Efficient Recovery of Low-Rank Matrix via Double Nonconvex Nonsmooth Rank Minimization

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

***Abstract*— Recently, there is a rapidly increasing attraction for the efficient recovery of low-rank matrix in computer vision and machine learning. The popular convex solution of rank minimization is nuclear norm-based minimization (NNM), which usually leads to a biased solution since NNM tends to overshrink the rank components and treats each rank component equally. To address this issue, some nonconvex nonsmooth rank (NNR) relaxations have been exploited widely. Different from these convex and nonconvex rank substitutes, this paper first introduces a general and flexible rank relaxation function named weighted NNR relaxation function, which is actually derived from the initial double NNR (DNNR) relaxations, i.e., DNNR relaxation function acts on the nonconvex singular values function (SVF). An iteratively reweighted SVF optimization algorithm with continuation technology through computing the supergradient values to define the weighting vector is devised to solve the DNNR minimization problem, and the closed-form solution of the subproblem can be efficiently obtained by a general proximal operator, in which each element of the desired weighting vector usually satisfies the nondecreasing order. We next prove that the objective function values decrease monotonically, and any limit point of the generated subsequence is a critical point. Combining the Kurdyka–Łojasiewicz property with some milder assumptions, we further give its global convergence guarantee. As an application in the matrix completion problem, experimental results on both synthetic data and real-world data can show that our methods are competitive with several state-of-the-art convex and nonconvex matrix completion methods.**

***Index Terms*— Double nonconvex nonsmooth rank (NNR) minimization, iteratively reweighted singular values function (SVF) algorithm, low-rank matrix recovery, nuclear norm-based minimization (NNM).**

Manuscript received July 30, 2017; revised July 21, 2018 and November 27, 2018; accepted February 10, 2019. Date of publication March 18, 2019; date of current version September 18, 2019. This work was supported in part by the National Science Fund of China under Grant U1713208, Grant 61876083, Grant 61602246, and Grant 61602244, in part by the Program for Changjiang Scholars, and in part by the National Science Fund of Jiangsu Province under Grant BK20171430. *(Corresponding authors: Jianjun Qian; Jian Yang.)*

H. Zhang, C. Gong, and J. Qian are with the PCA Laboratory, 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; chen.gong@njust.edu.cn; csjqian@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. Xu and J. Yang are with the PCA Laboratory, 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: cyx@njust.edu.cn; csjyang@njust.edu.cn).

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

Color versions of one or more of the figures in this article are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TNNLS.2019.2900572

I. Introduction

LOW-RANK matrix recovery problem can be viewed as one of the most visible and challengeable tasks in computer vision and machine learning. Furthermore, low-rank matrix can be efficiently recovered by matrix factorization-based methods (see [1]–[8]) and matrix rank minimization-based methods (see [9]–[18]), respectively. Specifically, this paper only focuses on the latter category. It is well known that one of the most representative low-rank matrix relaxation functions is the tightest convex nuclear norm, which is the sum of all singular values. Under certain incoherence conditions, the low-rank matrix can be successfully recovered in a higher probability [19], [20] than some general conditions. By virtue of the guarantee in theoretical convergence, NNM problem has been solved by several first-order algorithms such as accelerated proximal gradient method [21]–[25], augmented Lagrangian multiplier method [26], and alternating direction method of multiplier [27]. However, NNM methods usually lead to the relatively lower performance (see [12], [17], [18], [28], [29]) than the nonconvex ones due that their relaxations seriously deviate from the true solution of the original low-rank matrix. In other words, the nuclear norm can lead to a biased estimator of rank function as the l_0-norm overly relaxed by the l_1-norm [30], [31].

To overcome this limitation, many nonconvex nonsmooth rank (NNR) relaxation functions have been proposed in recent years. Examples of them include Schatten p-norm $(0 < p < 1)$ [12], [13], [32], truncated nuclear norm [15], [29], weighted nuclear norm [14], and so on. Moreover, some nonconvex counterparts of the l_0-norm such as l_p-norm [33], [34], minimax concave plus (MCP) [35], smoothly clipped absolute deviation (SCAD) [30], and log-sum penalty (LSP) [31], listed in [16], [36], and [37], have been extended to relax the rank function. These derivations mainly rely on the intimate relationship between them as l_1-norm and nuclear norm. The empirically numerous results have verified that these nonconvex surrogates obtain better performance than the convex nuclear norm in general. The reason is that these nonconvex relaxations can overcome the imbalanced penalization of different singular values by keeping larger singular values larger and shrinking smaller ones, which are benefit to preserve the major information since the larger singular values are dominant, especially for a low-rank matrix. Solving these NNR problems usually need to devise the iterative optimization algorithms such as efficient Schatten p-norm minimization (SPNM) [12], weighted nuclear norm-based minimization (WNNM) [14], truncated nuclear norm regularizer (TNNR)-based methods [15], iteratively reweighted nuclear norm (IRNN) algorithm [36], and generalized proximal gradient (GPG) algorithm [16]. The aforementioned functions and algorithms have been successfully applied to solve the low-rank matrix recovery problems. However, due to the absence of convexity, developing effective solutions with the convergence guarantees becomes very challenging. In addition, the values of objective function can be proved to decrease monotonically over the increasing of iterations, and the generated subsequence converges to a critical point.

To the best of our knowledge, none of these methodologies can establish the global convergence property, i.e., the generated sequence is a Cauchy sequence, which converges to a critical point. The great importance of global convergence property is not only for the theoretical analysis but also for the practical application since the intermediate results are useless in general. Fortunately, inspired by [16], [36], and [38] and the Kurdyka–Łojasiewicz (KŁ) property[1] [39]–[41] for real analytic functions (e.g., l_2-norm of the vector and Frobenius-norm of the matrix), semialgebraic functions (e.g., l_p-norm, MCP, and SCAD), and subanalytic functions (e.g., LSP), we can develop an iteratively reweighted algorithm scheme with both local and global convergence guarantees (see *supplementary materials*). Note that the derived algorithms aim to solve the double NNR (DNNR) minimization problem, which can further deduce to the weighted NNR (WNNR) minimization problem for obtaining nearly unbiased solution. This induced function is actually a general and flexible function by the proper choices of the weighting values and the nonconvex relaxation functions in problem (3). The aforementioned NNR relaxation functions can be regarded as its special examples as stated in Section II-A. The main contributions of this paper include the following two aspects.

1) We present a general and flexible WNNR relaxation function as the rank substitute and induce the WNNR minimization problem, which is actually originated from the DNNR minimization problem. By adaptively assigning the weighting values on the NNR relaxation function like [14], [18], and [36], the DNNR minimization problem can be optimized by the general iteratively reweighted singular values function (IRSVF) algorithm procedure with a newly proximal operator for the involved weighted SVF (WSVF) problem.
2) We obtain the local convergence guarantee under some milder assumptions and further give the global convergence guarantee with the help of KŁ property. As an application in the matrix completion problem, experimental results on both synthetic data and real-world data will demonstrate the superior performance of our proposed methods.

[1]The derived KŁ inequality is a very common tool for the convergence theory of nonconvex optimization though it is not novel.

A. Outlines

Section II first points out the research preliminaries and presents the DNNR minimization problem through the predefined WNNR relaxation function. Then, we also give some milder assumptions, remarks, and definitions. Section III mainly concentrates on devising the optimization scheme and its analysis for the convergence property and the involved parameters. In addition, the general WSVF thresholding operator is introduced and then two closed-form solutions of the l_p-norm with $p = 1/2$ and $2/3$ are given through [32] and [42]. Section IV conducts several experimental comparisons with some popular methods on the matrix completion problem. Finally, we conclude this paper in Section V.

B. Notations

The set $\mathbb{R}^n$ denotes the space of n-dimensional real column vector and the set $\mathbb{R}^{p\times q}$ denotes the space of $p\times q$ dimensional real matrix. For a matrix $X \in \mathbb{R}^{p\times q}$ with $p \geq q$, its singular values decomposition (SVD) is denoted by $X = U\Sigma V^T$ with $U \in \mathbb{R}^{p\times q}$, $V \in \mathbb{R}^{q\times q}$, and $\Sigma = \mathrm{Diag}\{\sigma_{i,i=1,2,\ldots,q}\}$, where Σ is the diagonal matrix and σ_i is the ith largest singular value of X. $\sigma(X) = (\sigma_1, \sigma_2, \ldots, \sigma_q)$ denotes the singular values vector. The distance from any point $x \in \mathbb{R}^n$ to the any subset $S \in \mathbb{R}^n$ is defined as $d(x, S) = \inf_{y\in S} \|x - y\|$, and we set $d(x, S) = +\infty$ if $S = \emptyset$. In addition, the bold $\mathbf{0}$ represents the null matrix.

II. Problem Formulation

This section first points out the research preliminaries to propose the WNNR minimization problem and its original version, which can be regarded as the general extension of several rank relaxation minimization problems. Subsequently, several assumptions, remarks, and definitions are introduced for further analysis of the proposed methodology.

A. Research Preliminaries

There are some commonly used convex and NNR relaxation functions and solutions, which are used to consider the following rank regularized optimization problem:

$$\min_X\{F(X) = \lambda \mathrm{rank}(X) + g(X)\} \tag{1}$$

where $\lambda > 0$ is a regularization parameter, $\mathrm{rank}(\cdot)$ is the l_0-norm of the singular value vector, and $g(\cdot) : \mathbb{R}^{p\times q} \to \mathbb{R}^+$ is a differential loss function, which may be nonconvex. Problem (1) can cover several applications such as matrix completion [20], [43] with $g(X) = \|\mathcal{P}_\Omega(X) - \mathcal{P}_\Omega(M)\|_F^2$ and multivariate regression [44]–[46] with $g(X) = \|PX - M\|_F^2$, where $\mathcal{P}_\Omega(\cdot)$ and P are the sampling operator and the given matrix, respectively. Specifically, solving problem (1) generally involves the following rank minimization problem:

$$\min_X \ \lambda \mathrm{rank}(X) + \frac{1}{2}\|X - Y\|_F^2. \tag{2}$$

It follows from [9], [20], and [47] that solving both problems (1) and (2) directly becomes very difficult due to the discontinuous and nonconvex property of $\mathrm{rank}(X)$, which leads to the fact that they are challenging optimization problems. To solve this issue, several representative relaxation functions including convex nuclear norm [48] and nonconvex relaxations [12], [15]–[17], [36], [43] have been widely used. As we know, the NNR functions usually outperform the convex nuclear norm, especially when the desired matrix has a large rank. However, the convergence property (e.g., the global convergence) of these NNR relaxation problems is not easy to guarantee, in general, though these nonconvex functions can lead to a nearly unbiased solution. Inspired by the weighted strategy on singular values (function) [14], [18], the rank function of both problems (1) and (2) can be substituted by the WSVF of matrix $X \in \mathbb{R}^{p\times q}$ denoted by

$$\boldsymbol{\rho}_{\mathbf{w}}(\sigma(X)) = \sum_{i=1}^{r} w_i \rho(\sigma_i), \quad r = \min(p, q) \tag{3}$$

where $\rho(\cdot) : \mathbb{R}^+ \to \mathbb{R}^+$ is the proper and lower semicontinuous[2] on $[0, +\infty]$, the weighting vector $\mathbf{w} = (w_1, w_2, \ldots, w_r)$ with $0 \le w_1 \le w_2 \le \cdots \le w_r$, and the singular values vector $\sigma(X) = (\sigma_1, \sigma_2, \ldots, \sigma_r)$ with $\sigma_1 \ge \sigma_2 \ge \cdots \ge \sigma_r \ge 0$. Intuitively, each weight w_i is inversely proportional to $\rho(\sigma_i)$ since $\rho(\sigma_1) \ge \rho(\sigma_2) \ge \cdots \ge \rho(\sigma_r) \ge 0$ holds from [36]. This property can show a different importance of different singular values to achieve good performance like the weighted Schatten p-norm $(0 < p < 1)$ in [18]. It should be noted that $\rho(\cdot)$ is the sparsity function (e.g., l_p-norm) when operating on the elementwise of vector, and $\boldsymbol{\rho}_{\mathbf{w}}(\sigma(\cdot))$ of (3) is a flexible rank relaxation function with different choices of w_i and $\rho(\cdot)$ when it acts on each singular value of the low-rank matrix. For example, $\boldsymbol{\rho}_{\mathbf{w}}(\sigma(\cdot))$ becomes nuclear norm [9] and weighted nuclear norm [14] when $\rho(\cdot)$ is the absolute function, i.e., l_p-norm with $p = 1$, Schatten p-norm [12], and weighted Schatten p-norm [18] when $\rho(\cdot)$ is the l_p-norm with $0 < p < 1$ by setting all $w_i = 1$ and not all $w_i = 1$ with $i = 1, 2, \ldots, r$, respectively. In addition, $\boldsymbol{\rho}_{\mathbf{w}}(\sigma(\cdot))$ can become truncated nuclear norm [15], [29] and truncated Schatten p-norm [49] with partial $w_i = 0$ for $i = 1, 2, \ldots, r$ as well as l_p-norm with $p = 1$ and $0 < p < 1$, respectively.

B. DNNR Minimization Problem

It follows from (3) and problem (1) that a general rank relaxation minimization problem can be achieved by

$$\min_X \{F_{\mathbf{w}}(X) = \lambda \boldsymbol{\rho}_{\mathbf{w}}(\sigma(X)) + g(X)\} \tag{4}$$

where $\boldsymbol{\rho}_{\mathbf{w}}(\sigma(X))$ can lead to the nearly unbiased low-rank solution, especially when each weighting value is inversely proportional to the involved singular values, such similar conclusions can be found in [14], [18], and [31]. Moreover, problem (4) can degrade into the NNR minimization problem (see [16], [36]) for different choices of $\mathbf{w}$ and $\rho(\cdot)$ defined in (3). Motivated by the reweighted strategies [18], [36] and the supergradient concepts [50], problem (4) can be derived from the following DNNR minimization problem:

$$\min_X \left\{ F(X) = \lambda \sum_{i=1}^{r} \rho_1(\rho(\sigma_i)) + g(X) \right\} \tag{5}$$

where $\rho_1(\rho(\cdot))$ is the DNNR relaxation function. Without loss of generality, we choose $\rho_1(\cdot) = \rho(\cdot)$. Throughout this paper, both $\rho(\cdot)$ and $g(\cdot)$ satisfy the following assumptions.

A1: The penalty function $\rho(\cdot)$: $\mathbb{R}^+ \to \mathbb{R}^+$ is a proper and lower semicontinuous function on $[0, +\infty)$.

A2: The loss function $g(\cdot)$ is continuously differentiable with the Lipschitz continuous gradient $\nabla g(\cdot)$, i.e., there exists a Lipschitz constant $L_g > 0$ for any $X_1, X_2 \in \mathbb{R}^{p\times q}$, such as

$$\|\nabla g(X_1) - \nabla g(X_2)\|_F \le L_g \|X_1 - X_2\|_F. \tag{6}$$

A3: $F(\cdot)$ is coercive and bounded from below, that is,

$$\lim_{\|X\|\to+\infty} F(X) = +\infty \quad \text{and} \quad \liminf_{\|X\|\to+\infty} F(X) > -\infty. \tag{7}$$

It should be mentioned that these assumptions are usually considered for the convergence analysis of nonconvex optimization algorithms. Subsequently, we recall several remarks for some illustrations of these assumptions, definitions of subdifferential for the NNR relaxation functions, and the KŁ property for further analysis of the convergence theory.

Remark 1: It follows from A1 that the "proper" property can be guaranteed if $\emptyset \ne \mathrm{dom}\rho(\cdot) = \{x \in \mathbb{R} : g(x) < +\infty\}$, and the "lower semicontinuous" property holds at point t_0 if

$$\liminf_{x\to t_0} g(x) = g(t_0). \tag{8}$$

It follows from [16] and [36] that both gradient and subgradient are not applied to the general NNR relaxation functions. To overcome this disadvantage, the following *supplementary materials* will provide the definition of subdifferential from [51] and introduce the KŁ inequality by [39]–[41]. Especially for a proper lower semicontinuous function, we call it KŁ function if it satisfies the KŁ inequality at each point of $\mathrm{dom}\partial\rho(\cdot)$. They will play a critical role in the convergence theory of the nonconvex optimization algorithms.

III. Proposed Algorithm Scheme

This section first gives the relationship of both problems (4) and (5) through the properties of supergradients for a class of NNR relaxation functions listed in [16], [36], and [37], which satisfy the assumption A1, and second introduces the WSVF thresholding operator. Subsequently, the general IRSVF algorithm with the continuation technology is proposed according to the assumption A2. More importantly, we further give the analysis of the involved parameters. Finally, we present the convergence analysis theoretically for optimizing the concave–convex problem (5) by combining the assumptions A1–A3 with the KŁ inequality. It is especially noted that the proposed WNNR problem (4) is a more general version than some existing methods due to the flexible representation of (3), and the choice of function $\rho(\cdot)$ is not merely limited to the l_p-norm (e.g., $p = 1/2$ and $2/3$) in this paper; many other NNR relaxation functions (e.g., MCP, SCAD, and LSP) can also be adopted to substitute the rank function. However, it is not easy

[2]It is a relatively weaker property than the nonconvexity for most NNR relaxation functions, e.g., l_p-norm, MCP, SCAD, and LSP.

to claim which one is better for the rank relaxations in general. The comparisons of several commonly used NNR functions and methods can be found in [17], [36], [37], and [49].

A. Relationship of Both Problems (4) and (5)

Considering that the superdifferential of nonconvex function $\rho(\cdot)$ [50] satisfies the antimonotone property, i.e., $\langle u - v, x - y\rangle \leq 0$, for $u \in \partial\rho(x)$ and $v \in \partial\rho(y)$, this property indicates that the supergradient of any given function $\rho(\cdot)$ satisfying A1 is monotonically nondecreasing on $[0, +\infty)$, that is,

$$u \geq v, \quad \text{if and only if} \quad 0 \leq x \leq y. \tag{9}$$

Using the definition of supergradient of the function $\rho(\cdot)$ again and [28, Proposition 1], we have

$$\rho(t) \leq \rho(s) + w_s(t - s), \quad \forall\ w_s \in \partial\rho(s) \tag{10}$$

where $\rho(\cdot)$ is the nondecreasing on $[0, +\infty)$ and the singular values satisfy $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r \geq 0$, it is easy to get $\rho(\sigma_1) \geq \rho(\sigma_2) \geq \cdots \geq \rho(\sigma_r) \geq 0$ and the nonincreasing property of the supergradient value of $\rho(\cdot)$. Then, we have

Lemma 1: If $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r \geq 0$, then we have

$$0 \leq w_1 \leq w_2 \leq \cdots \leq w_r. \tag{11}$$

Furthermore, by (10) and letting $w_s \in \partial\rho(\rho(\sigma_s))$, we can get

$$\rho(\rho(\sigma_t)) \leq \rho(\rho(\sigma_s)) + w_s(\rho(\sigma_t) - \rho(\sigma_s)). \tag{12}$$

It follows from Lemma 1 that the insignificant SVF have larger weights, otherwise inverse. Thus, $\boldsymbol{\rho}_{\mathbf{w}}(\sigma(X))$ can be induced to relax the rank function in both problems (1) and (2). In addition, the relationship of problems (4) and (5) can be established by (12), which further extends the NNR function to the WNNR function such as from Schatten p-norm to weighted Schatten p-norm [18]. In addition, the supergradient may not be unique due that a nonsmooth point x may exist for the function $\rho(\cdot)$. While it is differentiable at x, the supergradient will be unique. These conclusions have been illustrated in [36, Fig. 2.].

B. WSVF Thresholding Operator

We next give the WSVF thresholding operator for solving problem (5). At first, two closed-form solutions of the l_p-norm with $p = 1/2$ and $2/3$ are given through [32] and [42].

Proposition 1: Let $\rho(\cdot) : \mathbb{R}^+ \to \mathbb{R}^+$ be a function such that the proximal operator denoted by $\text{Prox}_\rho(\cdot)$ is monotone. For any $\lambda > 0$, let $Y = U\text{Diag}(\sigma(Y))V^T$ be the SVD of $Y \in \mathbb{R}^{p\times q}$ and all weighting values satisfy $0 \leq w_1 \leq w_2 \leq \cdots \leq w_r$. Then, the optimal solution X_* to the following problem:

$$\min_X\ \lambda\boldsymbol{\rho}_{\mathbf{w}}(\sigma(X)) + \frac{1}{2}\|X - Y\|_F^2 \tag{13}$$

can be given by the general WSVF thresholding operator $X_* = U\text{Diag}(\delta^*(Y))V^T$, where $\delta^*(Y) = (\delta_1^*, \delta_2^*, \ldots, \delta_r^*)$ satisfies $\delta_i^* \geq \delta_j^*$ for $1 \leq i \leq j \leq r$, and then δ_i^* is obtained by solving the problem as follows:

$$\delta_i^* \in \text{Prox}_\rho(\sigma_i) = \text{argmin}_{\delta_i \geq 0} \lambda w_i \rho(\delta_i) + \frac{1}{2}(\delta_i - \sigma_i)^2. \tag{14}$$

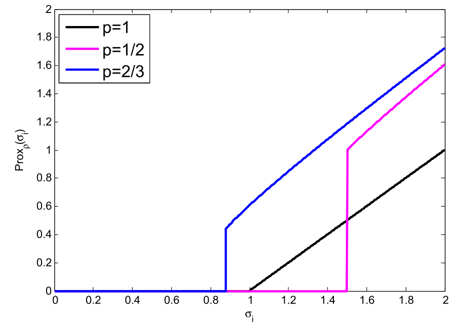


Fig. 1. Proximal operators for l_p-norm with $p = 1, 1/2$, and $2/3$.

Note that Proposition 1 can be achieved by the proof procedure of Theorem 1 [16]. Thus, it is not repeated here. However, there exist some differences to the rank relaxations that problem (13) is more general and extends the nonconvex relaxations to its weighted version. Moreover, using the von Neumann's trace inequality [52] and the separable property, problem (13) can be reformulated into an equivalent formulation with singular values (14), which can be solved by the fixed-point iteration algorithm [16]. Different from the iteratively thresholding solution [53], we consider to get the closed-form solution of (14) for the special choices of $\rho(\cdot)$ such as l_p-norm with $p = 1, 2/3$, and $1/2$, respectively. It can be shown in Fig. 1. The reason for choosing the popular l_p-norm is due that it can be extended to the Schatten p-norm, which has some inspiring properties [54], [55]. Other nonconvex functions listed in [36], [37], and [49] satisfying A1 can also be used here, especially when they have the closed-form solutions. For notational simplicity, we write λw_i as ξ, σ_i as σ, and δ_i as δ, respectively. To achieve the optimal solutions of problem (13), we compute the closed-form solutions of (14) as follows.

1) l_p-norm with $p = 1/2$, then (14) becomes

$$\delta^* \in \text{argmin}_{\delta\geq 0}\ \xi\delta^{\frac{1}{2}} + \frac{1}{2}(\delta - \sigma)^2. \tag{15}$$

It follows from [32], [42], and [56] that (15) has the following closed-form solution denoted by:

$$\delta^* = \begin{cases} \frac{2}{3}\sigma(1 + \cos(\frac{2\pi}{3} - \frac{2\phi(\sigma)}{3})), & \sigma > \varphi(\xi) \\ 0, & \text{otherwise} \end{cases} \tag{16}$$

where $\phi(\sigma) = \arccos(\xi/4(\sigma/3)^{-3/2})$ and $\varphi(\xi) = 3\sqrt[3]{2}/4(2\xi)^{2/3}$.

2) l_p-norm with $p = 2/3$, then (14) becomes

$$\delta^* \in \text{argmin}_{\delta\geq 0}\ \xi\delta^{\frac{2}{3}} + \frac{1}{2}(\delta - \sigma)^2. \tag{17}$$

It follows from [32], [42], and [56] that (17) has the following closed-form solution denoted by:

$$\delta^* = \begin{cases} ((\varpi + \sqrt{2\sigma/\varpi - \varpi^2})/2)^3 & \sigma > \varphi(\xi), \\ 0, & \text{otherwise} \end{cases} \tag{18}$$

where $\varpi = 2/3^{1/2}(2\xi)^{1/4}\cosh(\text{arccosh}(27\sigma^2/16(2\xi)^{-1.5})/3)^{1/2}$ and $\varphi(\xi) = 2/3(3(2\xi)^3)^{1/4}$.

By both (16) and (18) and using the relationship between (13) and (14), it is easy to obtain the closed-form solution of (13) for the l_p-norm with $p = 1/2$ and $2/3$, respectively. The assumption A1 can guarantee that the monotone property of $\text{Prox}_\rho(\cdot)$ holds for any lower bounded function $\rho(\cdot)$ with the help of [16], and this property plays a key role for making (13) separable to reformulate (14). Although we give the general thresholding operator, the unified formulation of solutions for (14) cannot be easy to obtain for the general NNR relaxation functions listed in [36], [37], and [49].

C. IRSVF Algorithm

Using the relationship of problems (4) and (5), this section will derive a detailed optimization scheme named IRSVF algorithm to solve problem (4). We first linearize $g(X)$ at $X_k \in \mathbb{R}^{p\times q}$ and add a proximal term, then obtain

$$g(X) \approx g(X_k) + \langle \nabla g(X_k), X - X_k\rangle + \frac{\mu}{2}\|X - X_k\|_F^2 \quad (19)$$

where μ is larger than the Lipschitz constant L_g, i.e., $\mu > L_g$. Such a proper choice of μ is very important to guarantee the algorithmic convergence. If L_g is unknown or uncomputable, the backtracking rule can be used to estimate μ in each iteration [57]. Inspired by both IRNN [36] and GPG [16], the right-hand side of (19) can be used to substitute $g(X)$ in problem (4). Then, X_{k+1} can be updated by

$$\begin{aligned} X_{k+1} = \text{argmin}_X\ & \lambda \boldsymbol{\rho}_{\mathbf{w}^k}(\sigma(X)) + \frac{\mu}{2}\|X - X_k\|_F^2 \\ & + \langle \nabla g(X_k), X - X_k\rangle + g(X_k). \end{aligned} \quad (20)$$

It follows from (3) that (20) can be written as

$$X_{k+1} = \text{argmin}_X\ \lambda \sum_{i=1}^{r} w_i^k \rho(\sigma_i(X)) + \frac{\mu}{2}\|X - Y_k\|_F^2 \quad (21)$$

where $Y_k = X_k - 1/\mu \nabla g(X_k)$. It is obvious that (21) can be solved by the Proposition 1. After updating X_{k+1}, we need to compute the weighting vector $\mathbf{w}^{k+1}$ by

$$w_i^{k+1} \in \partial\rho(\rho(\sigma_i(X_{k+1}))), \quad i = 1, 2, \ldots, r. \quad (22)$$

Actually, the above-mentioned optimization process can fall into the majorization–minimization algorithm scheme [38], [58], [59]. In detail, it includes iteratively updating X_{k+1} through solving a WSVF thresholding operator by the gradient step (19) and proximal step (21) and then updating the weighting vector $\mathbf{w}^k$ (22) by computing the supergradients of nonconvex relaxations, which is similar to both IRNN and GPG. However, there exist two main differences as follows.

1) Both IRNN and GPG can be regarded as the special cases of the proposed IRSVF algorithm by the proper choices of weighting vector and NNR functions. Thus, our proposed method and algorithm is more general than them.

2) Iteratively updating X_{k+1} and weighting vector $\mathbf{w}^k$ in IRNN mainly depends on the singular values $\sigma_i(\cdot)$'s, while our IRSVF algorithm is based on the SVF $\rho(\sigma_i(\cdot))$'s. Thus, the general thresholding operator to the WNNR minimization problem (21) can be obtained, which is also different from the proximal step in both IRNN and GPG. Specifically, it is not easy to give the general representation of solutions for (21) by (14) in general.

Algorithm 1 IRSVFc for Solving DNNR Problem

Input: $\lambda, \lambda_{\min} > 0$, $\mu > L_g$, $0 < \tau < 1$ and $\varepsilon > 0$.
Initialization: $k = 0$, $X_k \in \mathbb{R}^{p\times q}$ and $\mathbf{w}^k \in \mathbb{R}^r$.

A. **Outer loop**
1. **for** $j = 0, 1, 2...$ do
2. $k = 0$
B. **Inner loop**
3. Compute Y_k by $X_k - \frac{1}{\mu}\nabla g(X_k)$;
4. Update X_{k+1} by solving (21);
5. Update $\mathbf{w}^{k+1}$ by computing (22);
6. **if** $\frac{|F_{\lambda_k}(X_{k+1}) - F_\lambda(X_k)|}{|F_{\lambda_k}(X_{k+1})|} > \tau \cdot \lambda_k$,
7. then $k = k + 1$, Return to 3;
8. **else**
9. Update λ by $\lambda = \tau \cdot \lambda_{k+1}$, then $j = j + 1$;
10. Update X by $X_k = X_{k+1}$;
11. **until convergence**

Output: $X_* \leftarrow X_{k+1}$.

D. Adjustment of the Involved Parameters

Similar to the devised strategy in both algorithms IRNN and GPG, the continuation technique is used to enhance the quality of low-rank solution and accelerate the convergence speed of the proposed IRSVF algorithm. Moreover, we empirically observe that the obtained solution is very sensitive to the regularizer parameter λ, and the proper choice of λ plays a key role in determining the recovery ability of low-rank matrix. It involves an initial value λ_0 and a target parameter $\lambda_{\min}$ and dynamically decreases it over the increasing iteration numbers by $\lambda = \tau^k \cdot \lambda_0$ with the reduction factor $0 < \tau < 1$ until reaching the predefined target value, i.e., $\lambda \le \lambda_{\min}$, which determines the stopping tolerance. In addition, the larger values of τ, the more timing costs and iterations number in the experiments. This updating formula for λ also follows the fact that a larger value on $\lambda_{\min}$ in each iteration will lead to the lower relative error (RE) values, while a smaller value on $\lambda_{\min}$ will have the opposite effect. Such conclusions have been verified in the related methods [16], [36]. We summarize the mainly iterated optimization procedure of the IRSVF algorithm with the continuation technique (IRSVFc) in Algorithm 1.

E. Convergence Analysis

As we know, both the convergence theory and the computational complexity are the critical evaluation criteria for the first-order optimization algorithms. In the derived Algorithm 1, the computational complexity mainly depends on the computations of SVD. More importantly, under some milder assumptions, we focus on establishing the convergence guarantees from local to global of the devised algorithm. The main results can be found in the *supplementary materials*, and they can provide the theoretical support for the practical applications in computer vision and machine learning.

TABLE I

RE VALUES OF DIFFERENT LOW-RANK MATRIX COMPLETION METHODS ON THE SYNTHETIC DATA UNDER THE PARAMETERS CHOICE IN $\{p_r, p_m, n_l\}$

Setting	//	APGL	TNNR	SPNM	WNNM	l_p $(p = 1/2)$			l_p $(p = 2/3)$		
						IRNN	GPG	DNNR	IRNN	GPG	DNNR
(p_r,0.50,0.50,300)	0.15	1.850e-1	**8.762e-2**	9.603e-2	1.113e-1	9.285e-2	9.078e-2	8.870e-2	9.403e-2	9.123e-2	9.054e-2
	0.20	2.714e-1	1.152e-1	1.379e-1	1.181e-1	1.137e-1	1.083e-1	**1.007e-1**	1.245e-1	1.115e-1	1.112e-1
	0.25	3.760e-1	2.734e-1	2.256e-1	2.847e-1	2.770e-1	2.323e-1	**1.913e-1**	2.310e-1	2.205e-1	2.029e-1
	0.30	5.473e-1	5.238e-1	3.548e-1	4.864e-1	4.281e-1	3.915e-1	3.736e-1	3.847e-1	3.786e-1	**3.532e-1**
(0.30,p_m,0.50,300)	0.10	1.137e-1	5.822e-2	5.906e-2	5.923e-2	5.606e-2	5.541e-2	5.515e-2	5.520e-2	5.515e-2	**5.447e-2**
	0.20	1.382e-1	8.052e-2	7.512e-2	6.498e-2	6.113e-2	6.098e-2	**5.998e-2**	6.155e-2	6.068e-2	6.022e-2
	0.30	1.934e-1	1.785e-1	1.098e-1	7.666e-2	7.950e-2	7.692e-2	**7.291e-2**	8.444e-2	7.836e-2	7.810e-2
	0.40	3.091e-1	2.381e-1	1.917e-1	1.783e-1	1.892e-1	1.702e-1	**1.329e-1**	1.759e-1	1.554e-1	1.501e-1
(0.10,0.50,n_l,300)	0.10	7.194e-2	2.717e-2	3.478e-2	2.226e-2	2.051e-2	1.942e-1	**1.615e-2**	1.943e-2	2.201e-2	2.057e-2
	0.20	7.398e-2	4.559e-2	4.251e-2	4.544e-2	4.280e-2	3.738e-2	**3.118e-2**	4.080e-2	3.863e-2	3.311e-2
	0.30	8.276e-2	6.413e-2	5.725e-2	6.777e-2	6.442e-2	5.448e-2	4.994e-2	6.150e-2	5.487e-2	**4.933e-2**
	0.40	9.452e-2	9.024e-2	7.288e-2	9.165e-2	8.672e-2	7.275e-2	6.883e-2	8.253e-2	7.410e-2	**6.674e-2**

IV. EXPERIMENTS

This section will compare the efficacy and effectiveness of our method with other approaches on both synthetic and real-world data for the matrix completion problem

$$\min_X \lambda \operatorname{rank}(X) \ + \frac{1}{2}\|\mathcal{P}_\Omega(X) - \mathcal{P}_\Omega(M)\|_F^2 \qquad (23)$$

where $\mathcal{P}_\Omega(\cdot)$ represents the linear projection operator, i.e., $\mathcal{P}_\Omega(M)_{ij} = M_{ij}$ if $(i, j) \in \Omega$ and $\mathcal{P}_\Omega(M)_{ij} = 0$ otherwise. All the algorithms and experiments are implemented by MATLAB code on a PC with 4.00 GB of RAM and Intel Core i3-4150 CPU @ 3.50 GHz. The parameter choices of the compared methods rely on the authors' suggestions of the published papers or the default parameters of the released codes to obtain the best performance, respectively.

The involved matrix completion approaches are mainly based on convex and nonconvex rank relaxations as follows: 1) the convex NNM and 2) the NNR-based minimization for Schatten p-norm, truncated nuclear norm, and weighted nuclear norm. In addition, the nonconvex relaxation functions of l_0-norm listed in [16] and [36] have been extended to approximate the rank function in problems (1), (2), and (23). Such these convex and nonconvex problems mentioned earlier have been solved by several state-of-the-art algorithms such as Accelerated Proximal Gradient Line Search (APGL) [3] [21], SPNM[4] [12], WNNM[5] [14], TNNR[6] [15], IRNN[7] [36], and GPG [16], which are tested as the compared matrix completion methods. In addition, in the *supplementary materials*, we provide the experimental comparisons with the matrix factorization-based matrix completion methods, which mainly decompose the large-scale matrix into two or three small matrices to reduce the computational complexity. Note that, for the compared methods such as IRNN, GPG, and DNNR, the l_p-norm with $p = 1/2$ and $2/3$ are extended to relax the rank function in their respective settings. Moreover, the closed-form solutions for these cases are essential for IRNN, GPG, and DNNR algorithms when computing the proximal operators. The derived solutions mainly depend on (15) and (17). This is not different from the fixed-point iteration thresholding solver for the GPG algorithm [16].

[3]http://www.math.nus.edu.sg/mattohkc/NNLS.html

[4]http://www.escience.cn/people/fpnie/papers.html

[5]http://www.comp.polyu.edu.hk/cslzhang/

[6]http://sites.google.com/site/zjuyaohu/

[7]http://www.escience.cn/people/CanyiLu/index.html

A. Synthetic Data

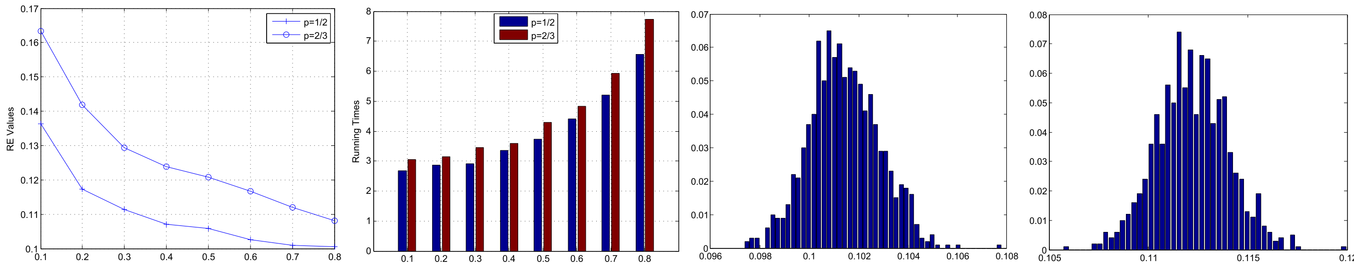
This section will quantitatively evaluate the performance on the synthetic data, especially for the DNNR method, which can be solved by the IRSVFc algorithm for the Schatten p-norm with $p = 1/2$ and $2/3$, respectively. We first generate the ground truth by the low-rank matrix $M_g = M_L M_R^T \in \mathbb{R}^{p \times q}$, where both $M_L \in R^{p \times r}$ and $M_R \in R^{q \times r}$ are obtained by `randn`, and the upper bound of the rank(M_g) is constrained by $r = p_r \times p$ and the number of missing entries in the corrupted matrix $M = M_g + n_l E \in \mathbb{R}^{p \times q}$ is computed by $p_m \times p^2$, where E is obtained by `randn` and n_l represents the noise level in this task.

We fix $p = q = 300$, and set $p_m, n_l \in [0.1 : 0.1 : 0.4]$ and $p_r \in [0.10 : 0.05 : 0.25]$. For $\{p_r, p_m, n_l\}$, we generate the synthetic data 20 times and report their averaging values as the final results. The matrix recovery performance is measured by computing $\text{RE} = \|\widehat{X} - M_g\|_F / \|M_g\|_F$, where $\widehat{X}$ is the recovered data matrix by different matrix completion algorithms. In particular, using the l_p-norm with $p = 1/2$ and $2/3$ on the singular values vector as the rank relaxations in IRNN, GPG, and DNNR methods. Here, Table I lists the RE values of all the methods, and Fig. 2 gives the demonstration for the parameter τ, running times, and initial value of our algorithm. Among them, we have the following observations.

1) These nonconvex methods can obtain lower RE values than the convex case. Using the Schatten p-norm with $p = 1/2$ and $2/3$ to the rank relaxations, we can achieve a better competitive performance than both IRNN and GPG though the improvements are slight. However, our methods can obtain relatively lower RE values than the compared methods.

2) The values of RE achieved by all of the methods become slightly worse with the increasing of p_r, p_m, and n_l, respectively. Specifically, when these values are small, the REs become relatively low. This phenomenon also reflects the fact that when the rank number becomes very large, the possibility of recovery becomes very difficult. Such similar conclusions have been verified in [14], [18], and [36].

3) It follows from the statements of Section III-D that τ influences both the efficacy and the efficiency for the proposed



(a) (b) (c) (d)

Fig. 2. Various choices of τ for (a) RE values and (b) running times (seconds); the distribution of RE values with 1000 different random initializations for (c) $p = 1/2$ and (d) $p = 2/3$ under the special case ($p_r = 0.2$, 0.5, 0.5, 300).

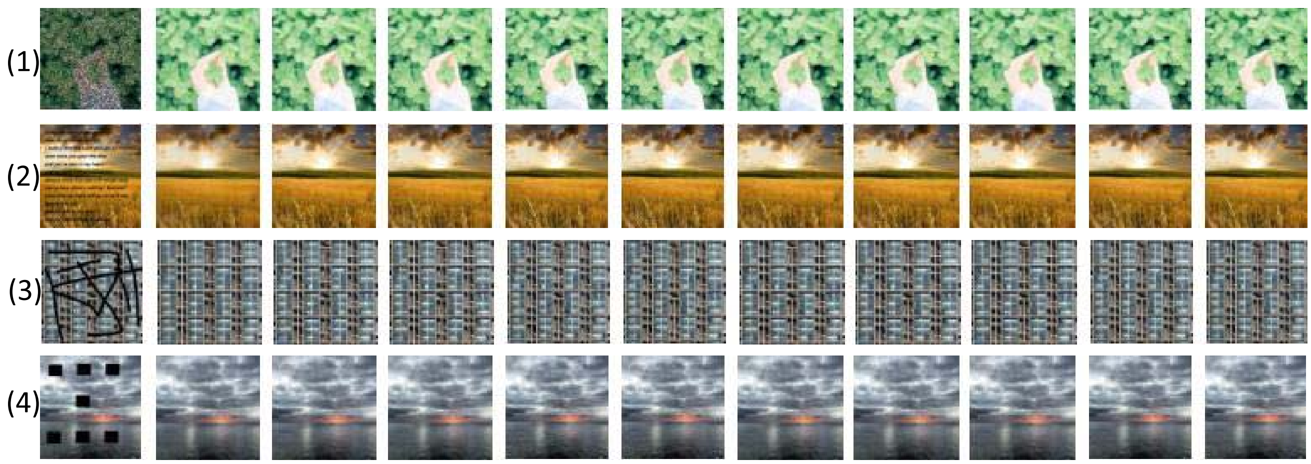


Fig. 3. Original and incomplete images (first column) are used for the recovered images (second to fifth column) generated by APGL, TNNR, SPNM, and WNNM, with other images by IRNN, GPG, and DNNR for l_p-norm with $p = 1/2$ (sixth to eighth column) and $p = 2/3$ (ninth to eleventh column) in sequence.

methods. Based on Fig. 2(a) and (b), we can conclude that the larger of the value of τ, the lower of the RE value, and the greater of the timing costs. Since the involved Schatten p-norm with $p = 1/2$ and $2/3$ is nonconvex, the converged solution may be different due to various choices of initializations. We conduct 1000 runs with random initialization for these cases, Fig. 2(c) and (d) shows that most of the solutions are concentrically distributed in the median regions.

B. Real Images Data

This section mainly applies several convex and nonconvex low-rank matrix completion methods for the recovery of real images. Similar to the experimental settings in [15], we test four incomplete types (i.e., random, text, curve, and block) on four natural images accordingly as shown in the first column of Fig. 3. All of the methods and algorithms are used to recover the missing entries of those partially damaged images. Consider that each of the color images has three channels (i.e., red, green, and blue), we need to recover the missing pixels by exploiting those matrix completion methods on each channel of the image independently and then combine them to get the final recovery results. Besides the RE values defined earlier, the recovered images are evaluated by the peak signal-to-noise ratio (PSNR) [15], [36] defined by

$$\mathrm{PSNR} = 10\log_{10}\left(\frac{255^2}{\frac{1}{3pq}\sum_{i=1}^{3}\|\widehat{X}_i - M_i\|_F^2}\right) \tag{24}$$

where both $\widehat{X}_i$ and M_i are the original image matrix and the recovered image matrix of the ith channel, and its size is $p \times q$. Higher PSNR values and lower RE values can indicate a better recovery performance of the involved methods than the compared methods. It follows from Table II that our methods can obtain both higher PSNR values and lower RE values than other methods. In addition, these nonconvex methods still outperform the convex nuclear norm-based methods (e.g., APGL), and DNNR outperforms GPG, which outperforms IRNN. This phenomenon is consistent with the conclusion [16], [18], [36] due to the tight relaxation. The recovered images are shown in Fig. 3 though the visual differences seem to be minor. In accordance with the derived sufficient decrease condition, the objective function values plotted in Fig. 4 are monotonically decreasing for IRNN, GPG, and DNNR when $p = 2/3$ and $1/2$, respectively. We observe that our methods decrease the objective function much faster since they need less iteration numbers, especially for $p = 2/3$.

C. Jester Joke Data

To further compare with these rank relaxation matrix completion methods, we perform the experiments on the Jester joke data set,[8] which contains 4.1 million ratings for 100 jokes from 73 421 users. The experimental data set includes four files: Jester-1, Jester-2, Jester-3, and Jester-all

[8]http://www.ieor.berkeley.edu/ goldberg/jester-data/

TABLE II

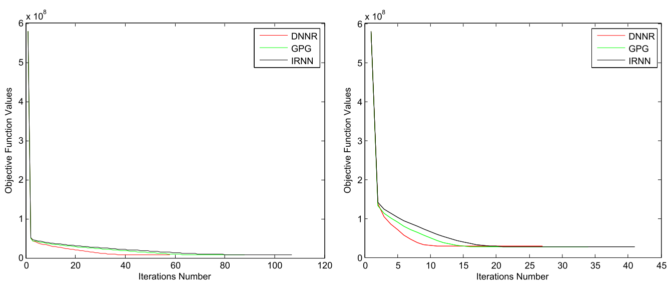
PSNR (DECIBEL) VALUES AND RE VALUES OF THE LOW-RANK MATRIX RECOVERY RESULTS BY COMPARING OUR METHODS WITH SEVERAL OTHER METHODS ON THE ABOVE IMAGES WITH MISSING ELEMENTS GENERATED BY RANDOM/TEXT/BLOCK/CURVE MASK

//	APGL	TNNR	SPNM	WNNM	l_p $(p = 1/2)$			l_p $(p = 2/3)$		
					IRNN	GPG	DNNR	IRNN	GPG	DNNR
(1)	30.233 (0.140)	31.284 (0.124)	31.618 (0.119)	31.911 (0.115)	30.994 (0.128)	31.835 (0.116)	32.458 (0.108)	31.558 (0.120)	32.528 (0.107)	**32.628** (**0.106**)
(2)	23.543 (0.374)	24.797 (0.328)	24.345 (0.343)	24.361 (0.345)	24.303 (0.348)	24.463 (0.340)	24.860 (0.325)	24.798 (0.327)	24.898 (0.324)	**24.975** (**0.321**)
(3)	19.789 (0.587)	20.257 (0.556)	20.525 (0.539)	20.067 (0.569)	20.424 (0.546)	20.482 (0.542)	20.698 (0.529)	20.644 (0.532)	20.738 (0.526)	**20.772** (**0.524**)
(4)	18.860 (0.639)	19.785 (0.575)	19.769 (0.577)	20.257 (0.544)	19.338 (0.605)	19.886 (0.568)	20.091 (0.574)	19.798 (0.563)	19.966 (0.555)	**20.389** (**0.536**)

TABLE III

NMAE VALUES AND RUNNING TIMES (SECONDS) OF ALL THE MATRIX RECOVERY METHODS ON THE JESTER JOKE DATA

Methods	(p, q)	APGL	TNNR	SPNM	WNNM	l_p $(p = 1/2)$			l_p $(p = 2/3)$		
						IRNN	GPG	DNNR	IRNN	GPG	DNNR
Jester-1	(24983,100)	1.688e-1 (32)	1.642e-1 (381)	1.632e-1 (267)	1.664e-1 (52)	1.678e-1 (409)	1.670e-1 (308)	1.601e-1 (222)	1.658e-1 (224)	1.645e-1 (172)	**1.582e-1** (132)
Jester-2	(23500,100)	1.692e-1 (34)	1.654e-1 (369)	1.634e-1 (251)	1.670e-1 (43)	1.683e-1 (394)	1.675e-1 (295)	1.616e-1 (214)	1.667e-1 (215)	1.656e-1 (155)	**1.599e-1** (127)
Jester-3	(24938,100)	1.923e-1 (98)	1.880e-1 (494)	1.881e-1 (270)	1.886e-1 (65)	1.916e-1 (468)	1.909e-1 (345)	1.893e-1 (305)	1.891e-1 (375)	1.795e-1 (271)	**1.754e-1** (265)
Jester-all	(73421,100)	1.827e-1 (136)	1.675e-1 (1413)	1.647e-1 (776)	1.676e-1 (96)	1.776e-1 (1555)	1.766e-1 (1079)	1.717e-1 (968)	1.690e-1 (1263)	1.685e-1 (809)	**1.631e-1** (638)



(a) (b)

Fig. 4. Convergence curves of the objective functions for three related algorithms to recover the natural image with number of (1). (a) $p = 1/2$. (b) $p = 2/3$.

(combining Jester-1,2,3 together). For each data set, we have an incomplete data matrix, and randomly choose half of entries in Γ to construct the entries set Ω, where Γ is the known ratings set of the entries by users. Different from the above measured criteria, the recovery performance of these methods are evaluated by the normalized mean absolute error (NMAE) [21]. More generally, a small value of NMAE usually indicates a good recovery performance. We compute the value of NMAE by

$$\text{NMAE} = \frac{\Sigma_{(i,j)\in\Gamma\setminus\Omega}|X_{ij} - M_{ij}|}{(M_{\max} - M_{\min})|\Gamma \setminus \Omega|} \tag{25}$$

where $M_{\max} = \max_{ij} M_{ij}$ and $M_{\min} = \min_{ij} M_{ij}$, respectively. The values of NMAE are demonstrated in Table III, which can verify the achievable superiority of the proposed methods. We observe that the values of NMAE on the Jester-3 data set are slightly higher than other Jester data sets for all the used methods. Since Jester-all is the combination of Jester-1,2,3, the NMAE values will fall into the median values. Due to the relatively tighter approximation of the rank function than the convex relaxations, the proposed DNNR methods can achieve the lower NMAE values than other compared methods. The timing costs are listed in Table III. It is easy to observe that the IRNN, GPG, and DNNR methods share a much higher timing cost than both APGL and WNNM. In addition, both SPNM and TNNR methods consume much timing costs due to the computations of the inverse matrix and the two-stages of optimization, respectively. Although our methods can achieve better recovery efficacy than the compared methods, they also suffer from higher computational loads than the matrix factorization methods. This motivates us to develop faster and accelerated proximal operators based on optimization algorithms like [21]–[25], which can improve the computational efficiency by reducing the total number of iterations.

V. CONCLUSION

This paper presents a general WNNR relaxation function to substitute the rank function. The involved strategy actually derives from the DNNR relaxation function by computing the supergradient of the nonconvex function. The main merits of WNNR relaxation function are its relatively flexible and nearly exact relaxation over some convex and nonconvex relaxations of rank function. The IRSVFc algorithm is further devised to solve our nonconvex optimization problem efficiently. Most importantly, we further give its local and global convergence analysis by combining the milder assumptions and the KŁ inequality. Experimental results on both synthetic and real-world data can show their superior performance over several existing matrix completion methods. Furthermore, developing faster and accelerated optimization algorithms of the proposed DNNR methods will be one of the future research studies for solving large-scale matrix completion problems.

ACKNOWLEDGMENT

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

REFERENCES

[1] N. Srebro and T. Jaakkola, "Weighted low-rank approximations," in *Proc. Int. Conf. Mach. Learn. (ICML)*, 2003, pp. 720–727.

[2] 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.

[3] R. Cabral, F. Da La Torre, J. P. Costeira, and A. Bernardino, "Unifying nuclear norm and bilinear factorization approaches for low-rank matrix decomposition," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Dec. 2013, pp. 2488–2495.

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

[5] T. Hastie, R. Mazumder, J. D. Lee, and R. Zadeh, "Matrix completion and low-rank SVD via fast alternating least squares," *J. Mach. Learn. Res.*, vol. 16, no. 1, pp. 3367–3402, 2015.

[6] C. Dorffer, M. Puigt, G. Delmaire, and G. Roussel, "Fast nonnegative matrix factorization and completion using Nesterov iterations," in *Proc. Int. Conf. Latent Variable Anal. (ICLVA)*, 2017, pp. 26–35.

[7] Y. Liu, L. Jiao, and F. Shang, "An efficient matrix factorization based low-rank representation for subspace clustering," *Pattern Recognit.*, vol. 46, no. 1, pp. 284–292, Jan. 2013.

[8] 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.

[9] 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.

[10] F. Zhang, J. Yang, J. Qian, and Y. Xu, "Nuclear norm-based 2-DPCA for extracting features from images," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 26, no. 10, pp. 2247–2260, Oct. 2015.

[11] S. Li, K. Li, and Y. Fu, "Self-taught low-rank coding for visual learning," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 29, no. 3, pp. 645–656, Mar. 2018.

[12] F. Nie, H. Huang, and C. Ding, "Low-rank matrix recovery via efficient schatten p-norm minimization," in *Proc. 26th AAAI Conf. Artif. Intell. (AAAI)*, 2012, pp. 655–661.

[13] 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.

[14] 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.

[15] 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.

[16] C. Lu, C. Zhu, C. Xu, S. Yan, and Z. Lin, "Generalized singular value thresholding," in *Proc. 29th AAAI Conf. Artif. Intell. (AAAI)*, 2015, pp. 1805–1811.

[17] 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.

[18] 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.

[19] E. J. Candès and B. Recht, "Exact matrix completion via convex optimization," *Found. Comput. Math.*, vol. 9, no. 6, pp. 717–772, 2009.

[20] E. J. Candès and T. Tao, "The power of convex relaxation: Near-optimal matrix completion," *IEEE Trans. Inf. Theory*, vol. 56, no. 5, pp. 2053–2080, May 2010.

[21] K. Toh and S. Yun, "An accelerated proximal gradient algorithm for nuclear norm regularized least squares problems," *Pacific J. Optim.*, vol. 6, pp. 615–640, 2010.

[22] X. Lan, A. J. Ma, and P. C. Yuen, "Multi-cue visual tracking using robust feature-level fusion based on joint sparse representation," in *Proc. IEEE Int. Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2014, pp. 1194–1201.

[23] X. Lan, M. Ye, S. Zhang, and P. C. Yuen, "Robust collaborative discriminative learning for RGB-infrared tracking," in *Proc. 32nd AAAI Conf. Artif. Intell. (AAAI)*, 2018, pp. 7008–7015.

[24] 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.

[25] 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.

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

[27] 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.

[28] S. Wang, D. Liu, and Z. Zhang, "Nonconvex relaxation approaches to robust matrix recovery," in *Proc. Int. Joint Conf. Artif. Intell. (IJCAI)*, 2013, pp. 1764–1770.

[29] T. H. Oh, Y. 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.

[30] 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.

[31] 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.

[32] 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. doi: 10.1109/TCYB.2018.2811764.

[33] L. E. Frank and J. H. Friedman, "A statistical view of some chemometrics regression tools," *Technometrics*, vol. 35, no. 2, pp. 109–135, 1993.

[34] J. H. Friedman, "Fast sparse regression and classification," *Int. J. Forecasting*, vol. 28, no. 3, pp. 722–738, Jul./Sep. 2012.

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

[36] C. Lu, J. Tang, S. Yan, and Z. Lin, "Generalized nonconvex nonsmooth low-rank minimization," in *Proc. IEEE Int. Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2014, pp. 4130–4137.

[37] 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.

[38] H. Zhang, J. Qian, J. Gao, J. Yang, and C. Xu, "Scalable proximal Jacobian iteration method with global convergence analysis for nonconvex unconstrained composite optimizations," *IEEE Trans. Neural Netw. Learn. Syst.*, to be published. doi: 10.1109/TNNLS.2018.2885699.

[39] 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, 2013.

[40] H. Attouch, J. Bolte, P. Redont, and A. Soubeyran, "Proximal alternating minimization and projection methods for nonconvex problems: An approach based on the Kurdyka-Łojasiewicz inequality," *Math. Oper. Res.*, vol. 35, no. 2, pp. 438–457, 2010.

[41] 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.

[42] W. Cao, J. Sun, and Z. Xu, "Fast image deconvolution using closed-form thresholding formulas of L_q $(q = \frac{2}{3}, \frac{1}{2})$ regularization," *J. Vis. Commun. Image. Represent.*, vol. 24, no. 1, pp. 31–41, Jan. 2013.

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

[44] 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.

[45] C. J. Hsieh and P. A. Olsen, “Nuclear norm minimization via active subspace selection,” in *Proc. Int. Conf. Mach. Learn. (ICML)*, 2014, pp. 575–583.

[46] B. Bao, G. Liu, C. Xu, and S. Yan, “Inductive robust principal component analysis,” *IEEE Trans. Image Process.*, vol. 21, no. 8, pp. 3794–3800, Aug. 2012.

[47] Z. Li, F. Nie, X. Chang, L. Nie, H. Zhang, and Y. Yang, “Rank-constrained spectral clustering with flexible embedding,” *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 29, no. 12, pp. 6073–6082, Dec. 2018.

[48] 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. 208–214, Jan. 2018.

[49] Z. Lu and X. Li, “Sparse recovery via partial regularization: Models, theory, and algorithms,” *Math. Oper. Res.*, vol. 43, no. 4, pp. 1290–1316, 2018.

[50] K. Border. (2001). *The Supergradient of a Concave Function*. [Online]. Available: http://www.hss.caltech.edu/kcb/Notes/Supergrad.pdf

[51] R. T. Rockafellar and R. J. B. Wets, *Variational Analysis*. Berlin, Germany: Springer, 2004.

[52] H. Miranda and R. C. Thompson, “A trace inequality with a subtracted term,” *Linear Algebra Appl.*, vol. 185, no. 93, pp. 165–172, May 1993.

[53] W. Zuo, D. Meng, L. Zhang, X. Feng, and D. Zhang, “A generalized iterated shrinkage algorithm for non-convex sparse coding,” in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Dec. 2013, pp. 217–224.

[54] M. Zhang, Z. Huang, and Y. Zhang, “Restricted p-isometry properties of nonconvex matrix recovery,” *IEEE Trans. Inf. Theory*, vol. 59, no. 7, pp. 4316–4323, Jul. 2013.

[55] M. Mohammadi, M. Zadeh, and M. Skoglund, “Performance guarantees for Schatten-p quasi-norm minimization in recovery of low-rank matrices,” *Signal Process.*, vol. 114, pp. 225–230, Sep. 2015.

[56] L. Luo, J. Yang, J. Qian, Y. Tai, and G. Lu, “Robust image regression based on the extended matrix variate power exponential distribution of dependent noise,” *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 28, no. 9, pp. 2168–2182, Sep. 2017.

[57] 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.

[58] K. Lange, D. R. Hunter, and I. Yang, “Optimization transfer using surrogate objective functions,” *J. Comput. Graph. Statist.*, vol. 9, no. 1, pp. 1–20, Mar. 2000.

[59] D. R. Hunter and R. Li, “Variable selection using MM algorithms,” *Ann. Statist.*, vol. 33, no. 4, pp. 1617–1942, 2005.

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

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, 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, Nanjing, China.

His current research interests include computer vision and machine learning.

Chen Gong (M’16) received the B.E. degree from the East China University of Science and Technology, Shanghai, China, in 2010, and the dual Ph.D. degree from Shanghai Jiao Tong University, Shanghai, and the University of Technology Sydney, Ultimo, NSW, Australia, 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, Nanjing, China. His current research interests include machine learning, data mining, and learning-based vision problems.

Jianjun Qian received the Ph.D. degree in pattern recognition and intelligence systems from the School of Computer Science and Engineering, 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, especially for face recognition.

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 Post-Doctoral Researcher with the Department of Electrical and Computer Engineering, Carnegie Mellon University, Pittsburgh, PA, USA. He is currently an Assistant Professor with the Department of Computer and Information Science, University of Macau, Zhuhai, China. His current research interests include medical biometrics, pattern recognition, and image processing.

Chunyan Xu received the Ph.D. degree from the School of Computer Science and Technology, Huazhong University of Science and Technology, Wuhan, China, in 2015.

From 2013 to 2015, she was a Visiting Scholar with the Department of Electrical and Computer Engineering, National University of Singapore, Singapore. She is currently a Lecturer with the School of Computer Science and Engineering, Nanjing University of Science and Technology, Nanjing, China. Her current research interests include computer vision, manifold learning, and deep learning.

Jian Yang (M’08) received the Ph.D. degree in pattern recognition and intelligence systems from the Nanjing University of Science and Technology (NUST), Nanjing, China, in 2002.

In 2003, he was a Post-Doctoral Researcher with the University of Zaragoza, Zaragoza, Spain. From 2004 to 2006, he was a Post-Doctoral Fellow with the Biometrics Centre, The Hong Kong Polytechnic University, Hong Kong. From 2006 to 2007, he was a Post-Doctoral Fellow with the Department of Computer Science, New Jersey Institute of Technology, Newark, NJ, USA. He is currently a Chang-Jiang Professor with the School of Computer Science and Technology, NUST. He has authored more than 200 scientific papers in pattern recognition and computer vision. His papers have been cited more than 6000 times in the Web of Science and 15 000 times in the Scholar Google. His current research interests include pattern recognition, computer vision, and machine learning.

Dr. Yang is a Fellow of IAPR. He is currently an Associate Editor of *Pattern Recognition*, *Pattern Recognition Letters*, the IEEE TRANSACTIONS ON NEURAL NETWORKS AND LEARNING SYSTEMS, and *Neurocomputing*.