Matrix completion via modified schatten 2/3-norm

Jincai Ha[1], Chuan Li[2], Xiaohu Luo[3] and Zhi Wang[2,3]*

*Correspondence:
chiw@swu.edu.cn

[1] School of Mathematics
and Information Science, North
Minzu University, Yinchuan,
China
[2] Big Data and Intelligence
Engineering School, Chongqing
College of International Business
and Economics, Chongqing,
China
[3] College of Computer
and Information Science,
Southwest University,
Chongqing 400715, China

Abstract

Low-rank matrix completion is a hot topic in the field of machine learning. It is widely used in image processing, recommendation systems and subspace clustering. However, the traditional method uses the nuclear norm to approximate the rank function, which leads to only the suboptimal solution. Inspired by the closed-form formulation of $L_{2/3}$ regularization, we propose a new truncated schatten 2/3-norm to approximate the rank function. Our proposed regularizer takes full account of the prior rank information and achieves a more accurate approximation of the rank function. Based on this regularizer, we propose a new low-rank matrix completion model. Meanwhile, a fast and efficient algorithm are designed to solve the proposed model. In addition, a rigorous mathematical analysis of the convergence of the proposed algorithm is provided. Finally, the superiority of our proposed model and method is investigated on synthetic data and recommender system datasets. All results show that our proposed algorithm is able to achieve comparable recovery performance while being faster and more efficient than state-of-the-art methods.

Keywords: Matrix completion, $L_{2/3}$ regularization, A new truncated schatten 2/3-norm, Recommender system

1 Introduction

The problem of recovering an incomplete low rank or approximately low rank matrix with missing values, namely low rank matrix completion (LRMC), has attracted significant attention in recent years. Such a problem is a central issue in the field of computer vision and machine learning, and can be found in various practical applications, such as recommender systems [1, 2], motion capture [3], video denoising [4], subspace clustering [5], and hyperspectral imaging [6]. Roughly, the methods for LRMC can be classified into two categories: the low rank matrix factorization methods and the rank minimization methods. In this work, we only focus on the latter category. It is because the factorization based algorithms are heavily rely on a prespecified rank [7], which is difficult to preestimated in some real applications.

It is well known that the rank function has nonconvex and discontinuous properties. Therefore, the rank minimization problem is NP-hard and is difficult to optimize. To

Ha *et al. EURASIP Journal on Advances in Signal Processing*   (2023) 2023:62

Page 2 of 19

alleviate this problem, many researchers have suggested to relax the rank function and, instead, to consider the nuclear norm. Theoretical analysis illustrate that the nuclear norm, i.e., the sum of singular values of the matrix, is the tightest convex lower bound of the rank [8]. Candès and Recht have proven that [9], if the observed entries of the matrix are sampled uniformly at random and the matrix satisfies restricted isometry property condition, the target low rank matrix can be exactly recovered by nuclear norm minimization. Because of this, the nuclear norm minimization gets its popularity and has been accepted as a very powerful method for the solution of low rank problems. During the past decades, a variety of algorithms have been proposed to solve the nuclear norm based model with strong theoretical guarantees, such as singular value thresholding (SVT) [10], accelerated proximal gradient with line search algorithm (APGL) [11], soft-impute [12] and its accelerated version (AIS-Impute) [13]. Nevertheless, the relaxation of the nuclear norm is too loose to approximate the rank function. Thus, the algorithms mentioned above may only yield suboptimal performance in practice. One important reason is that the nuclear norm treats all singular values equally. Intuitively, large singular values should shrink less, and small singular values should shrink more. All in all, a further improvement is required.

A very natural idea is the suggestion of the use of nonconvex surrogate functions to approximate the rank function. The representative nonconvex surrogate functions include the schatten p-norm ($0 < p < 1$) [14], capped-l_1 norm [15], log-sum penalty (LSP) [16], smmothly clipped absolute deviation (SCAD) [17], transformed l_1 penalty [18, 19], and Laplace [20]. The empirical results demonstrate that these nonconvex surrogate functions can achieve better performance than that of its convex counterpart. However, the resultant optimization problem is nonconvex, nonsmooth, and non-Lipschitz. It is a big challenge to solve these optimization problems efficiently. To this end, a number of algorithms, such as iteratively reweighted nuclear norm (IRNN) [21], fast nonconvex low rank learning (FaNCL) [22], matrix completion based on nonconvex relaxation (MC-NC) [23], double nonconvex nonsmooth rank (DNNR) relaxations function based method [24], and block-wise model dubbed differentiable low-rank learning (DLRL) [25], have been proposed to solve the nonconvex low rank approximation problems.

Another parallel research is to consider the different contributions of different rank components, with the weighted nuclear norm minimization (WNNM) [26, 27] being the most representative one. Comparing with the traditional nuclear norm minimization, the weighted nuclear norm minimization scheme assigns different weights to different singular values such that decrease the punishment on larger singular values. In order to achieve better recovery performance, the weighted schatten p-norm minimization (WSNM) [28] is proposed to solve LRMC problem. By setting appropriate values for the weights and p, the weighted nuclear norm minimization can be viewed as a special case of the weighted schatten p-norm minimization. The WNNM and WSNM models have been successfully applied to deal with typical low level vision tasks, such as image denoising and background subtraction [27, 28]. However, both WNNM and WSNM do not take into consideration a priori rank information. The variance of data distribution within the target rank does not need to minimize, which means

that we only need to minimize the singular values in residual ranks. Along this line of research, the truncated nuclear norm (TNN) [29] and partial sum of singular values (PSSV) [30] have been proposed for low rank matrix recovery problems. Indeed, the TNN and PSSV can be regared as one of the concrete examples of WNNM and WSNM. Although TNN can achieve a more accurate and robust approximation to the rank function, it still suffer from some drawbacks. More specifically, the algorithms for solving the traditional TNN-based models are time-consuming and a prespecified parameter is difficult to preestimated. Recent studies in [31–34], and [35] have addressed partially these issues.

In this work, we continue such a study. Our aim is to establish a novel continuous but nonconvex regularizer namely Modified Schatten 2/3-Norm Minimization with Reweighting strategy (TSNMR) for LRMC problem. Subsequently, a more accurate and flexible model with TSNMR is build. As can be seen latter, our proposed model is fully consider the priori rank information, and achieves robust approximation to the rank function. Furthermore, its solution can be analytically expressed in a thresholding form. Based on this finding, a computationally efficient optimization method is designed for solving matrix completion problems. The contributions of this work are highlighted as follows:

1. By virtue of the idea of TNN and WSNM, a nonvel continuous but nonconvex regularizer namely TSNMR is proposed for LRMA problem. Armed with it, a more accurate and flexible model is obtained. Meanwhile, the property of TSNMR is also analysed, and its closed-form solutions can be derived from a thresholding operator. By involving this finding, the resultant optimization model becomes more tractable.
2. An efficient and fast optimization algorithm with inexact proximal steps and Nesterov's acceleration rules is designed to optimize the proposed model. Rigorous mathematical proof of the proposed algorithm demonstrating that any accumulation point of its generated sequence is a first-order stationary point.
3. We apply the proposed TSNMR model to solve some typical low rank matrix completion problems, e.g., image inpainting.
4. Experimental results on synthetic data and color images demonstrate that our proposed model can achieve superior performance than the state-of-the-art models.

The rest of this paper is organized as follows. Section 2 briefly reviews some related works. Section 3 presents the proposed model and develops its optimization method with rigorous convergence guarantees. Section 4 introduces the applications of our proposed model to low level tasks. Section 5 reports and analyzes the experimental results. Finally, several concluding remarks are provided in Sect. 6.

Notations: Some notations used in this paper are listed in Table 1.

Ha *et al. EURASIP Journal on Advances in Signal Processing*   (2023) 2023:62

Page 4 of 19

Table 1 The summarization of notations

Notation	Description
$M \in \mathbb{R}^{m \times n}$	Matrix with size $m \times n$
$\sigma_i(M) = \sigma_i$	ith largest singular value of M
$u_i(M) = u_i$	ith left singular vector of M
$v_i(M) = v_i$	ith right singular vector of M
$\|M\|_* = \sum_{i=1}^{\min\{m,n\}} \sigma_i(M)$	Nuclear norm of M
$\|M\|_F = \sqrt{Tr(M^T M)}$	Frobenius norm of M
$\|M\|_p = \left(\sum_{i=1}^{\min\{m,n\}} \sigma_i(M)^p \right)^{1/p}$	Schatten p-norm of M
$\langle M_1, M_2 \rangle = Tr(M_1^T M_2)$	Standard inner product
∇f	Gradient of a differentiable function f

2 Background

In this section, we briefly introduce the closed-form thresholding formula for $L_{2/3}$ regularization and some widely used nonconvex low rank regularizers.

2.1 Thresholding formulas for $L_{2/3}$ regularization

The $L_{2/3}$ regularization model was recently proposed by Xu et al. [37] for solving the image deconvolution problem. It is believe that the $L_{2/3}$ regularization is more effective than $L_{1/2}$ regularization [36] in many practical applications. Mathematically, the $L_{2/3}$ regularization model can be represented as

$$h_{\lambda, \frac{2}{3}}(a) = \arg\min_{x \geq 0} \left\{ \frac{1}{2}(x - a)^2 + \lambda x^{\frac{2}{3}} \right\}, \tag{1}$$

where $a \geq 0$ is a constant in $\mathbb{R}$. It follows from [37] that the solutions of (1) can be analytically expressed by

$$h_{\lambda, \frac{2}{3}}(a) = \begin{cases} \frac{\left(|\phi_\lambda(a)| + \sqrt{\frac{2a}{|\phi_\lambda(a)|} - |\phi_\lambda(a)|^2} \right)}{8}, & \text{if } a > \gamma \\ 0, & \text{if } a \leq \gamma \end{cases} \tag{2}$$

where

$$\phi_\lambda(a) = \frac{2}{\sqrt{3}}(2\lambda)^{\frac{1}{4}} \left(\cosh\left(\frac{\text{arccosh}(\frac{27a^2}{16}(2\lambda)^{-\frac{3}{2}})}{3} \right) \right)^{\frac{1}{2}} \tag{3}$$

and $\gamma = 2/3(3(2\lambda)^3)^{1/4}$.

2.2 Existing nonconvex low rank regularizers for LRMA

(1) Weighted nuclear norm With the aim of improving the flexibility of nuclear norm minimization, Gu et al. [27] proposed the weighted nuclear norm (WNN), which can be represented as

$$\|X\|_{w,*} = \sum_i w_i \sigma_i(X),$$
(4)

where $X \in \mathbb{R}^{m \times n}$, $w = [w_1, w_2, \cdots, w_n]^T$, and $w_1 \geq w_2 \geq \cdots \geq w_n \geq 0$. Therefore, the WNNM model is obtained and it can be solved by weighted nuclear norm proximal (WNNP) operator

$$X^* = \text{prox}_{\|X\|_{w,*}}(W) = \arg\min_{X \in \mathbb{R}^{m \times n}} \|W - X\|_F^2 + \|X\|_{w,*}.$$
(5)

However, it is difficult to solve (5) due to the nonconvexity of WNNM. Fortunately, theoretical analysis of (5) reveals that it is actually a quadratic programming problem with linear constraints. Thus, the globally optimal solution of (5) can be achieved in closed-form.

Lemma 1   ([27]) *Suppose that $W \in \mathbb{R}^{m \times n}$ admits singular value decomposition (SVD) as $U\Sigma V^T$, where $\Sigma = \text{Diag}(\sigma)$, $\sigma = [\sigma_1, \sigma_2, \cdots, \sigma_r]^T$, and $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r \geq 0$. The global solution to*

$$\arg\min_{X \in \mathbb{R}^{m \times n}} \|W - X\|_F^2 + \|X\|_{w,*}$$
(6)

is given by

$$X^* = \text{prox}_{\|X\|_{w,*}}(W) = U\Sigma'V^T,$$
(7)

where $\Sigma'_{ii} = \max(\Sigma_{ii} - w_i/2, 0)$.

(2) Weighted schatten p-norm Inspired by the Schatten p-norm and WNN, Xie et al. [28] proposed the weighted shatten p-norm (WSN), which can be represented as

$$\|X\|_{w,S_p} = \left(\sum_i w_i \sigma_i^p \right)^{1/p}.$$
(8)

WSN can be seen as a generalization of WNN, but it can approximate the rank function better than WNN. By this relaxation, the WSNM model could be obtained. To handle such models efficiently, one need to consider the following nonconvex optimization problem.

$$X^* = \text{prox}_{\|X\|_{w,S_p}}(W) = \arg\min_{X \in \mathbb{R}^{m \times n}} \|W - X\|_F^2 + \|X\|_{w,S_p}.$$
(9)

Intuitively, solving (9) is nontrivial due to the noncovexity and nonsmoothness of the objective function. However, the following lemma shows that the optimal solution of (9) can be achieved by solving r independent subproblems, where r is the rank of W.

Lemma 2   ([28]) *Suppose that $W \in \mathbb{R}^{m \times n}$ admits SVD as $U\Sigma V^T$, the optimal solution to*

Ha *et al. EURASIP Journal on Advances in Signal Processing* (2023) 2023:62

Page 6 of 19

$$\underset{X \in \mathbb{R}^{m \times n}}{\arg \min} \|W - X\|_F^2 + \|X\|_{w,S_p} \tag{10}$$

is given by

$$X^* = \text{prox}_{\|X\|_{w,*}}(W) = U\text{Diag}(\varrho^*)V^T, \tag{11}$$

where

$$\varrho^* \in \underset{\varrho_1,\cdots,\varrho_r}{\arg \min} \sum_{i=1}^{r} \left[(\varrho_i - \sigma_i)^2 + w_i \varrho_i^p \right], i = 1, \cdots, r$$

$$s.t. \quad \varrho_i \geq 0, \quad and \quad \varrho_i \geq \varrho_{i+1} \tag{12}$$

It follows from [28] that (12) can be decoupled into *r* independent subproblems, and these subproblems can be effectively solved by generalized soft-thresholding (GST) algorithm (for more details about GST, please refer to [28]).

Comparing with the NNM models, the WNNM and WSNM models are fully consider the difference between different singular values, and achieve better approximation to the rank function. Nevertheless, these models do not take into consideration a priori rank information for the practical applications. Thus, they are still not accurate enough for solving real LRMC problems.

3 The proposed model and its optimization method

In this section, we first introduce the definition of TSNMR and then establish the low rank matrix completion problem. By analysing the property of TSNMR, the optimization method for the resultant model is proposed and its convergence property is analysed. Furthermore, we also discuss the adaptive regularization parameter.

3.1 Problem formulation

In this work, we devise a novel continuous but nonconvex surrogate function, namely truncated schatten 2/3-norm minimization with reweighting strategy. More precisely, the TSNMR is defined as

$$P_{2/3,\alpha}^{\epsilon}(X) = \sum_{i=r+1}^{q} \psi_{2/3,\alpha}^{\epsilon}(\sigma_i(X)) = \sum_{i=r+1}^{q} \frac{C\sigma_i(X)^{2/3}}{(\sigma_i(X) + \epsilon_i)^{2/3-\alpha}}, \tag{13}$$

where $X \in \mathbb{R}^{m \times n}$, r is the target rank, $q = \min\{m, n\}$, $\epsilon_{r+1} \geq \epsilon_{r+2} \geq \cdots \geq \epsilon_q > 0$ are set to sufficiently small positive numbers to avoid dividing by 0, and $C > 0$ is a constant. Our proposed TSNMR not only takes into consideration the importance of different rank components, but also fully considers the priori rank information.

Obviously, the function $\psi_{2/3,\alpha}^{\epsilon}(|t|) = C|t|^{2/3}/(|t| + \epsilon)^{2/3-\alpha}$ is concave for any $\alpha \in (0, 2/3]$ and $\epsilon > 0$. With the change of parameters α and ϵ, it is easy to verify that

$$\lim_{\alpha \to 0^+} \lim_{\epsilon \to 0^+} \psi_{2/3,\alpha}^{\epsilon}(|t|) = \begin{cases} 0, & \text{if } t = 0 \\ 1, & \text{otherwise.} \end{cases} \tag{14}$$

where $C = 1$. Therefore, with the proper choices of α and ϵ_i, we have

$$
\begin{aligned}
\lim_{\alpha \to 0^+} \lim_{\epsilon \to 0^+} P_{2/3,\alpha}^{\epsilon}(X) &= \lim_{\alpha \to 0^+} \lim_{\epsilon \to 0^+} \sum_{i=r+1}^{q} \psi_{2/3,\alpha}^{\epsilon}(\sigma_i(X)) \\
&= \lim_{\alpha \to 0^+} \lim_{\epsilon \to 0^+} \sum_{i=r+1}^{q} \frac{(\sigma_i(X))^{2/3}}{(\sigma_i(X) + \epsilon)^{2/3-\alpha}} \\
&= \mathrm{rank}(X) - r.
\end{aligned}
\tag{15}
$$

In other words, if we set $\alpha \to 0^+$ and $\epsilon \to 0^+$, then $P_{p,\alpha}^{\epsilon}(X)$ is degraded to TNN in [29] and [30].

Armed with TSNMR, in this paper, we mainly focus on the following low-rank minimization problem, which can be formulated as the form

$$
\min_{X \in \mathbb{R}^{m \times n}} F(X) = \frac{1}{2} \|\mathcal{P}_{\Omega}(X) - \mathcal{P}_{\Omega}(M)\|_F^2 + \lambda P_{2/3,\alpha}^{\epsilon}(X),
\tag{16}
$$

where $\lambda > 0$ is given parameter, Ω denotes the set of the locations of the observed entries, and $\mathcal{P}_{\Omega}$ denotes the orthogonal projector onto the span of matrices vanishing outside of Ω, i.e.,

$$
[\mathcal{P}_{\Omega}(M)]_{ij} = \begin{cases} M_{ij}, & \text{if } (i, j) \in \Omega \\ 0, & \text{otherwise} \end{cases}
$$

3.2 Solving scheme

Directly solving the nonconvex and nonsmooth optimization problem (16) is difficult. To make this issue tractable, we first define the following quadratic function

$$
F_{\lambda,\mu}(X, Y) = f(Y) + \langle \nabla f, X - Y \rangle + \frac{1}{2\mu} \|X - Y\|_F^2 + \lambda \sum_{i=r+1}^{q} \frac{C\sigma_i(X)^{2/3}}{(\sigma_i(Y) + \epsilon_i)^{2/3-\alpha}},
\tag{17}
$$

where $f(Y) = (1/2)\|\mathcal{P}_{\Omega}(Y) - \mathcal{P}_{\Omega}(M)\|_F^2$ and $X, Y \in \mathbb{R}^{m \times n}$. For any $\mu > 0$, it is easy to find that $F(X) = F_{\lambda,\mu}(X, X)$. In what follows, we will reveal that any global minimizer of $F(X)$ is also a global minimizer of $F_{\lambda,\mu}(X, Y)$. The following lemma addresses this issue.

Lemma 3 *Assume that $\mu \leq 1/L_f$ and X^* is the global minimizer of F(X), then we have*

$$
F_{\lambda,\mu}(X^*, X^*) \leq F_{\lambda,\mu}(X, X^*), \quad \forall X \in \mathbb{R}^{m \times n}.
\tag{18}
$$

Proof Considering the objective function $F(X)$ at $X = X^*$, we have

$$
F_{\lambda,\mu}(X^*, X^*) = F(X^*) \leq f(X) + \lambda \sum_{i=r+1}^{q} \frac{C\sigma_i(X)^{2/3}}{(\sigma_i(X^*) + \epsilon_i)^{2/3-\alpha}}.
\tag{19}
$$

Ha *et al. EURASIP Journal on Advances in Signal Processing* (2023) 2023:62

Page 8 of 19

Although f is possibly nonconvex, from the assumption that f is differentiable with L_f-Lipschitz continuous gradient, we can obtain that [40, 41]

$$f(X) \leq f(Y) + \langle \nabla f, X - Y \rangle + \frac{1}{2\mu} \|X - Y\|_F^2. \tag{20}$$

Substituting (20) into (19) and setting $Y = X^*$, we have

$$
\begin{aligned}
F_{\lambda,\mu}(X^*, X^*) &\leq f(X^*) + \langle \nabla f, X - X^* \rangle \\
&\quad + \frac{1}{2\mu} \|X - X^*\|_F^2 + \lambda \sum_{i=r+1}^{q} \frac{C\sigma_i(X)^{2/3}}{(\sigma_i(X^*) + \epsilon_i)^{2/3-\alpha}} \\
&= F_{\lambda,\mu}(X, X^*).
\end{aligned}
\tag{21}
$$

We complete the proof. $\square$

By Lemma 3, we can conclude that the global minimizer of optimization problem (16) can be obtained by computing the optimal solution of $F_{\lambda,\mu}(X, Y)$ in optimization problem (17). Using the basic algebra calculation, we obtain that

$$F_{\lambda,\mu}(X, Y) = f(Y) - \frac{\mu}{2}\|\nabla f(Y)\|_F^2 + \frac{1}{2\mu}\|X - B_\mu(Y)\|_F^2 + \lambda \sum_{i=r+1}^{q} \frac{C\sigma_i(X)^{2/3}}{(\sigma_i(Y) + \epsilon_i)^{2/3-\alpha}}, \tag{22}$$

where $B_\mu(Y) = Y - \mu\nabla f(Y)$. Ignoring constant terms of (22), the global minimizer of $F_{\lambda,\mu}(X, Y)$ can be obtained by solving the following optimization problem

$$\arg\min_{X \in \mathbb{R}^{m \times n}} \frac{1}{2}\|X - B_\mu(Y)\|_F^2 + \lambda\mu \sum_{i=r+1}^{q} \frac{C\sigma_i(X)^{2/3}}{(\sigma_i(Y) + \epsilon_i)^{2/3-\alpha}}. \tag{23}$$

Now the crucial thing we need to deal with is how to obtain the global minimizer of optimization problem (23). Thus, we extend the aforementioned well-known $L_{2/3}$ regularization to solve the resultant nonconvex optimization problem. Additionally, in the next section, we will show that its global optimal solution can be easily obtained in closed-form.

3.3 Optimization

In this subsection, we will exploit an efficient and fast optimization method to optimize problem (16). The main obstacle in this method is how to solve the optimization problem (23). As mentioned above, owing to the nonconvexity of TSNMR, this problem is much more challenging. To this end, we first show that the global optimal solution of such problem can be efficiently achieved. In order to better address this issue, we introduce the following lemma.

Lemma 4 (von Neumann [42, 43]) *For any matrices A and B in $\mathbb{R}^{m \times n}$ and assume that $\sigma(A)$ and $\sigma(B)$ are the singular value vector of A and B, respectively, then*

Ha *et al. EURASIP Journal on Advances in Signal Processing*   (2023) 2023:62

Page 9 of 19

$$\langle A, B \rangle \leq \langle \sigma(A), \sigma(B) \rangle. \tag{24}$$

The case of equality occurs iff there exists a simultaneous SVD U and V^T of A and B in the following form

$$A = U\mathrm{Diag}(\sigma(\mathrm{A}))\mathrm{V}^{\mathrm{T}}, \quad \mathrm{B} = \mathrm{U}\mathrm{Diag}(\sigma(\mathrm{B}))\mathrm{V}^{\mathrm{T}}. \tag{25}$$

By means of von Neummann's lemma, we establish the following theorem, which reveals that the global minimizer of optimization problem (23) can be obtained in closed-form.

Theorem 3.1 *Suppose that $\lambda > 0$, $B = Y - \mu \nabla f(Y)$ admits SVD as $U\mathrm{Diag}(\sigma)\mathrm{V}^{\mathrm{T}}$. Let $B = \hat{B} + \tilde{B} = \hat{U}\mathrm{Diag}(\hat{\sigma})\hat{V}^T + \tilde{U}\mathrm{Diag}(\tilde{\sigma})\tilde{V}^T$, where $\hat{\sigma} = (\sigma_1, \cdots, \sigma_r, 0, \cdots, 0)$, $\tilde{\sigma} = (0, \cdots, 0, \sigma_{r+1}, \cdots, \sigma_q)$, $\hat{U}$ and $\hat{V}$ are the singular vector matirces correspongding to the r largest singular values, $\tilde{U}$ and $\tilde{V}$ from the $(r+1)$th to the last singular values. Then, the optimal solutions to*

$$\underset{X \in \mathbb{R}^{m \times n}}{\arg \min} \frac{1}{2} \|X - B\|_F^2 + \lambda \mu P_{2/3,\alpha}^{\epsilon}(X) \tag{26}$$

are given by

$$X^* = \mathrm{prox}_{\lambda', \mathrm{P}_{2/3,\alpha}^{\epsilon}(\cdot)}(\mathrm{B}), \tag{27}$$

where $\lambda' = 2\lambda\mu C/(\sigma_i(Y) + \epsilon_i)^{2/3-\alpha}$, and $\mathrm{prox}_{\lambda, P_{2/3,\alpha}^{\epsilon}(\cdot)}(B) = \hat{B} + \tilde{U}\big(\mathrm{Diag}(H_\lambda(\tilde{\sigma}))\big)\tilde{V}^T$ with

$$H_\lambda(\tilde{\sigma}) = \big(h_{\lambda,2/3}(\sigma_{r+1}), \cdots, h_{\lambda,2/3}(\sigma_q)\big)^T. \tag{28}$$

Proof Assume that $\tau = \lambda\mu$ and X admits SVD as $U'\mathrm{Diag}(\sigma')\mathrm{V}'^{\mathrm{T}}$. Note that

$$\begin{aligned}
\frac{1}{2}\|X - B\|_F^2 + \tau \sum_{i=r+1}^{q} \phi_i &= \frac{1}{2}(\langle B, B \rangle - 2\langle X, B \rangle + \langle X, X \rangle) + \tau \sum_{i=r+1}^{q} \phi_i \\
&= \frac{1}{2}\left(\sum_{i=1}^{q} \sigma_i^2 - 2\langle X, B \rangle + \sum_{i=1}^{q} \sigma_i'^2\right) + \tau \sum_{i=r+1}^{q} \phi_i.
\end{aligned} \tag{29}$$

where $\phi_i = (C\sigma_i'^{2/3})/((\sigma_i(Y) + \epsilon_i)^{2/3-\alpha})$.

By applying the von Neumann trace inequality in Lemma 4, we can obtain that $\langle X, B \rangle$ reaches its maximum value $\sum_i^q \sigma_i' \sigma_i$ if $U = U'$ and $V = V'$. Therefore, we can get

$$\begin{aligned}
\frac{1}{2}\|X - B\|_F^2 + \tau \sum_{i=r+1}^{q} \phi_i &\geq \frac{1}{2}\left(\sum_{i=1}^{q} \sigma_i^2 - 2\sum_{i=1}^{q} \sigma_i' \sigma_i + \sum_i^q \sigma_i'^2\right) + \tau \sum_{i=r+1}^{q} \phi_i \\
&= \frac{1}{2}\sum_{i=1}^{q}(\sigma_i' - \sigma_i)^2 + \tau \sum_{i=r+1}^{q} \phi_i.
\end{aligned} \tag{30}$$

Moreover, the Eq. (30) can be further rewritten as

$$\frac{1}{2}\|X - B\|_F^2 + \tau \sum_{i=r+1}^{q} \phi_i \geq \frac{1}{2} \sum_{i}^{q} (\sigma_i' - \sigma_i)^2 + \tau \sum_{i=r+1}^{q} \phi_i$$

$$= \frac{1}{2} \sum_{i=1}^{r} (\sigma_i' - \sigma_i)^2 + \frac{1}{2} \sum_{i=r+1}^{q} \left((\sigma_i' - \sigma_i)^2 + 2\tau \phi_i \right) \tag{31}$$

It is easy to observe that Eq. (31) consists of simple quadratic equations for each σ_i' independently. Thus, by using the first-order optimality condition and the closed-form thresholding formula for $L_{2/3}$ regularization, we can obtain

$$\sigma_i' = \begin{cases} \sigma_i, & \text{if i} \leq \text{r} \\ h_{2\tau C/(\sigma_i(Y)+\epsilon_i)^{2/3-\alpha}, 2/3}(\sigma_i), & \text{if i} > \text{r} \end{cases} \tag{32}$$

Hence, the global optimal solutions to (26) can be achieved as

$$X^* = \hat{B} + \tilde{U}\left(\text{Diag}(H_{\lambda'}(\tilde{\sigma}))\right)\tilde{V}^T, \tag{33}$$

where $\lambda' = 2\lambda\mu C/(\sigma_i(Y) + \epsilon_i)^{2/3-\alpha}$, and

$$H_\lambda(\tilde{\sigma}) = \left(h_{\lambda, \frac{2}{3}}(\sigma_{r+1}), \cdots, h_{\lambda, \frac{2}{3}}(\sigma_q) \right)^T, \tag{34}$$

which are the desired results. We complete the proof. $\square$

As can be seen from Theorem 3.1, solving the optimization problem (23) involves a full SVD step. As we all know, for any matrix $B \in \mathbb{R}^{m \times n}$, computing its SVD takes $O(mn^2)$ time. Therefore, when the scale of matrix B is large, directly computing its SVD may be time-consuming. Fortunately, from (29) in Theorem 3.1, we only need to compute the singular values larger than γ, which can be made more efficient by using partial SVD. Specifically, we first employ the power method [44, 45] algorithm to achieve a orthogonal matrix $Q \in \mathbb{R}^{m \times t}$, and then perform SVD on a much smaller matrix. Inspired by [22, 44], and [46], we establish the following lemma to address this issue.

Lemma 5 *Assume that B has $\hat{r} \leq t$ singular values that are larger than γ, and let $U_{\hat{r}}\text{Diag}(\sigma_{\hat{r}})V_{\hat{r}}^T$ be the rank-$\hat{r}$ SVD of B, then there exists an orthonormal matrix $Q \in \mathbb{R}^{m \times t}$ ($t \ll n$), such that*

(1) *$span(U_{\hat{r}}) \subseteq span(Q)$, and*
(2) *$prox_{\lambda, P_{p,\alpha}^\epsilon(\cdot)}(B) = Q prox_{\lambda, P_{p,\alpha}^\epsilon(\cdot)}(Q^T B)$.*

Proof The proof follows the footsteps of Proposition 1 in [44], and we omit it here. $\square$

Since the partial SVD strategy is employed to compute the proxima operator in (27), this can be made the results inexact, meaning that

$$X_k = \text{prox}_{\lambda, P^\epsilon_{p,\alpha}(\cdot)}(B) = \{U | \lambda P^\epsilon_{p,\alpha}(U) + \frac{1}{2}\|U - B\|_F^2$$
$$\leq \xi_k + \lambda P^\epsilon_{p,\alpha}(V) + \frac{1}{2}\|V - B\|_F^2, \ \forall V \in \mathbb{R}^{m \times n}\}, \tag{35}$$

where ξ_k denotes the error in the proximal operator at the kth iteration.

Algorithm 3.1 TSNMR based algorithm for solving problem (16)

Input: choose $\mu \leq 1/L_f$, β, $R \in \mathbb{R}^{n \times t}$.
1: Initialize: $X_0 = Y_1 = 0$,
2: **for** $k = 1 : K$ **do**
3: $\quad B = Y_k - \mu\nabla f(Y_k)$;
4: $\quad Q = \text{PowerMehtod}(B, R)$;
5: $\quad X_k = Q\text{prox}_{\hat{\lambda}, P^\epsilon_{p,\alpha}(\cdot)}(Q^T B)$; $//\hat{\lambda} = \lambda'$
6: $\quad$ Update λ by (36);
7: $\quad V_k = X_k + \beta(X_k - X_{k-1})$;
8: $\quad$ **if** $F(X_k) \leq F(V_k)$ **then**
9: $\quad\quad Y_{k+1} = X_k$;
10: $\quad$ **else**
11: $\quad\quad Y_{k+1} = V_k$;
12: $\quad$ **end if**
13: **end for**
14: **return** X_K

With the representation (27), the TSNMR-based algorithm for solving the problem (16) is naturally proposed in Algorithm 1.

In convex optimization, the Nesterov's acceleration rules are commonly used to speed up the convergence of first-order methods. Recently, this acceleration strategy has been successfully extended to solve the nonconvex optimization problem [47–50]. In this work, we try to integrate Nesterov's acceleration strategy with our proposed algorithm. As can be seen from Algorithm 1, the accelerated iterate is obtained in step 8. Since the TSNMR is absent convexity, a monitor is needed to ensure that the objective value F can achieve a sufficient decrease (step 9). Specifically, if V_k is a good extrapolation, this iterate is accepted (step 12); otherwise, we discard it (step 10). In order to make the Nesterov's acceleration strategy more efficient, an alternative choice of the momentum stepsize [51] is employed (step 10 and step 12). When $F(X_k)$ is larger than $F(V_k)$, such a scheme provides the opportunity to further exploit acceleration by enlarging the momentum β. Due to the successful application of Nesterov's acceleration technique, the number of iterations of Algorithm 1 is greatly reduced.

Now the last issue is how to choose the regularization parameter λ, which plays an important role in a regularization problem. In general, it is hard to select an optimal λ. By virtue of the idea in [36], in this paper, we tune the optimal regularization parameter at kth iteration as

$$\lambda = \frac{\sqrt[3]{108}\left(\sigma_{r_0+1}(B)\right)^{4/3}\left(\sigma_{r_0+1}(X_k) + \epsilon_{r_0+1}\right)^{2/3-\alpha}}{8\mu}, \tag{36}$$

where r_0 is the rank of the optimal solution of problem (16). Accordingly, the regularization parameter λ can be selected more adaptive and intelligent. Thus, the Algorithm 1 is free from the choice of regularization parameter during iteration.

Ha *et al. EURASIP Journal on Advances in Signal Processing* (2023) 2023:62

Page 12 of 19

3.4 Convergence analysis

In this subsection, we will discuss the convergence of our proposed algorithm. First, we introduce some definitions that will be useful in this paper.

Definition 3.2 ([52]) The Frechét subdifferential of H at x is

$$\hat{\partial}H(x) = \left\{ u : \liminf_{\substack{y \neq x \\ y \to x}} \frac{H(y) - H(x) - u^T(y - x)}{\|y - x\|_2} \geq 0 \right\},$$

where $H : \mathbb{R}^d \to (-\infty, +\infty]$ is an extended real-valued function that is proper. The limiting subdifferential of H at x is $\partial H(x) = \{u : \exists x_k \to x, H(x_k) \to H(x), \hat{\partial}H(x_k) \ni u_k \to u,$ as $k \to \infty\}$.

Definition 3.3 ([52]) x is a critical point of H iff $0 \in \partial H(x)$.

Inspired by the pioneering works in [41, 47, 49, 52], we present the following lemma, which shows that X_k satisfies a sufficient decrease condition similar to lemma 1 in [53]. Its proof is provided in the Supplementary Material.

Lemma 6 *If $\{\xi_k\}$ is a decreasing sequence and $\sum_{k=1}^K \xi_k < \infty$, we have*

$$F(X_k) \leq F(X_{k-1}) - \left(\frac{1}{2\mu} - \frac{L_f}{2} \right) \|X_k - Y_k\|_F^2 + \frac{\xi_k}{\mu}. \tag{37}$$

In what follows, we give the following theorem to show that the Algorithm 1 achieves a bounded sequence making the objective function monotonically decreasing. The proof can be found in the Supplementary Material.

Theorem 3.4 *The sequence $\{X_k\}$ is generated by Algorithm 1 with $\mu \leq 1/L_f$. If for all $k \in \mathbb{N}, \xi_k \leq \delta\|X_k - Y_k\|_F^2$, where $\delta \leq 1/2 - \mu L_f/2$. Then, we have*

(1) *$\{X_k\}$ is bounded, and has at least one limit point.*
(2) *The objective function F is monotonically decreasing.*
(3) *$\sum_{k=1}^{+\infty} \|X_k - Y_k\|_F^2 \leq +\infty$, which implies that $\lim_{k \to +\infty} \|X_k - Y_k\|_F^2 = 0$.*

4 Experiments

To illustrate the effectiveness of our proposed algorithm, in this section, we conduct two types of experiments based on the synthetic data and recommendation datasets. Specifically, we compare the proposed method with the following state-of-the-art matrix completion methods.

(1) *APGL*[11] A nuclear norm-based matrix completion method uses the accelerated proximal gradient algorithm to solve the matrix completion problem.

(2) *AIS-Impute* [13] A nuclear norm-based matrix completion method uses the accelerated and inexact soft-impute to solve the large-scale matrix completion problem.

(3) *ASD* [54] A decomposition-based method uses alternating steepest descent algorithm to solve matrix completion problem.

(4) *IRNN-TNN* [21] A nonconvex-based matrix completion method uses iteratively reweighted nuclear norm algorithm to solve matrix completion problem.

(5) *FaNCL-LSP* [22] A nonconvex-based matrix completion method uses some accelerated scheme to solve matrix completion problem.

(6) DNNR($p = 2/3$) [24] A matrix completion method based on double nonconvex nonsmooth rank relaxations.

(7) *DLRL* [25] A nonconvex based back propagation method uses multi-schatten-p norm surrogate function to solve matrix completion problem.

In the following experiments, the parameters of these algorithms are set according to the recommendations of the original paper. For our algorithm, we set $\mu = 1.95$, $\beta = (k - 1)/(k + 2)$. All the algorithms are implemented in MATLAB R2014a on a Windows server 2008 system with Intel Xeon E5-2680-v4 CPU (3 cores, 2.4 GHz) and 256 GB memory.

4.1 Synthetic data

The synthetic matrices $M \in \mathbb{R}^{m \times n}$ with rank r are generated as $M = M_L M_R + N$, where the entries of random matrices $M_L \in \mathbb{R}^{m \times r}$ and $M_R \in \mathbb{R}^{r \times n}$ are sampled i.i.d. from the standard normal distribution $N(0, 1)$, and entries of N sampled from $N(0, 0.1)$. In the following test, we set $m = n$ and $r = 5$. The symbol Ω stands for the location of observations, which are sampled uniformly at random. We let $sr = |\Omega|/mn$ to denote the sample ratio.

The performance of all algorithms is evaluated as: (i) the normalized mean squared error $NMSE = \|\mathcal{P}_{\Omega^\perp}(X - M_L M_R)\|_F / \|\mathcal{P}_{\Omega^\perp}(M_L M_R)\|_F$, where X is the recovered matrix and $\Omega^\perp$ denotes the unobserved positions; (ii) rank of X; and (iii) running time. We vary m in the range {1000, 2000, 3000, 5000}. For each algorithm, we repeat 5 times and report its average *NMSE*, rank and running time.

We report the average NMSE, rank, and running time in Table 2. As can be seen from Table 2, although all algorithms can attain satisfactory results, our proposed algorithm achieves the lowest NMSE value among all problems, which indicates that our proposed algorithm has excellent performance. In terms of accuracy, we can find that our proposed algorithm runs fastest among all algorithms. Specifically, it is 2 and 8 times faster than ASD and FaNCL-LSP, respectively. We also observe that as the matrix size increases, the advantages of our algorithm become more pronounced. In addition, for large-scale low-rank matrix completion problem, our proposed algorithm can still solve it efficiently. For example, the running time of the proposed algorithm for solving problem with $m = 10^5$, $sr = 0.12\%$ is within 1163.7 s (NMSE is smaller than 0.0140), while other algorithms cannot solve it at all or cannot get satisfactory results within this time. Therefore, taking both accuracy and converge speed into consideration, our proposed algorithm has the best recovery performance among these algorithms.

Ha *et al. EURASIP Journal on Advances in Signal Processing*   (2023) 2023:62

Page 14 of 19

Table 2 Comparison of different algorithms on synthetic data, NMSE is scaled by 10^{-2}, CPU time is in seconds, and "sr" denotes the sample ratio

	m = 1000 sr = 6.91%			m = 2000 sr = 3.80%			m = 3000 sr = 2.67%			m = 5000 sr = 1.70%		
	NMSE	Rank	Time	NMSE	Rank	Time	NMSE	Rank	Time	NMSE	Rank	Time
APGL	3.82	68	45.18	3.81	104	447.22	3.66	132	2372.02	3.49	179	21448.57
AIS-impute	3.05	5	5.09	3.61	80	259.82	3.53	80	500.96	3.33	80	1579.33
ASD	4.75	5	1.60	4.81	5	3.29	4.73	5	5.07	4.77	5	9.37
IRNN-TNN	1.83	5	92.85	1.77	5	1046.51	1.70	5	4761.42	1.67	5	23906.09
FaNCL-LSP	1.86	5	2.15	1.78	5	8.81	1.73	5	19.12	1.61	5	39.94
DNNR($p = 2/3$)	1.84	5	53.04	1.77	5	511.33	1.72	5	3115.31	1.63	5	37198.66
DLRL	1.85	5	32.87	1.78	5	170.94	1.71	5	627.53	1.61	5	15043.28
Proposed	1.83	5	0.36	1.76	5	1.3467	1.70	5	2.0432	1.61	5	5.13

Ha *et al. EURASIP Journal on Advances in Signal Processing*   (2023) 2023:62

Page 15 of 19

Table 3 Characteristics of the recommendation datasets

Dataset	Row	Column	Rating
Jester1	24983	100	10^6
Jester2	23500	100	10^6
Jester3	24983	100	6×10^5
MovieLens100K	943	1682	10^5
MovieLens1M	6040	3706	10^6

4.2 Recommendation

In this section, the Jester and MovieLens datasets will be used in our experiments to further demonstrate the effectiveness of our proposed method. There are six datasets that will be considered, namely Jester1, Jester2, Jester3, Jester-all, MovieLens-100K, and MovieLens-1 M, whose characteristics are shown in Table 3. The Jester datasets are collected from the joke recommendation system, all of which are stored in three Excel files with the following characteristics.

(1) Jester-1: 24,983 users who have rated 36 or more jokes;
(2) Jester-2: 23,500 users who have rated 36 or more jokes;
(3) Jester-3: 24,938 users who have rated between 15 and 35 jokes.

The MovieLens datasets are collected from the MovieLens website, and these datasets are characterized as follows.

(1) Movie-100K: 100,000 ratings for 1682 movies by 943 users;
(2) Movie-1 M: 1 million ratings for 3900 movies by 6040 users.

The Jester1, Jester2, and Jester3 datasets are combined to form the Jester-all dataset. In the experiment, we follow the setup in [35], which is to randomly select 50% of the observations for training and use the remaining 50% for testing. We use the root mean squared error (RMSE) and running time to evaluate the recovery performance of the algorithms. The RMSE is defned as $RMSE = \sqrt{\|\mathcal{P}_{\bar{\Omega}}(X - M)\|_F^2 / |\bar{\Omega}|_1}$, where $\bar{\Omega}$ is the test set, X is the recovered matrix. The test of each algorithm is repeated 5 times.

The recovery results regarding RMSE and running time are shown in Table 4. From Table 4, we can see that our algorithm has the best performance, that is, it achieves the smallest RMSE value in all problems. In addition, we can also find that our proposed algorithm is the fastest among all algorithms. As the size of datasets increases, some algorithms cannot get the recovery results in a satisfactory time, while our algorithm can run on all six datasets. This proves once again that our proposed algorithm has excellent performance in the field of low-rank matrix completion.

Ha *et al. EURASIP Journal on Advances in Signal Processing*   (2023) 2023:62

Page 16 of 19

Table 4 Comparison of different algorithms on Jester and MovieLens datasets, CPU time is in seconds

	Jester1		Jester2		Jester3		Jester-all		MovieLens100K		MovieLens1M	
	RMSE	Time	RMSE	Time	RMSE	Time	RMSE	Time	RMSE	Time	RMSE	Time
APGL	4.1709	157.53	4.1976	145.35	4.6358	412.55	4.2547	795.66	1.0367	1264.9	–	–
AIS-impute	4.3158	42.88	4.3471	50.04	5.2217	83.84	4.6815	474.09	1.0390	15.04	0.8880	103.44
ASD	4.6943	106.28	4.7283	102.01	5.1348	41.79	4.9944	279.94	1.1552	5.67	0.9992	56.48
IRNN-TNN	4.3257	412.07	4.3512	370.13	6.8324	274.09	4.6352	1224.6	1.1026	368.84	–	–
FaNCL-LSP	4.3210	97.91	4.3589	93.97	4.8349	38.50	4.6251	257.90	1.0871	5.22	0.9190	52.03
DNNR($p = 2/3$)	4.3031	189.12	4.3207	172.43	4.7739	568.25	4.3592	898.16	1.0543	1418.63	0.8978	–
DLRL	4.2829	144.38	4.2985	129.92	4.6941	327.84	4.3029	643.32	1.0429	529.84	0.8892	1958.76
Proposed	4.1677	1.28	4.1970	0.87	4.6552	0.92	4.2497	4.41	1.0296	0.25	0.8764	5.79

Ha *et al. EURASIP Journal on Advances in Signal Processing*   (2023) 2023:62

Page 17 of 19

5 Conclusions

In this paper, we proposed a new non-convex regularizer for low-rank minimization problems. This regularizer is better able to induce low ranks, and the resulting optimization problem has a closed-form solution. Based on the proposed regularizer, we proposed a more reasonable matrix completion model. Meanwhile, we designed an efficient optimization algorithm based on the first-order gradient method to solve the proposed model. It is simple to use and more suitable for large-scale low-rank matrix completion problems. The rationality of our proposed model and the efficiency of the algorithm is verified on a series of synthetic data and recommender system datasets. All results show that our proposed algorithm is able to achieve comparable recovery performance while being faster and more efficient than state-of-the-art methods.

Availability of data and materials
Please contact any of the authors for data and materials.

Declarations

Competing interests
The authors declare that they have no competing interests.

References

1. H. Steck, Training and testing of recommender systems on data missing not at radom, in Proc. 16th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining, Washiongton, DC, USA, Jul. (2010), pp. 713–22
2. X. Luo, M. Zhou, S. Li, Y. Xia, Q. Zhu, A non-negative latent factor model for large-scale sparse matrices in recommender systems via alternating direction method. IEEE Trans. Neural Netw. Learn. Syst. **27**(3), 579–592 (2016)
3. G. Xia, H. Sun, B. Chen, Q. Liu, L. Feng, G. Zhang, R. Hang, Nonlinear low-rank matrix completion for human motion recovery. IEEE Trans. Image Process. **27**(6), 3011–3024 (2018)
4. H. Ji, C. Liu, Z. Shen, Y. Xu, Robust video denoising using low rank matrix completion, in Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit., Jun. (2010), pp. 1791–1798
5. G. Liu, Z. Lin, S. Yan, J. Sun, Y. Yu, Y. Ma, Robust recovery of subspace structures by low-rank representation. IEEE Trans. Pattern Anal. Mach. Intell. **35**(1), 171–184 (2013)
6. Y. Xie, Y. Qu, D. Tao, W. Wu, Q. Yuan, W. Zhang, Hyperspectral image restoration via iteratively regularized weighted schatten-p norm minimization. IEEE Trans. Geosci. Remote Sens. **54**(8), 4642–4659 (2016)
7. Z. Wen, W. Yin, Y. Zhang, Solving a low-rank factorization model for matrix completion by a nonlinear successive over-relaxation algorithm. Math. Program. Comput. **4**(4), 333–361 (2012)
8. B. Recht, M. Fazel, P.A. Parrilo, Guaranteed minimum-rank solutions of linear matrix equations via nuclear norm minimization. SIAM Rev. **52**(3), 471–501 (2010)
9. E.J. Candès, B. Recht, Exact matrix completion via convex optimization. Found. Comput. Math. **9**(6), 717–772 (2009)
10. J.-F. Cai, E.J. Candès, Z. Shen, A singular value thresholding algorithm for matrix completion. SIAM J. Optim. **20**(4), 1956–1982 (2010)
11. K.C. Toh, S. Yun, An accelerated proximal gradient algorithm for nuclear norm regularized linear least squares problems. Pacific J. Optim. **6**(615–640), 15 (2010)
12. R. Mazumder, T. Hastie, R. Tibshirani, Spectral regularization algorithms for learning large incomplete matrices. J. Mach. Learn. Res. **11**, 2287–2322 (2010)
13. Q. Yao, J.T. Kwok, Accelerated inexact soft-impute for fast large-scale matrix completion, in Proceedings of 24th International Joint Conference on Artificial Intelligence, (2015), pp. 4002–4008
14. F. Nie, H. Wang, H. Huang, C. Ding, Joint schatten p-norm and l_p-norm robust matrix completion for missing value recovery. Knowl. Inf. Syst. **42**(3), 525–544 (2015)
15. T. Zhang, Analysis of multi-stage convex relaxation for sparse regularization. J. Mach. Learn. Res. **11**, 1081–1107 (2010)

16. E.J. Candès, M.B. Wakin, S.P. Boyd, Enhancing sparsity by reweighted l_1 minimization. J. Fourier Anal. Appl. **14**(5–6), 877–905 (2008)

17. J. Fan, R. Li, Variable selection via nonconcave penalized likelihood and its oracle properties. J. Am. Stat. Assoc. **96**(456), 1348–1360 (2001)

18. S. Zhang, J. Xin, Minimization of transformed L_1 penalty: theory, difference of convex function algorithm, and robust application in compressed sensing. Math. Program. **169**(1–2), 307–336 (2018)

19. Z. Wang, D. Hu, X. Luo, W. Wang, J. Wang, W. Chen, Performance guarantees of transformed Schatten-1 regularization for exact low-rank matrix recovery. Int. J. Mach. Learn. Cyber. **12**, 3379–3395 (2021)

20. J. Weston, A. Elisseeff, B. Schölkopf, M. Tipping, Use of the zero-norm with linear models and kernel methods. J. Mach. Learn. Res. **3**, 1439–1461 (2003)

21. C. Lu, J. Tang, S. Yan, Z. Lin, Nonconvex nonsmooth low rank minimization via iteratively reweighted nuclear norm. IEEE Trans. Image Process. **25**(2), 829–839 (2016)

22. Q. Yao, J.T. Kwok, T. Wang, T.-Y. Liu, Large-scale low-rank matrix learning with nonconvex regularizers. IEEE Trans. Pattern Anal. Mach. Intell **41**(11), 2628–2643 (2019)

23. F. Nie, Z. Hu, X. Li, Matrix completion based on non-convex low-rank approximation. IEEE Trans. Image Process. **28**(5), 2378–2388 (2019)

24. H. Zhang, C. Gong, J. Qian, B. Zhang, C. Xu, J. Yang, Efficient recovery of low-rank matrix via double nonconvex nonsmooth rank minimization. IEEE Trans. Neural Netw. Learn. Syst. **30**(10), 2916–2925 (2019)

25. Z. Chen, J. Yao, J. Xiao, S. Wang, Efficient and differentiable low-rank matrix completion with back propagation. IEEE Trans. Multimed. **25**, 228–242 (2023)

26. S. Gu, L. Zhang, W. Zuo, X. Feng, Weighted nuclear norm minimization with application to image denoising, in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Jun. (2014), pp. 2862–2869

27. S. Gu, Q. Xie, D. Meng, W. Zuo, X. Feng, L. Zhang, Weighted nuclear norm minimization and its applications to low level vision. Int. J. Comput. Vis. **121**(2), 183–208 (2017)

28. Y. Xie, S. Gu, Y. Liu, W. Zuo, W. Zhang, L. Zhang, Weighted schatten p-norm minimization for image denoising and background subtraction. IEEE Trans. Image Process. **25**(10), 4842–4857 (2016)

29. Y. Hu, D. Zhang, J. Ye, X. Li, X. He, Fast and accurate matrix completion via truncated nuclear norm regularization. IEEE Trans. Pattern Anal. Mach. Intell. **35**(9), 2117–2130 (2013)

30. T.H. Oh, Y.W. Tai, J.C. Bazin, H. Kim, I.S. Kweon, Partial sum minimization of sigular values in robust PCA: algorithm and applications. IEEE Trans. Pattern Anal. Mach. Intell. **38**(4), 744–758 (2016)

31. X. Su, Y. Wang, X. Kang, R. Tao, Nonconvex truncated nuclear norm minimization based on adaptive bisection method. IEEE Trans. Circuits Syst. Video Technol. **29**(11), 3159–3172 (2019)

32. Q. Liu, Z. Lai, Z. Zhou, F. Kuang, Z. Jin, A truncated nuclear norm regularization method based on weighted residual error for matrix completion. IEEE Trans. Image Process. **25**(1), 316–330 (2016)

33. C. Lee, E. Lam, Computationally efficient truncated nuclear norm minimization for high dynamic range imaging. IEEE Trans. Image Process. **25**(9), 4145–4157 (2016)

34. T. Saeedi, M. Rezghi, A novel enriched version of truncated nuclear norm regularization for matrix completion of inexact observed data. IEEE Trans. Knowl. Data Eng. **34**(2), 519–530 (2022)

35. J. Zheng, M. Qin, X. Zhou, J. Mao, H. Yu, Efficient implementation of truncated reweighting low-rank matrix approximation. IEEE Trans. Ind. Inform. **16**(1), 488–500 (2020)

36. Z. Xu, X. Chang, F. Xu, H. Zhang, $L_{1/2}$ regularization: a thresholding representation theory and a fast solver. IEEE Trans. Neural. Netw. Learn. Syst. **23**(7), 1013–1027 (2012)

37. W. Cao, J. Sun, Z. Xu, Fast image deconvolution using closed-form thresholding fomulas of $l_q(q = 1/2, 2/3)$ regularization. J. Vis. Commun. Image Represent. **24**(1), 1529–1542 (2013)

38. B. Chen, H. Sun, J. Xia, L. Feng, B. Li, Human motion recovery utilizing truncated schatten p-norm and kinematic constraints. Inf. Sci. **450**, 80–108 (2018)

39. C. Wen, W. Qian, Q. Zhang, F. Cao, Algorithms of matrix recovery based on truncated schatten p-norm. Int. J. Mach. Learn. Cyber. **12**, 1557–1570 (2021)

40. Y. Nesterov, *Introductory Lectures on Convex Optimization: A Basic Course*, vol. 87 (Springer, New York, 2013)

41. T. Sun, H. Jiang, L. Cheng, Covergence of proximal iteratively reweighted nuclear norm algorithm for image processing. IEEE Trans. Image Process. **26**(2), 5632–5644 (2017)

42. E.M. de Sá, Exposed faces and duality for symmetric and unitarily invariant norms. Linear Algebra Appl. **197**, 429–450 (1994)

43. L. Mirsky, A trace inequality of John von Neumann. Monatshefte Math. **79**(4), 303–306 (1975)

44. T.H. Oh, Y. Matsushita, Y. Tai, H. Kim, I.S. Kweon, Fast randomized singular value thresholding for low-rank optimization. IEEE Trans. Pattern Anal. Mach. Intell. **40**(2), 376–391 (2018)

45. Z. Wang, M.-J. Lai, Z. Lu, W. Fan, H. Davulcu, J. Ye, Orthogonal rank-one matrix pursuit for low rank matrix completion. SIAM J. Sci. Comput. **37**(1), A488–A514 (2015)

46. Z. Wang, Y. Liu, X. Luo, J. Wang, C. Gao, D. Peng, W. Chen, Large-scale affine matrix rank minimization with a novel nonconvex regularizer. IEEE Trans. Neural Netw. Learn. Syst. **33**(9), 4661–4675 (2022)

47. H. Li, Z. Lin, Accelerated proximal gradient methods for nonconvex programming, in Proceedings of Advances in neural information processing systems, (2015), pp. 379–387

48. S. Ghadimi, G. Lan, Accelerated gradient methods for nonconvex nonlinear and stochastic programming. Math. Program. **156**(1–2), 59–99 (2016)

49. Q. Yao, J. T. Kwok, F. Gao, W, Chen, T.-Y. Liu, Efficient inexact proximal gradient algortihm for nonconvex problems, in Proc. 26th Int. Joint Conf. Artif. Intell., Aug. (2017), pp. 3308–3314

50. B. Gu, Z. Huo, H. Huang, Inexact proximal gradient methods for nonconvex and non-smooth optimization, in Proceedings 32nd AAAI Conference on Artificial Intelligence, (2018), pp. 3093–3100

51. Q. Li, Y. Zhou, Y. Liang, P. K. Varshney, Convergence analysis of proximal gradient with momentum for nonconvex optimization, in Proceedings 34th International Conference on Machine Learning, (2017), pp. 2111–2119

52. H. Attouch, J. Bolte, B.F. Svaiter, Convergence of descent methods for semi-algebraic and tame problems: proximal algorithms, forward-backward splitting, and regularized Gauss-Seidel methods. Math. Program. **137**(1–2), 91–129 (2013)

53. P. Gong, C. Zhang, Z. Lu, J. Z. Huang, J. Ye, A general iterative shrinkage and thresholding algorithm for nonconvex regularized optimization problems, In Proceedings 30th International Conference on Machine Learning, (2013), pp. 37–45

54. J. Tanner, K. Wei, Low rank matrix completion by alternating steepest descent methods. Appl. Comput. Harmon A. **40**, 417–420 (2016)

Publisher's Note