Low-Rank Matrix Recovery via Modified Schatten-p Norm Minimization With Convergence Guarantees

Hengmin Zhang, Jianjun Qian, Bob Zhang, *Member, IEEE*, Jian Yang, *Member, IEEE*, Chen Gong, *Member, IEEE*, and Yang Wei

***Abstract*— In recent years, low-rank matrix recovery problems have attracted much attention in computer vision and machine learning. The corresponding rank minimization problems are both combinational and NP-hard in general, which are mainly solved by both nuclear norm and Schatten-p ($0 < p < 1$) norm based optimization algorithms. However, inspired by weighted nuclear norm and Schatten-p norm as the relaxations of rank function, the main merits of this work firstly provide a modified Schatten-p norm in the affine matrix rank minimization problem, denoted as the modified Schatten-p norm minimization (MS$_p$NM). Secondly, its surrogate function is constructed and the equivalence relationship with the MS$_p$NM is further achieved. Thirdly, the iterative singular value thresholding algorithm (ISVTA) is devised to optimize it, and its accelerated version, i.e., AISVTA, is also obtained to reduce the number of iterations through the well-known Nesterov's acceleration strategy. Most importantly, the convergence guarantees and their relationship with objective function, stationary point and variable sequence generated by the proposed algorithms are established under some specific assumptions, e.g., Kurdyka-Łojasiewicz (KŁ) property. Finally, numerical experiments demonstrate the effectiveness of the proposed algorithms in the matrix completion problem for image inpainting and recommender systems. It should be noted that the accelerated algorithm has a much faster convergence speed and a very close recovery precision when comparing with the proposed non-accelerated one.**

***Index Terms*— Low-rank matrix recovery, modified Schatten-p norm, iterative singular value thresholding algorithm, Kurdyka-Łojasiewicz property, convergence guarantees.**

I. Introduction

THIS paper mainly considers to solve a class of nonconvex low-rank matrix recovery problem, which can be viewed as the following regularization problem

$$\min_{\mathbf{X}} \frac{1}{2}\|\mathcal{A}(\mathbf{X}) - \mathbf{b}\|_F^2 + \lambda \text{rank}(\mathbf{X}), \tag{1}$$

where $\mathbf{X}$ and $\mathbf{b}$ are both matrices, $\lambda > 0$ is the regularized parameter and rank($\mathbf{X}$) counts the number of nonzero singular values of the desired low-rank matrix $\mathbf{X}$, $\mathcal{A}(\cdot)$ is the linear mapping and $\mathcal{A}^*(\cdot)$ stands for its adjoint. Unfortunately, Problem (1) is a challenging nonconvex optimization problem, and is known as combinational and NP-hard so that it is not easy to solve directly in general, but it has attracted much attention in numerous applications such as image inpainting [1]–[3], collaborative filtering [4], [5], recommender and minimum order systems [6]–[8], subspace clustering [9]–[11], hyperspectral imaging [12] and turbulence removal [13].

To address this issue, one can usually consider the popular nuclear norm [14]–[17] as convex relaxation of rank function in Problem (1), named as nuclear norm regularized affine matrix minimization problem. As we know, several first-order optimization algorithms (e.g., alternating direction multiplier methods (ADMMs) [18]–[20], proximal gradient algorithm and its accelerated variants [15], [21], singular values thresholding algorithm [16], [22], fixed point and bregman iteration algorithm [23], [24]) have been adopted to achieve the optimal solution, and the global convergence guarantees are also established in theory due to the existence of convexity. However, this convex substitute may produce a biased solver for recovering a real low-rank matrix due to its loose relaxation, and shrinking all of the singular values toward zero simultaneously [25]–[27] when computing the singular values thresholding operator.

To overcome this disadvantage, most nonconvex rank surrogates are proposed such as nonconvex Schatten-p norm for $0 < p < 1$ [28], [29] and truncated/weighted nuclear

Manuscript received September 22, 2018; revised April 5, 2019; accepted November 18, 2019. Date of publication December 11, 2019; date of current version January 28, 2020. This work was supported in part by the Program for Changjiang Scholars, in part by the National Science Fund (NSF) of China under Grant U1713208, Grant 61876083, Grant 61906067, Grant 61602246, and Grant 61973162, in part by the China Postdoctoral Science Foundation under Grant 2019M651415, in part by the Fundamental Research Funds for the Central Universities under Grant 30918014108 and Grant 30918011319, in part by the NSF of Jiangsu Province under Grant BK20171430, in part by the Summit of the Six Top Talents Program under Grant DZXX-027, in part by the Young Elite Scientists Sponsorship Program by Jiangsu Province, and in part by the Young Elite Scientists Sponsorship Program by China Association for Science and Technology under Grant 2018QNRC001. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Weisheng Dong. *(Corresponding authors: Jianjun Qian; Jian Yang.)*

H. Zhang, J. Qian, and J. Yang are with the PCA Laboratory, Nanjing University of Science and Technology, Nanjing 210094, China, and also with the Key Laboratory of Intelligent Perception and Systems for High-Dimensional Information of Ministry of Education, Nanjing University of Science and Technology, Nanjing 210094, China (e-mail: zhanghengmin@126.com; csjqian@njust.edu.cn; csjyang@njust.edu.cn).

B. Zhang is with the Department of Computer and Information Science, University of Macau, Macau 999078, China (e-mail: bobzhang@umac.mo).

C. Gong and Y. Wei are with the PCA Laboratory, Nanjing University of Science and Technology, Nanjing 210094, China, and also with the Jiangsu Key Laboratory of Image and Video Understanding for Social Security, School of Computer Science and Engineering, Nanjing University of Science and Technology, Nanjing 210094, China (e-mail: chen.gong@njust.edu.cn; csywei@njust.edu.cn).

This article has supplementary downloadable material available at http://ieeexplore.ieee.org, provided by the authors.

Digital Object Identifier 10.1109/TIP.2019.2957925

norm [30]–[32]. It should be stressed that Schatten-p norm is the commonly used nonconvex rank relaxed function. Additionally, other nonconvex relaxations of l_0-norm (e.g., l_p-norm [28], [33]–[35], log-sum penalty (LSP) [25], Capped-l_1 norm [36], minimax concave penalty (MCP) [37] and smoothly clipped absolute deviation (SCAD) [26]) can also be used to replace the rank function listed in [2], [3], [5], [38]. Solving these nonconvex rank relaxation problems, some first-order optimization algorithms have successfully been considered such as ADMM variants [4], [27], [30]–[32], [39], iteratively reweighted nuclear norm algorithm [2], fast low-rank matrix learning algorithms [38], [40], [41] and singular values thresholding algorithm [5], [42]. These algorithms can help to achieve a lower low-rank solver for a better recovery precision, and also guarantee that the objective function has the monotonically decreasing property and the generated variable sequence converges to a limiting point satisfying the Karush-Kuhn-Tucker (KKT) condition. However, there exist two main limitations as follows:

- The algorithms may be computationally expensive due to too much number of iterations. Involving singular values decomposition (SVD) is necessary per iteration as updating the low-rank matrix without using the decomposable matrix strategy in [7], [8], [40], [43]. Thus, this enforces us to devise faster algorithms to reduce the number of iterations for improving the computational efficiency.
- The algorithms usually do not have the global convergence guarantees, i.e., the whole sequence is a Cauchy one, and thus converges to a critical point. It plays a key role for deciding the property of the solver, so achieving this objective is of great significance in general, not only for theoretical importance, but also for practical computation as many intermediate results are usually useless without global convergence guarantees.

The main merits of this paper focus on solving the aforementioned both disadvantages to devise faster optimization algorithms and then provide the global convergence analysis. Inspired by the great success of nonconvex relaxation approaches in sparse signal and low-rank matrix recovery problems, many researchers have shown that using the l_p-norm to approximate the l_0-norm [33], [34] and the Schatten-p norm to relax the rank function [4], [28] for $0 < p < 1$ is a better choice than using l_1-norm and nuclear norm as the substitutes in most cases. It is well known that both l_p-norm and Schatten-p norm have some good advantages for relaxing l_0-norm and rank function accordingly. Hence, as one of the substitutes for rank function in Problem (1), this work concentrates on devising a modified Schatten-p norm induced by weighted l_1-norm and nuclear norm [2], [25], [32]. Several relaxations of rank function are listed in TABLE I, such as (a) Nuclear Norm, (b) Schatten-p Norm, (c) Truncated Nuclear Norm, (d) Weighted Nuclear Norm and (e) Modified Schatten-p Norm. Note that (d) is very different from both (a) and (b), and has a slight difference with (c) due to different choice of weights according to [2], [32]. Specifically, for (e), that is, the modified Schatten-p norm with $0 < p < 1$ and $\epsilon = (\epsilon_1, \epsilon_2, \ldots, \epsilon_r) \succ 0$, it can be regarded as one of the

TABLE I
SEVERAL CONVEX AND NONCONVEX RANK RELAXATIONS

//	Relaxations $1 \le i \le r = \min(\text{size}(\mathbf{X}))$
rank($\mathbf{X}$)	(a) $\|\mathbf{X}\|_* = \sum_i \sigma_i(\mathbf{X})$; (b) $\|\mathbf{X}\|_{S_p}^p = \sum_i \sigma_i^p(\mathbf{X}) = \mathrm{Tr}\left((\mathbf{X}^T\mathbf{X})^{p/2}\right)$; (c) $\|\mathbf{X}\|_{r,*} = \sum_{i=r+1} \sigma_i(\mathbf{X})$; (d) $\|\mathbf{X}\|_{\mathbf{w},*} = \sum_i \mathbf{w}_i \sigma_i(\mathbf{X})$; (e) $\|\mathbf{X}\|_{S_p,\epsilon}^p = \sum_i \frac{\sigma_i(\mathbf{X})}{(\sigma_i(\mathbf{X})+\epsilon_i)^{1-p}}$ (Proposed).

concrete examples of (d) as explained later. For these convex and nonconvex minimization problems, some first-order algorithms are developed but they usually suffer from lower computational efficiency since too much iterations may be needed to make them converge. We would like to emphasize that, the modified Schatten-p norm does not have matrix decomposable formulations like nuclear norm [7], [43], [44] and Schatten-p norm for $p = 2/3$, $1/2$ and $1/3$ [8], [45], which can decrease the computational complexity at each iteration. Different from these operations, we consider the popular Nesterov's acceleration strategy [46] to reduce the total number of iterations. Due that the involved minimization problem is nonconvex and can not be optimized directly, we present the maximization of the objective function in Problem (1) by virtue of the idea in [47], where the rank function is substituted by the modified Schatten-p norm as shown in TABLE I. The desired optimization algorithms are given to solve this problem by series of equivalence transforms, and we further analyze both local and global convergence guarantees under some milder assumptions. The main contributions of this work are listed as follows:

- We devise the modified Schatten-p norm as the rank substitute in Problem (1). By maximizing its objective function, a novel MS$_p$NM problem is obtained. Meanwhile, a series of equivalent relationships are further established through the optimal solution, and the desired optimization problem can be guaranteed to achieve the closed-form solution via the weighted singular values thresholding (WSVT) operator.
- Both ISVTA and its accelerated version, i.e., AISVTA, with the Nesterov's acceleration strategy are devised to optimize Problem (1). Moreover, we prove that the objective function decreases monotonically over the iterations and any cluster point of the generated variable sequence is a stationary point. With the help of the Kurdyka-Łojasiewicz (KŁ) property[1] and some milder conditions, we further give a global convergence guarantee of the proposed algorithms by proving that the variable sequence is a Cauchy sequence.

[1]The KŁ property was first introduced by Łojasiewicz for real analytic functions [48], and then extended by Kurdyka to smooth functions [49], and recently further extended to nonsmooth subanalytic functions [50], [51]. Furthermore, the KŁ inequality holds for many convex and nonconvex functions [3], [52] like real analytic functions, semialgebraic functions and subanalytic functions (e.g., real polynomial functions, logistic loss function, l_p $(p \ge 0)$ and Schatten-p norm). Note that the KŁ property is a powerful tool for the convergence analysis of nonconvex optimizations.

- As the specific example of Problem (1), matrix completion, several numerical experiments are conducted on synthesized data, image inpainting and recommender system. The results can show the superiority of our methods over some mostly related state-of-the-art low-rank matrix recovery problems. Besides, some prospective convergence properties and other theoretical analysis are further verified, and AISVTA does indeed reduce the number of iterations over ISVTA in the experimental settings.

The remainder of this paper is organized as follows: Section II presents the modified Schatten-p norm and formulates the MS$_p$NM problem accordingly. The equivalence relations are achieved via the optimal solution. Section III first provides the ISVTA procedure and then gives its accelerated version through the Nesterov's acceleration strategy. The choice of involved parameters are further analyzed. Section IV first introduces the basic preliminaries and then gives the main convergence results for both ISVTA and AISVTA step by step. Section V conducts several numerical experiments on both synthesized and real-world data to verify the superiority of our approaches in the matrix completion problem. We finally conclude this paper in Section VI.

II. Problem Formulation

In this section, we mainly present the definition of modified Schatten-p norm and then establish the MS$_p$NM problem. By constructing its equivalent transformation step in step, we can further achieve the desired equivalence relations according to the optimal solutions for both minimization problems.

A. Modified Schatten-p Norm

It is well known that nuclear norm is a biased approximation of rank function [26] due that it treats each singular value equally to pursue the convexity in the regularized minimization problem. Those nonconvex rank relaxations listed in TABLE I can address this issue. Based on the commonly used Schatten-p norm for $0 < p < 1$ defined in [4], [28], [29], [35], we can achieve the following relation of rank relaxations

$$\begin{aligned}\mathrm{rank}(\mathbf{X}) &= \lim_{p\to 0^+}\left[\sum_i \sigma_i^p(\mathbf{X}) = \mathrm{Tr}\left((\mathbf{X}^T\mathbf{X})^{p/2}\right)\right]\\ &= \lim_{p\to 0^+}\|\mathbf{X}\|_{S_p}^p,\end{aligned} \tag{2}$$

where we set $0^0 = 0$. It follows from (2) that the Schatten 0-norm of a matrix $\mathbf{X}$ is exactly its rank, and the Schatten 1-norm is the nuclear norm of $\mathbf{X}$. Different from them, we next define a modified Schatten p-norm, which generalizes the minimal l_p norm [53] to this case, represented by

$$\|\mathbf{X}\|_{S_p,\epsilon}^p = \sum_i \frac{\sigma_i(\mathbf{X})}{(\sigma_i(\mathbf{X})+\epsilon_i)^{1-p}}, \tag{3}$$

where $\epsilon_1 \geq \epsilon_2 \geq \ldots \geq \epsilon_r > 0$, which can make the function at the zero singular values derivable. It should be noted that the variable $\mathbf{X}$ appeared in (3) are not the same in the process of iterations as shown later, and $\|\mathbf{X}\|_{S_p,\epsilon}^p$ can be regarded as the derivative of the smoothed function[2] $\frac{1}{p}\sum_i(\sigma_i(\mathbf{X})+\epsilon_i)^p$. With the proper choices of $\epsilon_i > 0$ for any i, we have $\|\mathbf{X}\|_{S_p}^p = \lim_{\epsilon\to 0^+}\|\mathbf{X}\|_{S_p,\epsilon}^p$ induced by (2). In addition, considering that $\sigma_1(\mathbf{X}) \geq \sigma_2(\mathbf{X}) \geq \ldots \geq \sigma_r(\mathbf{X}) > 0$, it is easy to obtain the following properties

$$0 < \mathbf{w}_1 \leq \mathbf{w}_2 \leq \ldots \leq \mathbf{w}_r \text{ with } \mathbf{w}_i = \frac{1}{(\sigma_i(\mathbf{X})+\epsilon_i)^{1-p}}, \tag{4}$$

which will play a key role for computing the WSVT operators due to the nondecreasing property of the weights sequence $\{\mathbf{w}_i\}$ defined from (4). For the involved variables and parameters, the modified Schatten p-norm (3) is actually a weighted nuclear norm as previously proposed. However, there exist several differences with the mostly related rank relaxation based recovery methods stated as follows:

- [2], [5], [54] adopt $\sum_i \sigma_i^p(\mathbf{X})$ as rank relaxation for the nonconvex nonsmooth low-rank minimization problem, the supergradient is $p\sigma_i^{p-1}(\mathbf{X})$ if $\sigma_i(\mathbf{X}) > 0$, otherwise, $+\infty$ if $\sigma_i(\mathbf{X}) = 0$. To achieve the low-rank solution, [2], [5] devise the algorithms IRNN and GPG accordingly whereas [54] gives the majorization minimization algorithm as well as a weaker restricted isometry property.
- [29], [55]–[57] employ the smoothed Schatten-p norm $\mathrm{Tr}(\mathbf{X}^T\mathbf{X} + \mu^2\mathbf{I})^{p/2}$ and $\sum_i(\sigma_i^2(\mathbf{X}) + \mu^2)^{p/2}$ as rank relaxations in low-rank matrix recovery problems, their derivative are $p\mathbf{X}(\mathbf{X}^T\mathbf{X}+\mu^2\mathbf{I})^{p/2-1}$ and $p\sum_i(\sigma_i^2(\mathbf{X}) + \mu^2)^{p/2-1}$, respectively. A family of Iterative Reweighted Least Squares (IRLS) algorithms [56], [57] are provided to optimize these Schatten-p $(0 < p < 1)$ norm based nonconvex recovery problems to promote low-rankness.
- [30], [31] present the truncated nuclear norm as rank substitute and apply it to the rank minimization problem. When r is the rank number of $\mathbf{X}$, the elements of weighting vector are assigned $\mathbf{w}_i = 0$ for $i \geq r+1$, and $\mathbf{w}_i = 1$ otherwise. Additionally, [42] studies $\sum_i(\sigma_i(\mathbf{X})+\epsilon_i)^p$ in the iterative reweighted singular value minimization for image inpainting problem, and computes the weighting values by $\mathbf{w}_i = \frac{p}{(\sigma_i(\mathbf{X})+\epsilon_i)^{1-p}}$ for any i.
- [58] proposes $\sum_i \mathbf{w}_i\sigma_i^p(\mathbf{X})$ as rank substitute and generalizes the weighted nuclear norm [32] by setting $p = 1$ for image denoising. Here, $\sigma_i(\mathbf{X})$ is assigned to a nonnegative weighting value $\mathbf{w}_i = \frac{c(>0)}{\sigma_i(\mathbf{X})+\epsilon}$, which may be nonascending, nondecreasing or arbitrary order like [32]. Specifically, when $\mathbf{w}_i = 1$ holds for any i in $\sum_i \mathbf{w}_i\sigma_i^p(\mathbf{X})$ [58], it will degenerate to the Schatten-p norm [59], which can be optimized by the projected gradient descent algorithm with the convergence analysis.

In addition, to better surrogate l_0-norm and rank function for characterizing the sparsity and the low-rankness, respectively, [60] studies a novel optimization algorithm, which is formulated in the latent space for recovering a simultaneously sparse and low-rank matrix with a sufficient number of noiseless linear measurements, and [61] proposes an alternating

[2] It implies that the weights achieved in [42] is much more general since (4) is actually its special case by setting $p = 1$, and linearizing the quadratic term is used while the proposed algorithms do not.

minimization algorithm called sparse power factorization for compressed sensing of sparse rank-one matrices.

B. Modified Schatten-p Norm Minimization (MS$_p$NM)

Different from some convex and nonconvex rank relaxations, we adopt the modified Schatten-p norm as rank substitute in Problem (1) to obtain the MS$_p$NM model, denoted as

$$\min_{\mathbf{X}} \left\{ \mathcal{H}_\lambda(\mathbf{X}) = \frac{1}{2}\|\mathcal{A}(\mathbf{X}) - \mathbf{b}\|_F^2 + \lambda \|\mathbf{X}\|_{S_p,\epsilon}^p \right\}, \tag{5}$$

where the objective function $\mathcal{H}_\lambda(\mathbf{X})$ is nonconvex and does not optimize directly. Most existing first-order algorithms (e.g., IRNN [2] and ADMMs [27], [30]) can be considered to solve Problem (5) through the linearized strategy or introducing more variables. These may guarantee that each subproblem has the closed-form solution, no matter when the rank substitutes are convex or nonconvex functions (e.g., nuclear norm [14], Schatten-p norm [29], weighted/truncated nuclear norm [30]–[32] and weighted Schatten-p norm [58]). Similarly, to obtain the closed-form solution directly, this work tries to minimize the surrogate function of $\mathcal{H}_\lambda(\mathbf{X})$ by adding the quadratic terms with auxiliary variable and parameter, represented by

$$\begin{aligned}\min_{\mathbf{X},\mathbf{Y}} \Bigg\{ & \mathcal{H}_{\lambda,\mu}(\mathbf{X},\mathbf{Y}) \\ &= \mu \left[\frac{1}{2}\|\mathcal{A}(\mathbf{X}) - \mathbf{b}\|_F^2 + \lambda \sum_i \frac{\sigma_i(\mathbf{X})}{(\sigma_i(\mathbf{Y}) + \epsilon_i)^{1-p}} \right] \\ &\quad - \frac{\mu}{2}\|\mathcal{A}(\mathbf{X}) - \mathcal{A}(\mathbf{Y})\|_F^2 + \frac{1}{2}\|\mathbf{X} - \mathbf{Y}\|_F^2 \Bigg\},\end{aligned} \tag{6}$$

where both $\mathbf{X}$ and $\mathbf{Y}$ have the same size. It follows from (3), (5) and (6) that $\mathcal{H}_{\lambda,\mu}(\mathbf{X},\mathbf{X}) = \mu\mathcal{H}_\lambda(\mathbf{X})$ holds for all $\mu > 0$ with any fixed parameters $\lambda > 0$, $\epsilon \succ 0$ and $0 < p < 1$. Different from some existing first-order algorithms for solving Problem (5), one of the main merits of this paper focuses on optimizing Problem (6) efficiently as an extension of [47].

C. The Equivalence Relation of Optimal Solutions

Instead of solving Problem (5) directly to achieve the optimal solution, we will convert to compute the optimal solution of $\min_{\mathbf{X},\mathbf{Y}}\mathcal{H}_{\lambda,\mu}(\mathbf{X},\mathbf{Y})$ in Problem (6). As a result, it is necessary to guarantee that if $\mathbf{X}^*$ is the optimal solution of $\min_{\mathbf{X}}\mathcal{H}_\lambda(\mathbf{X})$, then $\mathbf{X}^*$ is also the optimal solution of $\min_{\mathbf{X}}\mathcal{H}_{\lambda,\mu}(\mathbf{X},\mathbf{X}^*)$, i.e., $\mathcal{H}_{\lambda,\mu}(\mathbf{X}^*,\mathbf{X}^*) \le \mathcal{H}_{\lambda,\mu}(\mathbf{X},\mathbf{X}^*)$ holds for any $\mathbf{X}$. By the definitions of both $\min_{\mathbf{X}}\mathcal{H}_\lambda(\mathbf{X})$ and $\min_{\mathbf{X},\mathbf{Y}}\mathcal{H}_{\lambda,\mu}(\mathbf{X},\mathbf{Y})$, it is necessary to prove

$$\begin{aligned}&\mathcal{H}_{\lambda,\mu}(\mathbf{X},\mathbf{X}^*) \\ &= \mu \left[\frac{1}{2}\|\mathcal{A}(\mathbf{X}) - \mathbf{b}\|_F^2 + \lambda \sum_i \frac{\sigma_i(\mathbf{X})}{(\sigma_i(\mathbf{X}^*) + \epsilon_i)^{1-p}} \right] \\ &\quad - \frac{\mu}{2}\|\mathcal{A}(\mathbf{X}) - \mathcal{A}(\mathbf{X}^*)\|_F^2 + \frac{1}{2}\|\mathbf{X} - \mathbf{X}^*\|_F^2 \\ &\ge \mu \left[\frac{1}{2}\|\mathcal{A}(\mathbf{X}) - \mathbf{b}\|_F^2 + \lambda \sum_i \frac{\sigma_i(\mathbf{X})}{(\sigma_i(\mathbf{X}^*) + \epsilon_i)^{1-p}} \right] \\ &\ge \mu\mathcal{H}_\lambda(\mathbf{X}^*),\end{aligned} \tag{7}$$

which implies that $\mathcal{H}_{\lambda,\mu}(\mathbf{X},\mathbf{X}^*) \ge \mathcal{H}_{\lambda,\mu}(\mathbf{X}^*,\mathbf{X}^*)$ due to the fact that $\mu\mathcal{H}_\lambda(\mathbf{X}^*) = \mathcal{H}_{\lambda,\mu}(\mathbf{X}^*,\mathbf{X}^*)$. Moreover, it also implies that $\mu\nabla_{\mathbf{X}}\mathcal{H}_\lambda(\mathbf{X}) = \nabla_{\mathbf{X}}\mathcal{H}_{\lambda,\mu}(\mathbf{X},\mathbf{X}^*)$ holds naturally for $\mathbf{X} = \mathbf{X}^*$. Actually, the first inequality holds for $0 < \mu < \frac{1}{\|\mathcal{A}\|_2^2}$, the second inequality holds due to the fact that $\mathbf{X}^*$ is the optimal solution of $\min_{\mathbf{X}}\mathcal{H}_\lambda(\mathbf{X})$. Then we further get the equivalent problem of $\min_{\mathbf{X}}\mathcal{H}_{\lambda,\mu}(\mathbf{X},\mathbf{X}^*)$ as below.

Proposition 1: For any $\lambda > 0$, $0 < \mu < \frac{1}{\|\mathcal{A}\|_2^2}$, $\mathcal{B}_\mu(\boldsymbol{X}^*) = \boldsymbol{X}^* - \mu\mathcal{A}^*(\mathcal{A}(\boldsymbol{X}^*) - \boldsymbol{b})$ and let $\boldsymbol{X}^*$ be the optimal solution of $\min_{\boldsymbol{X}}\mathcal{H}_\lambda(\boldsymbol{X})$, then $\min_{\boldsymbol{X}}\mathcal{H}_{\lambda,\mu}(\boldsymbol{X},\boldsymbol{X}^*)$ is equivalent to

$$\min_X \left\{ \frac{1}{2}\|\boldsymbol{X} - \mathcal{B}_\mu(\boldsymbol{X}^*)\|_F^2 + \lambda\mu \sum_i \frac{\sigma_i(\boldsymbol{X})}{(\sigma_i(\boldsymbol{X}^*) + \epsilon_i)^{1-p}} \right\} \tag{8}$$

where $\epsilon_1 \ge \epsilon_2 \ge \ldots \ge \epsilon_r > 0$ and $0 < p < 1$.

Proof: It follows from the definition of $\mathcal{H}_{\lambda,\mu}(\mathbf{X},\mathbf{Y})$ in Problem (6) that we can rewrite $\mathcal{H}_{\lambda,\mu}(\mathbf{X},\mathbf{X}^*)$ as

$$\begin{aligned}&\mathcal{H}_{\lambda,\mu}(\mathbf{X},\mathbf{X}^*) \\ &= \frac{1}{2}\|\mathbf{X} - \mathbf{X}^* + \mu\mathcal{A}^*\left(\mathcal{A}(\mathbf{X}^*)\right) - \mu\mathcal{A}^*(\mathbf{b})\|_F^2 + \frac{1}{2}\|\mathbf{X}^*\|_F^2 \\ &\quad + \lambda\mu \sum_i \frac{\sigma_i(\mathbf{X})}{(\sigma_i(\mathbf{X}^*) + \epsilon_i)^{1-p}} + \frac{\mu}{2}\left[\|\mathbf{b}\|_F^2 - \|\mathcal{A}(\mathbf{X}^*)\|_F^2\right] \\ &\quad - \frac{1}{2}\|\mathbf{X}^* - \mu\mathcal{A}^*\left(\mathcal{A}(\mathbf{X}^*)\right) + \mu\mathcal{A}^*(\mathbf{b})\|_F^2 \\ &\widehat{=} \frac{1}{2}\|\mathbf{X} - \mathcal{B}_\mu(\mathbf{X}^*)\|_F^2 + \lambda\mu \sum_i \frac{\sigma_i(\mathbf{X})}{(\sigma_i(\mathbf{X}^*) + \epsilon_i)^{1-p}},\end{aligned} \tag{9}$$

where $\widehat{=}$ stands for omitting the terms without $\mathbf{X}$.

Due that $\mathbf{X}^*$ is the optimal solution of $\min_{\mathbf{X}}\mathcal{H}_\lambda(\mathbf{X})$, and then combining it with (7) we can conclude that $\mathbf{X}^*$ is also the optimal solution of $\min_{\mathbf{X}}\mathcal{H}_{\lambda,\mu}(\mathbf{X},\mathbf{X}^*)$. Thus $\mathbf{X}^*$ is the optimal solution of Problem (8) according to (9). ■

It should be specifically mentioned that the equivalence relation of (5), (6) and (8) can be established for the optimal solution through (7) and Proposition 1. To further obtain the closed-form solution of Problem (8), it is actually equivalent to computing the proximity operator of the modified Schatten-p norm for any $0 < p < 1$ instead of only for the specific values like $p = 1, 2/3$ and $1/2$ though they can obtain the closed-form solutions according to [16], [62]. Due to the nondecreasing property of the weights in (4), it is easy to obtain the closed-form solution from the following Proposition 2 in despite of the existence of nonconvexity. This will play a key role for developing the optimization algorithm to solve Problem (6) instead of Problem (5) in the following section.

Proposition 2: [2], [32] For any $\tau > 0$ and the weights sequence $\{\boldsymbol{w}_i\}_{i=1}^r$ satisfies (4), let $\boldsymbol{Z} = \boldsymbol{U}\Sigma\boldsymbol{V}^T$ be the SVD of $\mathbf{Y}$, both $\boldsymbol{U}$ and $\boldsymbol{V}$ are unitary matrices, then the WSVT operator, denoted as $\boldsymbol{X}^* = \boldsymbol{U}\mathcal{S}_{\tau\boldsymbol{w}}(\Sigma)\boldsymbol{V}^T$, can be regarded as a globally optimal solution of the following problem

$$\min_X \left\{ \frac{1}{2}\|\boldsymbol{X} - \mathbf{Z}\|_F^2 + \tau \sum_i \boldsymbol{w}_i\sigma_i(\boldsymbol{X}) \right\}, \tag{10}$$

where $\mathcal{S}_{\tau\boldsymbol{w}}(\Sigma) = Diag\{(\Sigma_{ii} - \tau\boldsymbol{w}_i)_+\}$ for $i = 1, 2, \ldots, r$, and Σ_{ii} is the i-th entry of the singular values vector of $\mathbf{Z}$.

To the best of our knowledge, the minimization Problem (10) can be regarded as the WSVT operator shown in [2], [32], which can be used to optimize Problem (8) directly by setting the corresponding variables and parameters. Using the von Neumanns trace inequality [63] and the variable separable property [64], Problems (8) and (10) can be converted to compute a weighted l_1-norm thresholding operator [25]. Besides, all the singular values of $\mathcal{B}_\mu(\mathbf{X}^*)$ have a nonincreasing property, i.e., $\sigma_1(\mathcal{B}_\mu(\mathbf{X}^*)) \geq \sigma_2(\mathcal{B}_\mu(\mathbf{X}^*)) \geq \ldots \geq \sigma_r(\mathcal{B}_\mu(\mathbf{X}^*)) > 0$, and it can make the singular values of the optimal solution of Problem (8) acting on the nonnegative part.

III. The Optimization Algorithms

This section will devise both ISVTA and AISVTA for the optimizations of Problem (6). Actually, the ideas of both are derived from the majorization-minimization strategy [64], [65]. Detailed statements and explanations are provided as below.

In the position, we begin to update the variable $\mathbf{X}$, which starts with the initial variable $\mathbf{X}^0$, then for $k = 0, 1, \ldots,$ it follows from Proposition 2 that

$$\mathbf{X}^{k+1} = \mathcal{G}_{\lambda\mu, MS_pN}(\mathcal{B}_\mu(\mathbf{X}^k)) = \mathbf{U}^k \mathcal{S}_{\tau\mathbf{w}}(\Sigma^k_{\mathcal{B}_\mu})(\mathbf{V}^k)^T, \quad (11)$$

where $\mathcal{G}_{\lambda\mu, MS_pN}(\mathcal{B}_\mu(\mathbf{X}^k))$ is denoted as the WSVT of Problem (8) for the modified Schatten-p norm (3) for $0 < p < 1$. Furthermore, by setting $\tau = \lambda\mu$, $\mathbf{Z} = \mathcal{B}_\mu(\mathbf{X}^*)$ and $\mathbf{w}_i = \frac{1}{(\sigma_i(\mathbf{X}^*)+\epsilon_i)^{1-p}}$ with $0 < p < 1$, it follows from Proposition 2 that the optimal solution of Problem (8) can be easily achieved in this way. Thus the updating rule in (11) can be achieved by setting $\mathcal{B}_\mu(\mathbf{X}^k) = \mathbf{U}^k \Sigma^k_{\mathcal{B}_\mu}(\mathbf{V}^k)^T$ and $\mathcal{S}_{\tau\mathbf{w}}(\Sigma^k_{\mathcal{B}_\mu}) = \mathrm{Diag}\{(\Sigma^k_{\mathcal{B}_\mu, ii} - \tau\mathbf{w}_i)_+\}$ for $i = 1, 2, \ldots, r$.

From the empirical analysis, we get the conclusion that the proper selection of regularization parameter λ will decide the quality of the desired solution. However, it is not easy to select the optimal value of λ for the best performance. The existing iterative thresholding algorithm usually use two strategies to choose the proper regularization parameter. One way is the continuous technology [2], [5], which sets a larger value of initial λ_0, and dynamically decreases until reaching a predefined target value λ_t, i.e.,

$$\lambda^{k+1} = \kappa^k \lambda_0 \leq \lambda_t, \quad 0 < \kappa < 1. \quad (12)$$

The other is the cross-validation strtegy [47], which chooses the proper regularization parameter. To make the selection more adaptive and intelligent, we also assume that the matrix $\mathbf{X}^*$ of rank r_0 is the optimal solution of Problem (5). Then, by the nonincreasing property of the singular values of matrix $\mathcal{B}_\mu(\mathbf{X}^*)$, we have the following inequalities:

$$\begin{cases} \sigma_i(\mathcal{B}_\mu(\mathbf{X}^*)) > \dfrac{\lambda\mu}{(\sigma_i(\mathbf{X}^*)+\epsilon_i)^{1-p}}; & (13)\\[2ex] \sigma_i(\mathcal{B}_\mu(\mathbf{X}^*)) \leq \dfrac{\lambda\mu}{(\sigma_i(\mathbf{X}^*)+\epsilon_i)^{1-p}}, & (14) \end{cases}$$

Algorithm 1 Optimizations of Problem (6)

Input: $\mathbf{b}$, μ, ϵ and $0 < p < 1$.
Initialization: $\mathbf{X}^0$, take $\mathbf{T}^0 = \mathbf{X}^0$, $t_0 = 1$ and $k = 0, 1, \ldots.$

While not converged do
- $\sharp$. **ISVTA**
  - Update $\mathcal{B}_\mu(\mathbf{X}^k) = \mathbf{X}^k - \mu\mathcal{A}^*(\mathcal{A}(\mathbf{X}^k) - \mathbf{b})$;
  - Update λ_k by (12) or (17);
  - Update $\mathbf{X}^{k+1}$ by (11).
- $\sharp$. **AISVTA**
  - Update $\mathcal{B}_\mu(\mathbf{T}^k) = \mathbf{T}^k - \mu\mathcal{A}^*(\mathcal{A}(\mathbf{T}^k) - \mathbf{b})$;
  - Update λ_k by (12) or (17) with $\mathbf{T}^k \to \mathbf{X}^k$;
  - Update $\mathbf{Z}^{k+1}$ by (18);
  - Update t_{k+1} by (20);
  - Update $\mathbf{X}^{k+1}$ by (19);
  - Update $\mathbf{T}^{k+1}$ by (21).

end while

Output: $\mathbf{X}^* \leftarrow \mathbf{X}^{k+1}$.

where (13) holds for $i \in \{1, 2, \ldots, r_0\}$ and (14) holds for $i \in \{r_0+1, r_0+2, \ldots, r\}$. These further conclude that

$$\lambda \in \left[\frac{\sigma_{r_0+1}(\mathcal{B}_\mu(\mathbf{X}^*))(\sigma_{r_0+1}(\mathbf{X}^*)+\epsilon_{r_0+1})^{1-p}}{\mu}, \frac{\sigma_{r_0}(\mathcal{B}_\mu(\mathbf{X}^*))(\sigma_{r_0}(\mathbf{X}^*)+\epsilon_{r_0})^{1-p}}{\mu}\right) \quad (15)$$

satisfies due to the basic assumption that $\sigma_i(\mathcal{B}_\mu(\mathbf{X}^*))(\sigma_i(\mathbf{X}^*)+\epsilon_i)^{1-p} \leq \sigma_j(\mathcal{B}_\mu(\mathbf{X}^*))(\sigma_j(\mathbf{X}^*)+\epsilon_j)^{1-p}$ holds for $r \geq i \geq j \geq 1$ and $0 < p < 1$. Thus, we achieve $\mathcal{B}_\mu(\mathbf{X}^*)$ by $\mathcal{B}_\mu(\mathbf{X}^k)$ and $\mathbf{X}^*$ by $\mathbf{X}^k$ accordingly, and then set a proper choice of λ from (15) in the region as below

$$\lambda_k \in \left[\frac{\sigma_{r_0+1}(\mathcal{B}_\mu(\mathbf{X}^k))(\sigma_{r_0+1}(\mathbf{X}^k)+\epsilon_{r_0+1})^{1-p}}{\mu}, \frac{\sigma_{r_0}(\mathcal{B}_\mu(\mathbf{X}^k))(\sigma_{r_0}(\mathbf{X}^k)+\epsilon_{r_0})^{1-p}}{\mu}\right). \quad (16)$$

Inspired by this updating rule, in each iteration, the optimal regularization parameter λ can be tuned by

$$\lambda_k = \frac{\sigma_{r_0+1}(\mathcal{B}_\mu(\mathbf{X}^k))(\sigma_{r_0+1}(\mathbf{X}^k)+\epsilon_{r_0+1})^{1-p}}{\mu}. \quad (17)$$

Incorporated with different parameter settings and implementation variable updating rules, we can develop the iteration procedure for ISVTA to solve Problem (6). Note that (17) is valid for any $0 < \mu < \frac{1}{\|\mathcal{A}\|_2^2}$, where we set $\mu = \mu_0 = \frac{1-\eta}{\|\mathcal{A}\|_2^2}$ with any smaller $0 < \eta < 1$. Based on these results, we summarize ISVTA in Algorithm 1.

Though ISVTA can be used to solve Problem (6), it may suffer from relatively higher computational complexity because of the SVD computations for large-scale matrix computations at each iteration and more iteration steps to reach the stopping criteria. In other words, given a matrix $\mathbf{X} \in \mathbb{R}^{m\times n}$ $(m \geq n)$, its computational complexity of SVD is $o(mn^2)$. The proposed algorithm is terminated to run until the k-th iteration step. More specifically, the modified Schatten-p norm (3) has no decomposable formulations, which is different from the mostly related Schatten-p norm [8], [45], [66] due to the fact that the latter can be decomposed into the operators of two smaller matrix factors for $p = 1, 1/2$ and $2/3$ derived by Frobenius norm and/or nuclear norm. Thus reducing the number of iterations is the only feasible way for improving the computational efficiency of ISVTA. Fortunately, inspired by the popular Nesterov's acceleration strategy, both l_1-norm and nuclear norm regularized problem can be optimized by accelerated first-order algorithms in [15], [21], [67]–[71], and then nonconvex programmings can also be successfully solved by this faster strategy in [52], [69]. Thus, we try to present the accelerated version of ISVTA, i.e., AISVTA, with Nesterov's acceleration strategy for solving Problem (6) efficiently. The detailed iteration steps of AISVTA are given as follows:

$$\begin{cases} \mathbf{Z}^{k+1} = \mathcal{G}_{\lambda\mu, MS_pN}(\mathcal{B}_\mu(\mathbf{T}^k)); & (18)\\ \mathbf{X}^{k+1} = \mathrm{argmin}_{\mathbf{X}\in\{\mathbf{Z}^{k+1},\ \mathbf{X}^k\}} \mathcal{H}_{\lambda,\mu}(\mathbf{X}, \mathbf{X}^k); & (19)\\ t_{k+1} = \dfrac{1+\sqrt{1+4t_k^2}}{2}; & (20)\\ \mathbf{T}^{k+1} = \mathbf{X}^{k+1} + t_{1,k}(\mathbf{Z}^{k+1} - \mathbf{X}^{k+1}) + t_{2,k}(\mathbf{X}^{k+1} - \mathbf{X}^k) & \\ \quad \text{with } t_{1,k} = \dfrac{t_k}{t_{k+1}} \text{ and } t_{2,k} = \dfrac{t_k - 1}{t_{k+1}}, & (21) \end{cases}$$

where the Nesterov's acceleration technique has been successfully applied to AISVTA by the updating rules in (18)-(21) as summarized in Algorithm 1. Most importantly, we stress out that (19) can guarantee the following inequality

$$\mathcal{H}_{\lambda,\mu}(\mathbf{X}^{k+1}, \mathbf{X}^k) \le \min\{\mathcal{H}_{\lambda,\mu}(\mathbf{Z}^{k+1}, \mathbf{X}^k), \mathcal{H}_{\lambda,\mu}(\mathbf{X}^k, \mathbf{X}^k)\} \quad (22)$$

holds for $k = 0, 1, \ldots$ in the whole iteration scheme. To better use the Nesterov's acceleration strategy, we need to monitor and correct $\mathbf{T}^{k+1}$ when it has the potential to fail, and the monitor $\mathbf{Z}^{k+1}$ should enjoy the property of sufficient descent which is critical to ensure the objective function has the nonincreasing property. Such similar strategy can be found in the accelerated proximal gradient algorithms [52], [68]–[71]. The main reason is due to the comparisons of $\mathcal{H}_{\lambda,\mu}(\mathbf{Z}^{k+1}, \mathbf{X}^k)$ and $\mathcal{H}_{\lambda,\mu}(\mathbf{X}^k, \mathbf{X}^k)$, which can make sure sufficient descent for converging to a critical point. This can motivate us to use a WSVT step as a monitor and then guarantee that the sequence $\{\mathcal{H}_{\lambda,\mu}(\mathbf{X}^k, \mathbf{X}^{k-1})\}$ is monotonically nonincreasing and $\{\mathbf{X}^k\}$ is a desired variable sequence of the original objective function $\mathcal{H}_\lambda(\mathbf{X}^k)$ in Problem (5). These properties will play a key role in the following convergence analysis.

IV. Convergence Analysis

In this section, we first introduce the basic preliminaries in the supplementary materials through [50], [51], [72], [73] including the subdifferential, the critical point, the distance and the uniformized KŁ property. The detailed descriptions are given in Definitions 1–3 and Propositions 3–4, respectively. Combining them with the iteration rules of Algorithm 1, we further establish both local and global convergence guarantees of ISVTA and AISVTA in Theorems 1 and 2.

It should be specially noted that the KŁ inequality appearing in Definition 3 is a popular tool and is studied in the existing works [50], [51], [72] for solving a class of nonsmooth nonconvex minimization problems. Moreover, the Proposition 4 can be regarded as the uniformized KŁ property in general. The detailed proof procedures of Theorem 1 and 2 for establishing the convergence guarantees of Algorithm 1 also relies on this routine with the help of some milder conditions. This can be regarded as the main merits of this work though using the popular KŁ property, the Bolzano-Weierstrass theorem [74] and the well-known Nesterov's acceleration strategy is not novel for analysing and solving the nonconvex nonsmooth optimization problems. To the best of our knowledge, some similar results for convergence analysis in [3], [33], [52], [75] have also been verified in the generalized cases.

V. Numerical Experiments

In this section, we will conduct the numerical experiments on both synthesized and real-world data, not only to analyze the convergence property of Algorithm 1 (i.e., ISVTA and AISVTA), but also to show their efficiency and efficacy over several state-of-the-art rank relaxed matrix completion methods[3] (e.g., APGL [15], SPNM [35], TNNR [30], PSVT [31], WNNM [32], IRucLq [56], IRNN [2] and $(S+L)_{1/2}$ [45]) for solving the matrix completion problem, i.e.,

$$\min_{\mathbf{X}} \frac{1}{2}\|\mathcal{P}_\Omega(\mathbf{X}-\mathbf{M})\|_F^2 + \lambda \mathrm{rank}(\mathbf{X}), \quad (23)$$

where Ω is a subset of indexes of all pairs (i, j), defined as $[\mathcal{P}_\Omega(\mathbf{X})]_{ij} = \mathbf{X}_{ij}$ when $(i, j) \in \Omega$, and $[\mathcal{P}_\Omega(\mathbf{X})]_{ij} = 0$ when $(i, j) \notin \Omega$. As the convex and nonxonvex relaxation of NP-hard Problem (23), several rank substitutes listed in TABLE I will be considered here though other rank relaxations listed in [2], [3], [5], [38] can also be used here by the further extension. The main contents of experiments are given as follows:

- The first experiment on the synthesized data will verify the convergence property of both ISVTA and AISVTA summarized in Algorithm 1 under some different settings. The main goal is to verify the fact that AISVTA can reduce the number of iterations over ISVTA for any values of $0 < p < 1$. Due to the existence of nonconvexity, the sensitivity of Algorithm 1 for the random initialization need to be analyzed, and the effects of different settings for the values of relative error are also discussed.
- The second experiment on the natural images will show that both ISVTA and AISVTA can obtain the relatively good performance compared with other solutions for recovering the low-rank solution in image inpainting with random and text missing. Meanwhile, the evaluation criteria, e.g., Peak Signal-to-Noise Ration (PSNR) and Relative Error, are both given over the number of iterations, and the values of PSNR over different choices of $0 < p < 1$ in Problem (5).
- The third experiment on the MovieLens100K dataset will investigate the effects of the devised Algorithm 1.

[3]The optimizations almost involve SVD per iteration except SPNM, which needs the matrix inverse computations. Besides, other matrix decomposition based algorithms can also be used to solve Problem (23). However, they usually involve more variables and an estimator of rank number.

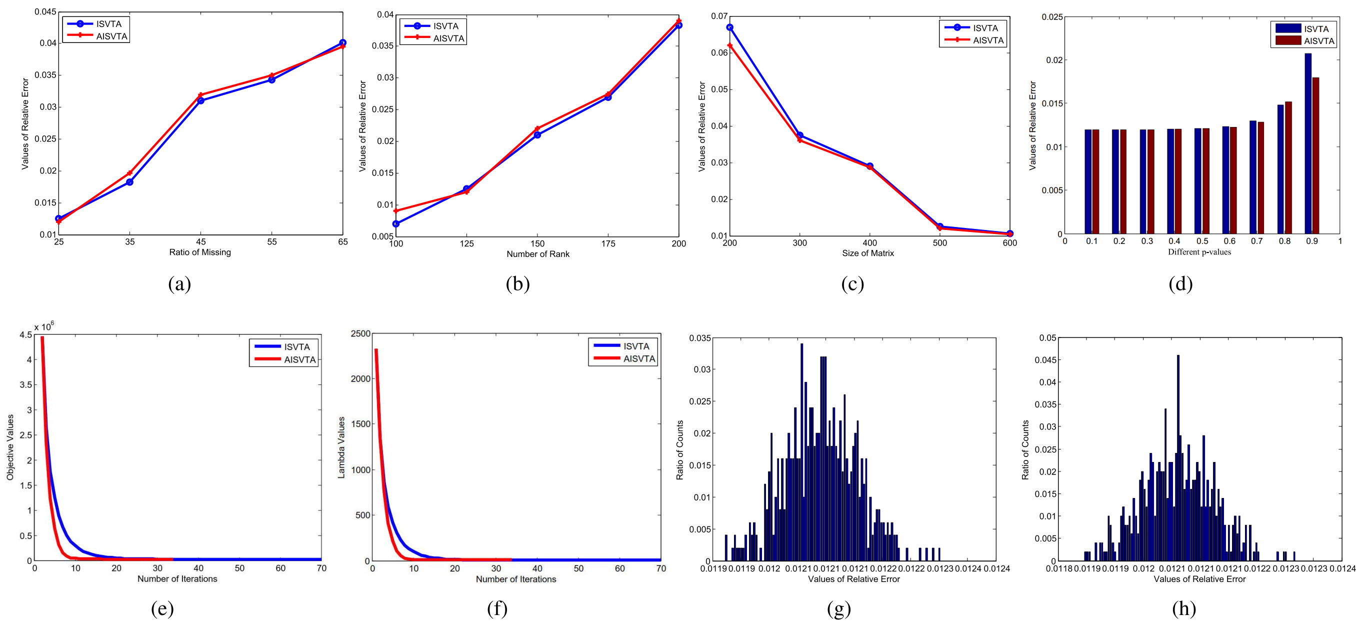


Fig. 1. Several numerical results on the synthesized data for the values of relative error vs different values for (a) l, (b) r, (c) $m = n$ and (d) p, the changes of (e) $\mathcal{H}_\lambda(\mathbf{X}^k)$ and (f) λ_k vs the number of iterations, the sensitivity to the randomly initialized values of (g) ISVTA and (h) AISVTA, respectively.

The further comparisons of time consumptions, root-mean-squared error (RMSE) [8] and normalized mean absolute error (NMAE) [15] are obtained to evaluate the testing performance in the recommender system.

All of the above-mentioned experiments are conducted by the Matlab2014a on a personal computer (PC) with 4.0GB of RAM and Intel(R) Core(TM) i3-4158 CPU@3.50GHZ. The experimental results can be achieved by tuning the involved parameters carefully of the proposed methods as well as the compared ones according to the default values of the released codes or the suggestions of the published papers.

A. Synthesized Data

Similar to the experimental settings in [32], we here verify some desired properties using the synthetic low-rank matrices. The ground truth matrix $\overline{\mathbf{X}}$ is generated by $\mathbf{A}\mathbf{B}^T$ with $\mathbf{A} \in \mathbb{R}^{m\times r}$ and $\mathbf{B} \in \mathbb{R}^{n\times r}$, respectively. All of their entries are independently sampled from the Gaussian distribution $\mathcal{N}(0, 1)$ by the Matlab command randn. The observed matrix $\mathbf{M}$ is assumed noisy, generated by $\mathcal{P}_\Omega(\mathbf{M}) = \mathcal{P}_\Omega(\overline{\mathbf{X}}) + 0.15 \times \mathbf{E}$, where $\mathbf{E}$ is also independently sampled from a Gaussian distribution. To further evaluate both ISVTA and AISVTA in this task, we next show how the performance varies when changing one of the parametric choices for (m, n, r, l). Note that we assume that (m, n) is the size of data matrix, r is the rank number and l the ratio of missing elements.

- We fix $m = n = 500$, $r = 125$, and vary l in the set $\{25\%, 35\%, 45\%, 55\%, 65\%\}$;
- We fix $m = n = 500$, $l = 25\%$, and vary r in the set $\{100, 125, 150, 175, 200\}$;
- We fix $r = 125$, $l = 25\%$, and vary $m = n$ in the set $\{200, 300, 400, 500, 600\}$.

Based on the aforementioned choice of several parameters, some typical experimental results are shown in Fig. 1, then we have the observations as follows:

- It follows from Fig. 1 (a) the smaller l-values, (b) the smaller r-values, (c) the larger of matrix size $m = n$ and (d) the smaller p-values for $p \in \{0.1, 0.2, \dots, 0.9\}$ that we can achieve the lower of relative error values, which accord to the conclusions in [4], [35].
- The nonincreasing property of the objective function $\mathcal{H}_\lambda(\mathbf{X}^k)$ in Problem (5) and regularized parameter λ_k in (17) can not only be proved in theory (see (30) of Theorem 1), but also be verified with in both (e) and (f) of Fig. 1. Such similar properties can be found in the recent works [2], [5]. Moreover, the AISVTA does indeed reduce the number of iterations over ISVTA, which shows the efficiency of the Nesterov's acceleration strategy.
- Due to the absence of convexity in the objective function of the modified Schatten-p norm, the convergence solvers, obtained by both ISVTA and AISVTA, may be different according to the initialization. Thus both (g) and (h) in Fig. 1 further study the sensitivity of the optimization against the random initialization variables. Most of solutions can be concentrically distributed in some a region near the ground-truth solution with smaller relative errors, measured by $\|\mathbf{X}^* - \overline{\mathbf{X}}\|_F / \|\overline{\mathbf{X}}\|_F$, where $\mathbf{X}^*$ is the desired low-rank solution. Note that this distribution with 500 different random initializations are considered for the proposed model and optimization algorithms.

B. Image Inpainting

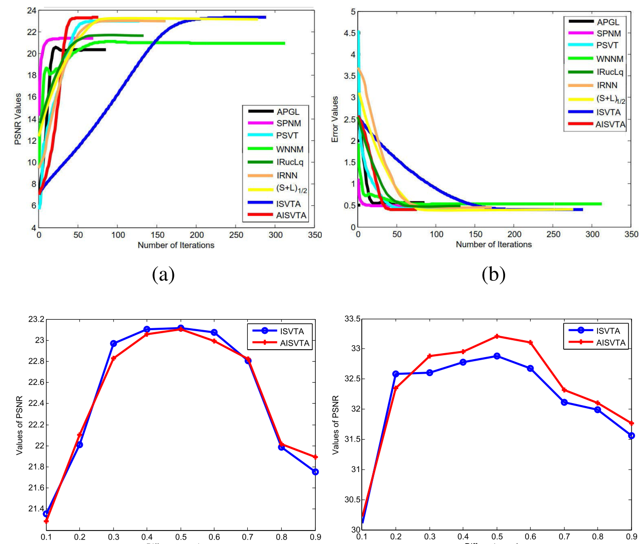
Similar to the similar settings in [30], this subsection mainly tests two incomplete types (i.e., random and text mask) on the natural images as shown in Fig. 2. All of the involved methods,

TABLE II

THE PSNR (dB) VALUES AND (ERROR VALUES, TIME CONSUMPTIONS (SECONDS)) OBTAINED BY ALL INVOLVED METHODS ON TWO NATURAL IMAGES WITH MISSING ENTRIES GENERATED BY RANDOM AND TEXT MASKS, RESPECTIVELY

//	APGL	SPNM	TNNR	PSVT	WNNM	IRucLq	IRNN	(S+L)$_{1/2}$	Ours	
									ISVTA	AISVTA
Random	20.288 (0.5577,7.3)	21.376 (0.4920,112.2)	22.697 (0.4226,44.9)	22.976 (0.4092,11.2)	20.944 (0.5171,51.1)	21.759 (0.4699,121.5)	22.953 (0.4103,56.4)	23.021 (0.4078,10.2)	**23.115** (0.4028,81.9)	<u>23.104</u> (0.4033,51.1)
Text	31.737 (0.1360,11.1)	32.412 (0.1260,69.2)	32.494 (0.1246,18.1)	32.638 (0.1228,14.7)	32.520 (0.1244,87.7)	32.613 (0.1221,72.0)	32.835 (0.1182,36.2)	32.779 (0.1197,9.3)	<u>32.880</u> (0.1175,73.8)	**33.206** (0.1148,38.5)

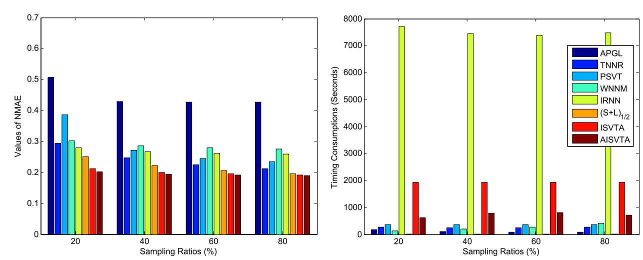
Fig. 2. The original natural images with/without random and text masks.



(a) (b)

(c) (d)

Fig. 3. The changes of values in (a) PSNR and (b) Relative Error vs the number of iterations over several related methods, and the changes of PSNR vs different p-values over random (c) and text (d) masks, respectively.



(a) (b)

Fig. 4. The values of NMAE (a) and timing consumptions (b) for all the involved methods except SPNM and IRucLq on the MovieLens100K data set.

induced by the convex and nonconvex rank substitutes listed in TABLE I, aim to to recover the missing entries of those partially damaged images. It should be noted that each of color images has three channels (i.e., red, green and blue), thus we need to recover the missing pixels by exploiting the low-rank matrix completion on each channel independently, and then combine them to get the final recovery results. The relative error values, PSNR values and time consumptions are used to evaluate the efficacy and efficiency of all the involved methods. Note that a higher value of PSNR corresponds to a lower value of error, these will indicate better recovery performance.

It follows from TABLE II that the proposed methods can obtain relatively higher or slightly same PSNR values and relative error values compared with several mostly related works. In addition, we observe that SPNM, IRucLq and ISVTA have more timing costs, while APGL, TNNR, PSVT and (S+L)$_{1/2}$ need less timing costs. Moreover, the nonconvex methods can outperform the convex method, i.e., APGL. As verified in theory, AISVTA can reduce the number of iterations over ISVTA about 3 $\sim$ 5 times in Fig. 3 (a) and (b), which can lead to less consumption costs. As an example, for the image with random mask, it should be noted that Fig. 3 (a) and (b) show the changes of values for both PSNR and relative error over the number of iterations for each involved algorithms except TNNR due to its external iterations. These results can verify the fact that the Nesterov's accelerated strategy is indeed effective for improving the computational efficiency. We here notice that AISVTA does not need to compute more SVDs than ISVTA. Further, we compare the changes of PSNR with different p-values, i.e., $p \in \{0.1, 0.2, \ldots, 0.9\}$, for two masks as shown in Fig. 3 (c) and (d), respectively.

C. Recommender System

In this subsection, we will conduct the experiments on the real world recommendation system data sets,[4] i.e., MovieLens100K with the size of (943, 1682), which contains ratings of different users on movies or music. It appears run out of memory when we trying the experiments on both MovieLens1M and MovieLens10M data sets, the reason is due to the existence of SVD computations for large-scale matrix and the memory limitation of the personal computer (PC). Thus we here neglect them. To further compare our algorithms with other aforementioned methods, we randomly sample 20%, 40%, 60% and 80% as the training set and the remaining as the testing set for the MovieLens100K data set. To avoid the influence of random initializations, the numerical results are reported over 10 independent runs in TABLE III as well as Fig. 4. Thus, the conclusions can be made as follows:

All involved methods with nonconvex cases can perform better than convex APGL. In addition, our algorithms consistently outperform the other matrix completion methods in terms of prediction accuracy evaluated by both RMSE and NMAE. This further confirms that our MS$_p$NM model

[4]http://www.grouplens.org/node/73

TABLE III
THE TESTING RMSE OF ALL INVOLVED METHODS ON THE MOVIELENS100K DATA SET WITH DIFFERENT SAMPLING RATIOS

//	MovieLens100K			
	20%	40%	60%	80%
APGL	2.3143	2.0181	2.0169	2.0163
SPNM	-	-	-	-
TNNR	1.4895	1.2804	1.1640	1.0954
PSVT	1.9688	1.4291	1.3076	1.2722
WNNM	1.5521	1.4656	1.4373	1.4260
IRucLq	-	-	-	-
IRNN	1.4406	1.3672	1.3420	1.3267
$(S+L)_{1/2}$	1.2932	1.1637	1.0820	1.0145
ISVTA	1.0928	1.0241	1.0071	0.9898
AISVTA	**1.0294**	**0.9874**	**0.9813**	**0.9718**

can provide a good estimation of a low-rank matrix. Moreover, the larger of the sampling ratio, the smaller the values of both RMSE and NMAE as can be seen for the AISVTA. This phenomenon actually has been verified in (a) and (b) of Fig. 1 and 3. The reason is due to the solver obtained by AISVTA is more close to the optimal one than ISVTA. These can imply that AISVTA slightly outperform its non-accelerated version, i.e., ISVTA. However, the consumption time of AISVTA is relatively higher than ISVTA since the former involves SVD many times. Most importantly, they can obtain the achievable performance compared with other involved methods. These are not conflicting the desired expectations. Hence, we can achieve the phenomenon for the timing consumptions and the values of NMAE for both the proposed methods and the related methods as shown in Fig. 4 (a) and (b). Also we conclude that APGL, TNNR, PSVT, WNNM and $(S+L)_{1/2}$ are the relatively faster solvers in updating the low-rank matrix through the Nesterov's technique and the decomposable strategy, while this work can achieve the closed-form solution of proximal operator according to the Proposition 2 in Section II-C.

In the testing process, it follows from TABLE III that both SPNM and IRucLq are the slowest methods in this task due to the inverse computations of matrix. Thus it is very difficult to report the experimental results of SPNM and IRucLq on more larger datasets (e.g., MovieLens1M and MovieLens10M) due to the runtime exceptions. However, $(S+L)_{1/2}$ has very good scalability and are suitable for real-world applications as verified in [8], [76]. However, the proposed methods can not obtain the final results in the task. The main reason is that iteratively solving the modified Schatten-p norm based methods for a large-scale dataset can fail to work well on the PC with limited resource due to the higher computational complexity. As a result, the proposed algorithms are not adopted to address some large-scale dataset efficiently.

VI. CONCLUSION

In this paper, we mainly studied the modified Schatten-p norm minimization based low-rank matrix recovery problem and devised the optimization algorithms entitled ISVTA and its accelerated version (i.e., AISVTA) to solve it. Meanwhile, the involved equivalence relations are established to obtain the closed-form solution using the WSVT operator instead of the commonly used linearized technique. We further prove step by step that any cluster point of the generated sequence are a stationary point. With the help of some additional constraints and the KŁ property, we can achieve the global convergence guarantees of the whole sequence generated by ISVTA and AISVTA, respectively. Finally, experimental results are performed to show the effectiveness of our algorithms in the matrix completion problem for synthesized data, image inpainting and recommender system, respectively.

In the future works, one may try the further studies, e.g., the linear mapping $\mathcal{A}(\cdot)$ of Problem (1) can be studied from linear projection [14] to random sampling [77], and some nonconvex low-rank matrix recovery problems can also be extended to the tensor case for handling 3-way color images [78].

ACKNOWLEDGMENT

The authors would like to thank the editor and the anonymous reviewers for their critical and constructive comments and suggestions.

REFERENCES

[1] E. J. Candès and Y. Plan, "Matrix completion with noise," *Proc. IEEE*, vol. 98, no. 6, pp. 925–936, Jun. 2010.
[2] C. Lu *et al.*, "Nonconvex nonsmooth low-rank minimization via iteratively reweighted nuclear norm," *IEEE Trans. Image Process.*, vol. 25, no. 2, pp. 829–839, Feb. 2016.
[3] T. Sun, H. Jiang, and L. Cheng, "Convergence of proximal iteratively reweighted nuclear norm algorithm for image processing," *IEEE Trans. Image Process.*, vol. 26, no. 12, pp. 5632–5644, Dec. 2017.
[4] F. Nie, H. Wang, H. Huang, and C. Ding, "Joint Schatten-p norm and ℓ_q-norm robust matrix completion for missing value recovery," *Knowl. Inf. Syst.*, vol. 42, no. 3, pp. 525–544, 2015.
[5] C. Lu, C. Zhu, C. Xu, S. Yan, and Z. Lin, "Generalized singular value thresholding," in *Proc. Assoc. Adv. Artif. Intell (AAAI)*, 2015, pp. 1805–1811.
[6] M. Fazel, H. Hindi, and S. P. Boyd, "A rank minimization heuristic with application to minimum order system approximation," in *Proc. Conf. Amer. Control (ACC)*, Jun. 2001, pp. 4734–4739.
[7] F. Shang, Y. Liu, H. Tong, J. Cheng, and H. Cheng, "Robust bilinear factorization with missing and grossly corrupted observations," *Inf. Sci.*, vol. 307, pp. 53–72, Jun. 2015.
[8] F. Shang, Y. Liu, and J. Cheng, "Scalable algorithms for tractable schatten quasi-norm minimization," in *Proc. Assoc. Adv. Artif. Intell (AAAI)*, 2016, pp. 2016–2022.
[9] G. Liu, Z. Lin, S. Yan, J. Sun, Y. Yu, and Y. Ma, "Robust recovery of subspace structures by low-rank representation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 1, pp. 171–184, Jan. 2013.
[10] H. Zhang, J. Yang, F. Shang, C. Gong, and Z. Zhang, "LRR for subspace segmentation via tractable schatten-p norm minimization and factorization," *IEEE Trans. Cybern.*, vol. 49, no. 5, pp. 1722–1734, May 2019.
[11] R. Vidal and P. Favaro, "Low rank subspace clustering (LRSC)," *Pattern Recognit. Lett.*, vol. 43, pp. 47–61, Jul. 2013.
[12] Y. Xie, Y. Qu, D. Tao, W. Wu, Q. Yuan, and W. Zhang, "Hyperspectral image restoration via iteratively regularized weighted schatten-p norm minimization," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 8, pp. 4642–4659, Aug. 2016.
[13] Y. Xie, W. Zhang, D. Tao, W. Hu, Y. Qu, and H. Wang, "Removing turbulence effect via hybrid total variation and deformation-guided kernel regression," *IEEE Trans. Image Process.*, vol. 25, no. 10, pp. 4943–4958, Oct. 2016.
[14] B. Recht, M. Fazel, and P. A. Parrilo, "Guaranteed minimum-rank solutions of linear matrix equations via nuclear norm minimization," *SIAM Rev.*, vol. 52, no. 3, pp. 471–501, 2010.
[15] K. C. Toh and S. Yun, "An accelerated proximal gradient algorithm for nuclear norm regularized least squares problems," *Pacific J. Optim.*, vol. 6, no. 3, pp. 615–640, 2010.

[16] J. Cai, E. Candes, and Z. Shen, "A singular value thresholding algorithm for matrix completion," *SIAM J. Optim.*, vol. 20, no. 4, pp. 1956–1982, 2010.

[17] X. Peng, C. Lu, Z. Yi, and H. Tang, "Connections between nuclear-norm and Frobenius-norm-based representations," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 29, no. 1, pp. 218–224, Jan. 2018.

[18] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Found. Trends Mach. Learn.*, vol. 3, no. 1, pp. 1–122, Jan. 2011.

[19] Z. Lin, M. Chen, L. Wu, and Y. Ma, "The augmented Lagrange multiplier method for exact recovery of corrupted low-rank matrices," 2010, *arXiv:1009.5055*. [Online]. Available: https://arxiv.org/abs/1009.5055

[20] J. Wang, C. Lu, M. Wang, P. Li, S. Yan, and X. Hu, "Robust face recognition via adaptive sparse representation," *IEEE Trans. Cybern.*, vol. 44, no. 12, pp. 2368–2378, Dec. 2014.

[21] S. Ji and J. Ye, "An accelerated gradient method for trace norm minimization," in *Proc. Int. Conf. Mach. Learn (ICML)*, 2009, pp. 457–464.

[22] T. H. Oh, Y. Matsushita, Y. W. Tai, and I. S. Kweon, "Fast randomized singular value thresholding for low-rank optimization," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 40, no. 2, pp. 376–391, Feb. 2018.

[23] S. Ma, D. Goldfarb, and L. Chen, "Fixed point and Bregman iterative methods for matrix rank minimization," *Math. Program.*, vol. 128, nos. 1–2, pp. 321–353, 2011.

[24] D. Goldfarb and S. Ma, "Convergence of fixed-point continuation algorithms for matrix rank minimization," *Found. Comput. Math.*, vol. 11, no. 2, pp. 183–210, 2011.

[25] E. J. Candès, M. B. Wakin, and S. P. Boyd, "Enhancing sparsity by reweighted L_1-minimization," *J. Fourier Anal. Appl.*, vol. 14, nos. 5–6, pp. 877–905, 2008.

[26] J. Fan and R. Li, "Variable selection via nonconcave penalized likelihood and its oracle properties," *J. Amer. Statist. Assoc.*, vol. 96, no. 456, pp. 1348–1360, 2001.

[27] H. Zhang, J. Yang, J. Qian, and W. Luo, "Nonconvex relaxation based matrix regression for face recognition with structural noise and mixed noise," *Neurocomputing*, vol. 269, no. 20, pp. 188–198, Dec. 2017.

[28] G. Marjanovic and V. Solo, "On l_q-optimization and matrix completion," *IEEE Trans. Signal Process.*, vol. 60, no. 11, pp. 5714–5724, Nov. 2012.

[29] C. Lu, Z. Lin, and S. Yan, "Smoothed low rank and sparse matrix recovery by iteratively reweighted least squares minimization," *IEEE Trans. Image Process.*, vol. 24, no. 2, pp. 646–654, Feb. 2015.

[30] Y. Hu, D. Zhang, J. Ye, X. Li, and X. He, "Fast and accurate matrix completion via truncated nuclear norm regularization," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 9, pp. 2117–2130, Sep. 2013.

[31] T. H. Oh, Y. W. Tai, J. C. Bazin, H. Kim, and I. S. Kweon, "Partial sum minimization of singular values in robust PCA: Algorithm and applications," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 38, no. 4, pp. 744–758, Apr. 2016.

[32] S. Gu, Q. Xie, D. Meng, W. Zuo, X. Feng, and L. Zhang, "Weighted nuclear norm minimization and its applications to low level vision," *Int. J. Comput. Vis.*, vol. 121, no. 2, pp. 183–208, Jan. 2017.

[33] F. Wen, P. Liu, Y. Liu, R. C. Qiu, and W. Yu, "Robust sparse recovery in impulsive noise via ℓ_1-ℓ_1 optimization," *IEEE Trans. Signal Process.*, vol. 65, no. 1, pp. 105–118, Jan. 2017.

[34] W. Zuo, D. Meng, L. Zhang, X. Feng, and D. Zhang, "A generalized iterated shrinkage algorithm for non-convex sparse coding," in *Proc. IEEE Conf. Comput. Vis. Pattern. Recogn (CVPR)*, Dec. 2013, pp. 217–224.

[35] F. Nie, H. Huang, and C. H. Ding, "Low-rank matrix recovery via efficient Schatten p-norm minimization," in *Proc. Assoc. Adv. Artif. Intell (AAAI)*, 2012, pp. 655–661.

[36] T. Zhang, "Analysis of multi-stage convex relaxation for sparse regularization," *J. Mach. Learn. Resear.*, vol. 14, nos. 5–6, pp. 1081–1107, 2010.

[37] C. H. Zhang, "Nearly unbiased variable selection under minimax concave penalty," *Ann. Statist.*, vol. 38, no. 2, pp. 894–942, 2010.

[38] Q. Yao, J. T. Kwok, and W. Zhong, "Fast low-rank matrix learning with nonconvex regularization," in *Proc. IEEE Int. Conf. Data. Mining (ICDM)*, Nov. 2015, pp. 539–548.

[39] H. Zhang, J. Yang, J. Xie, J. Qian, and B. Zhang, "Weighted sparse coding regularized nonconvex matrix regression for robust face recognition," *Inf. Sci.*, vols. 394–395, pp. 1–17, Jul. 2017.

[40] Z. Wen, W. Yin, and Y. Zhang, "Solving a low-rank factorization model for matrix completion by a nonlinear successive over-relaxation algorithm," *Math. Program. Comput.*, vol. 4, no. 4, pp. 333–361, 2012.

[41] X. Peng, R. Yan, B. Zhao, H. Tang, and Z. Yi, "Fast low rank representation based spatial pyramid matching for image classification," *Knowl.-Based Syst.*, vol. 90, pp. 14–22, Dec. 2015.

[42] Z. Lu, Y. Zhang, and J. Lu, "l_q regularized low-rank approximation via iterative reweighted singular value minimization," *Comput. Optim. Appl.*, vol. 117, no. 3, pp. 1–24, 2017.

[43] R. Cabral, F. D. L. Torre, J. P. Costeira, and A. Bernardino, "Unifying nuclear norm and bilinear factorization approaches for low-rank matrix decomposition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Dec. 2014, pp. 2488–2495.

[44] G. Liu and S. Yan, "Active subspace: Toward scalable low-rank learning," *Neural Comput.*, vol. 24, no. 12, pp. 3371–3394, 2012.

[45] F. Shang, J. Cheng, Y. Liu, Z. Luo, and Z. Lin, "Bilinear factor matrix norm minimization for robust PCA: Algorithms and applications," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 40, no. 9, pp. 2066–2080, Sep. 2018.

[46] Y. Nesterov, "A method for unconstrained convex minimization problem with the rate of convergence $o(1/k^2)$," *Nauk SSSR*, vol. 269, pp. 543–547, 1983.

[47] Z. Xu, X. Chang, F. Xu, and H. Zhang, "$L_{1/2}$ regularization: A thresholding representation theory and a fast solver," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 23, no. 7, pp. 1013–1027, Jul. 2012.

[48] S. Lojasiewicz, "Une propriété topologique des sous-ensembles analytiques réels," *Les Équations Aux Dérivées Partielles*, vol. 117, pp. 87–89, 1963.

[49] K. Kurdyka, "On gradients of functions definable in o-minimal structures," *Annales de L'Institut Fourier*, vol. 48, no. 3, pp. 769–784, 1998.

[50] J. Bolte, A. Daniilidis, and A. Lewis, "The ıojasiewicz inequality for nonsmooth subanalytic functions with applications to subgradient dynamical systems," *SIAM. J. Optim.*, vol. 17, no. 4, pp. 1205–1223, 2007.

[51] H. Attouch, J. Bolte, and 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.*, vol. 137, nos. 1–2, pp. 91–129, Feb. 2013.

[52] H. Li and Z. Lin, "Accelerated proximal gradient methods for nonconvex programming," in *Proc. Adv. Neural. Inf. Process. Syst (NIPS)*, 2015, pp. 379–387.

[53] S. Foucart and M.-J. Lai, "Sparsest solutions of underdetermined linear systems via l_q-minimization for $0 < q \leqslant 1$," *Appl. Comput. Harmon. Anal.*, vol. 26, no. 3, pp. 395–407, 2009.

[54] L. Liu, W. Huang, and D. Chen, "Exact minimum rank approximation via Schatten-p norm minimization," *J. Comput. Appl. Math.*, vol. 267, no. 6, pp. 218–227, 2014.

[55] K. Mohan and M. Fazel, "Iterative reweighted algorithms for matrix rank minimization," *J. Mach. Learn. Res.*, vol. 13, no. 1, pp. 3441–3473, 2012.

[56] M. J. Lai, Y. Xu, and W. Yin, "Improved iteratively reweighted least squares for unconstrained smoothed ℓ_q minimization," *SIAM J. Numer. Anal.*, vol. 51, no. 2, pp. 927–957, Mar. 2013.

[57] C. Kümmerle and J. Sigl, "Harmonic mean iteratively reweighted least squares for low-rank matrix recovery," *J. Mach. Learn. Res.*, vol. 19, no. 1, pp. 1815–1863, 2018.

[58] Y. Xie, S. Gu, Y. Liu, W. Zuo, W. Zhang, and L. Zhang, "Weighted schatten p-norm minimization for image denoising and background subtraction," *IEEE Trans. Image Process.*, vol. 25, no. 10, pp. 4842–4857, Oct. 2016.

[59] Y. Cai and S. Li, "Convergence analysis of projected gradient descent for Schatten-p nonconvex matrix recovery," *Sci. Chin. Math.*, vol. 58, no. 4, pp. 845–858, 2015.

[60] W. Chen, "Simultaneously sparse and low-Rank matrix reconstruction via nonconvex and nonseparable regularization," *IEEE Trans. Signal Process.*, vol. 66, no. 20, pp. 5313–5323, Oct. 2018.

[61] K. Lee, Y. Wu, and Y. Bresler, "Near-optimal compressed sensing of a class of sparse low-rank matrices via sparse power factorization," *IEEE Trans. Inf. Theory.*, vol. 64, no. 3, pp. 1666–1698, Mar. 2018.

[62] W. Cao, J. Sun, and Z. Xu, "Fast image deconvolution using closed-form thresholding formulas of l_p (2/3, 1/2) regularization," *J. Vis. Commun. Image. Represent.*, vol. 24, no. 1, pp. 31–41, 2013.

[63] D. Rhea, "The case of equality in the von Neumann trace inequality," Tech. Rep., 2011. [Online]. Available: http://www.drhea.net/wpcontent/uploads/2011/01/vonNeumann.pdf

[64] C. Lu, J. Feng, S. Yan, and Z. Lin, "A unified alternating direction method of multipliers by majorization minimization," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 40, no. 3, pp. 527–541, Mar. 2018.

[65] Z. Lin, C. Xu, and H. Zha, "Robust matrix factorization by majorization minimization," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 40, no. 1, pp. 208–220, Jan. 2017.

[66] Y. Liu, F. Shang, L. Jiao, J. Cheng, and H. Cheng, "Trace norm regularized CANDECOMP/PARAFAC decomposition with missing data," *IEEE Trans. Cybern.*, vol. 45, no. 11, pp. 2437–2448, Nov. 2015.

[67] A. Beck and M. Teboulle, "A fast iterative shrinkage-thresholding algorithm for linear inverse problems," *SIAM J. Imag. Sci.*, vol. 2, no. 1, pp. 183–202, 2009.

[68] A. Beck and M. Teboulle, "Fast gradient-based algorithms for constrained total variation image denoising and deblurring problems," *IEEE Trans. Image Process.*, vol. 18, no. 11, pp. 2419–2434, Nov. 2009.

[69] Q. Yao, J. T. Kwok, F. Gao, W. Chen, and T. Liu, "Efficient inexact proximal gradient algorithm for nonconvex problems," in *Proc. Int. Joint. Conf. Artif. Intell (IJCAI)*, 2016, pp. 3308–3314.

[70] X. Lan, S. Zhang, P. C. Yuen, and R. Chellappa, "Learning common and feature-specific patterns: A novel multiple-sparse-representation-based tracker," *IEEE Trans. Image Process.*, vol. 27, no. 4, pp. 2022–2037, Apr. 2018.

[71] X. Lan, A. J. Ma, P. C. Yuen, and R. Chellappa, "Joint sparse representation and robust feature-level fusion for multi-cue visual tracking," *IEEE Trans. Image Process.*, vol. 24, no. 12, pp. 5826–5841, Dec. 2015.

[72] J. Bolte, S. Sabach, and M. Teboulle, "Proximal alternating linearized minimization for nonconvex and nonsmooth problems," *Math. Program.*, vol. 146, nos. 1–2, pp. 459–494, 2014.

[73] R. T. Rockafellar and R. J. Wets, *Variational Analysis*, vol. 317. Berlin, Germany: Springer, 2009.

[74] J. A. Fridy, "Statistical limit points," *Proc. Amer. Math. Soc.*, vol. 118, no. 4, pp. 1187–1192, 1993.

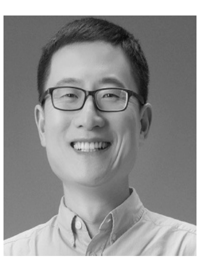
[75] Y. Wang, W. Yin, and J. Zeng, "Global convergence of ADMM in nonconvex nonsmooth optimization," *J. Sci. Comput.*, vol. 29, no. 1, pp. 29–63, 2019.

[76] F. Shang, Y. Liu, J. Cheng, and Y. Da, "Fuzzy double trace norm minimization for recommendation systems," *IEEE Trans. Fuzzy. Syst.*, vol. 26, no. 4, pp. 2039–2049, Aug. 2018.

[77] M. Rahmani and G. K. Atia, "High dimensional low rank plus sparse matrix decomposition," *IEEE Trans. Signal Process.*, vol. 65, no. 8, pp. 2004–2019, Apr. 2017.

[78] C. Lu, J. Feng, Y. Chen, L. Wei, Z. Lin, and S. Yan, "Tensor robust principal component analysis with a new tensor nuclear norm," *IEEE Trans. Pattern Anal. Mach. Intell.*, to be published, doi: 10.1109/TPAMI.2019.2891760.

Hengmin Zhang received the B.S. degree from the School of Mathematics and Statistics, Heze University, Heze, China, in 2010, and the M.S. degree from the College of Science, China University of Mining and Technology University (CUMT), Xuzhou, China, in 2013. He is currently pursuing the Ph.D. degree with the Key Laboratory of Intelligent Perception and Systems for High-Dimensional Information of Ministry of Education, School of Computer Science and Engineering, Nanjing University of Science and Technology (NUST), Nanjing, China. His research interests include nonconvex and stochastic optimization algorithms, statistical machine learning theory, and large scale representation learning methods.



Jianjun Qian received the Ph.D. degree from the Nanjing University of Science and Technology (NUST), Nanjing, China, in 2014. He is currently an Associate Professor with the Key Laboratory of Intelligent Perception and Systems for High-Dimensional Information of Ministry of Education, School of Computer Science and Engineering, NUST. His current research interests include pattern recognition, computer vision, and machine learning. He has served as Guest Editors for *Neural Processing Letter* and *The Visual Computer*, respectively. He was selected as a Hong Kong Scholar in China in 2018.

Bob Zhang (M'12) received the Ph.D. degree in electrical and computer engineering from the University of Waterloo, Waterloo, ON, Canada, in 2011. He was with the Center for Pattern Recognition and Machine Intelligence and later was a Postdoctoral Researcher with the Department of Electrical and Computer Engineering, Carnegie Mellon University, Pittsburgh, PA, USA. He is currently an Associate Professor with the Department of Computer and Information Science, University of Macau, Macau, China. His current research interests include medical biometrics, pattern recognition, and image processing.

Jian Yang (M'08) received the Ph.D. degree from the Nanjing University of Science and Technology (NUST) on the subject of pattern recognition and intelligence systems in 2002. In 2003, he was a Postdoctoral Researcher with the University of Zaragoza. From 2004 to 2006, he was a Postdoctoral Fellow with Biometrics Centre, The Hong Kong Polytechnic University. From 2006 to 2007, he was a Postdoctoral Fellow with the Department of Computer Science, New Jersey Institute of Technology. He is currently a Chang-Jiang Professor with the School of Computer Science and Technology, NUST. He is the author of more than 200 scientific articles in pattern recognition and computer vision. His articles have been cited more than 6000 times in the Web of Science, and 15000 times in the Scholar Google. His research interests include pattern recognition, computer vision, and machine learning. He is a Fellow of IAPR. Currently, he is/was an Associate Editor of *Pattern Recognition*, *Pattern Recognition Letters*, the IEEE TRANSACTION NEURAL NETWORKS AND LEARNING SYSTEMS, and *Neurocomputing*.

Chen Gong (M'16) received the B.E. degree from the East China University of Science and Technology (ECUST) in 2010, and the dual doctoral degree from Shanghai Jiao Tong University (SJTU) and the University of Technology Sydney (UTS) in 2016 and 2017, respectively. He is currently a Full Professor with the School of Computer Science and Engineering, Nanjing University of Science and Technology (NUST). He serves as the reviewer for more than 20 international journals. His research interests mainly include machine learning, data mining, and learning-based vision problems. He also serves as the SPC/PC member for several top-tier conferences. He received the Excellent Doctoral Dissertation awarded by Shanghai Jiao Tong University (SJTU) and Chinese Association for Artificial Intelligence (CAAI). He was enrolled by the Young Elite Scientists Sponsorship Program of Jiangsu Province and China Association for Science and Technology. He was a recipient of Wu Wen-Jun AI Excellent Youth Scholar Award.

Yang Wei received the B.S. degree from the School of Computer Science and Engineering, Nanjing University of Science and Technology (NUST), Nanjing, China, in 2015, where she is currently pursuing the Ph.D. degree with the Key Laboratory of Intelligent Perception and Systems for High-Dimensional Information of Ministry of Education. She is also a Visiting Researcher with Carleton University, Canada. She has published several technical articles at prominent journals, such as IEEE T-NNLS, IEEE T-CYB, and IEEE T-IP. Her research interests include pattern recognition, incomplete data based learning, and deep learning.