied numerical analysis (2004). Pearson Education India
26. Boyd S, Parikh N, Chu E (2011) Distributed optimization and statistical learning via the alternating direction method of multipliers. Now Publishers Inc
27. Blaschke B, Neubauer A, Scherzer O (1997) On convergence rates for the iteratively regularized Gauss-Newton method. *IMA J Numer Anal* 17(3):421–436. <https://doi.org/10.1093/imanum/17.3.421>
28. Kou J, Li Y, Wang X (2006) A modification of Newton method with third-order convergence. *Appl Math Comput* 181(2):1106–1111. <https://doi.org/10.1016/j.amc.2006.01.076>
29. Li M, Kwok J T Y, Lü B (2010) Making Large-Scale Nyström Approximation Possible. In: ICML 2010-proceedings, 27th International Conference on Machine Learning, 631
30. Bertsekas D P (2014) Constrained optimization and lagrange multiplier methods. Academic press
31. Toh KC, Yun S (2010) An accelerated proximal gradient algorithm for nuclear norm regularized linear least squares problems. *Pacific J Optim* 6(615–640):15
32. F. Maxwell Harper and Joseph A. Konstan (2015). The movielens datasets: History and context. *ACM transactions on interactive intelligent systems (TiiS)* 5, 4, Article 19 (December 2015), 19 pages. <https://doi.org/10.1145/2827872>
33. Shamir O, Shalev-Shwartz S (2014) Matrix completion with the trace norm: learning, bounding, and transducing. *J Mach Learn Res* 15(1):3401–3423. <https://doi.org/10.5555/2627435.2697073>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.