Generalized Nonconvex Nonsmooth Low-Rank Matrix Recovery Framework With Feasible Algorithm Designs and Convergence Analysis

Hengmin Zhang, Feng Qian, *Senior Member, IEEE*, Peng Shi, *Fellow, IEEE*, Wenli Du, *Member, IEEE*, Yang Tang, *Senior Member, IEEE*, Jianjun Qian, *Member, IEEE*, Chen Gong, *Member, IEEE*, and Jian Yang, *Member, IEEE*

***Abstract*— Decomposing data matrix into low-rank plus additive matrices is a commonly used strategy in pattern recognition and machine learning. This article mainly studies the alternating direction method of multiplier (ADMM) with two dual variables, which is used to optimize the generalized nonconvex nonsmooth low-rank matrix recovery problems. Furthermore, the minimization framework with a feasible optimization procedure is designed along with the theoretical analysis, where the variable sequences generated by the proposed ADMM can be proved to be bounded. Most importantly, it can be concluded from the Bolzano–Weierstrass theorem that there must exist a subsequence converging to a critical point, which satisfies the Karush–Kuhn–Tucher (KKT) conditions. Meanwhile, we further ensure the local and global convergence properties of the generated sequence relying on constructing the potential objective function. Particularly, the detailed convergence analysis would be regarded as one of the core contributions besides the algorithm designs and the model generality. Finally, the numerical simulations and the real-world applications are both provided to verify the consistence of the theoretical results, and we also validate the superiority in performance over several mostly related solvers to the tasks of image inpainting and subspace clustering.**

***Index Terms*— Algorithm designs, convergence analysis, low-rank matrix recovery, multiple variables, nonconvex alternating direction method of multiplier (ADMM).**

Manuscript received October 7, 2021; revised March 22, 2022; accepted June 11, 2022. This work was supported in part by the National Natural Science Fund for Distinguished Young Scholars under Grant 61725301; in part by the National Key Research and Development Program of China under Grant 2021YFB1714300; in part by the General and Youth Programs under Grant 61906067, Grant 62176124, Grant 61973162, and Grant 61876083; in part by the China Postdoctoral Science Foundation under Grant 2019M651415 and Grant 2020T130191; and in part by the UM Macao Talent Programme under Grant UMMTP-2020-01. *(Corresponding authors: Feng Qian; Wenli Du.)*

Hengmin Zhang was with the Key Laboratory of Advanced Smart Manufacturing in Energy Chemical Process, Ministry of Education, School of Information Science and Engineering, East China University of Science and Technology, Shanghai 200237, China. He is now with the Department of Computer and Information Science, University of Macau, Macau (e-mail: zhanghengmin@126.com).

Feng Qian, Wenli Du, and Yang Tang are with the Key Laboratory of Advanced Smart Manufacturing in Energy Chemical Process, Ministry of Education, School of Information Science and Engineering, East China University of Science and Technology, Shanghai 200237, China, and also with the Shanghai Institute of Intelligent Science and Technology, Tongji University, Shanghai 200092, China (e-mail: fqian@ecust.edu.cn; wldu@ecust.edu.cn; yangtang@ecust.edu.cn).

Peng Shi is with the School of Electrical and Electronic Engineering, The University of Adelaide, Adelaide, SA 5005, Australia, and also with the College of Engineering and Science, Victoria University, Melbourne, VIC 8001, Australia (e-mail: peng.shi@adelaide.edu.au).

Jianjun Qian, Chen Gong, and Jian Yang are with the PCA Laboratory and 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: csjqian@njust.edu.cn; chen.gong@njust.edu.cn; csjyang@njust.edu.cn).

This article has supplementary material provided by the authors and color versions of one or more figures available at https://doi.org/10.1109/TNNLS.2022.3183970.

Digital Object Identifier 10.1109/TNNLS.2022.3183970

I. Introduction

IT IS well-known that low-rank matrix recovery problems have a mass of real-world applications (e.g., recommendation system [1], social network [2], background substraction [3], [4], subspace clustering [5], [6], and image classification/inpainting [7], [8]) in pattern recognition and machine learning. The popular methods consist of matrix completion (MC) [9], robust principle component analysis (RPCA) [10], low-rank representation (LRR) [11], robust MC (RMC) [12], and their series of variants [13], [14]. In these methods, the involved rank function (i.e., the number of nonzero singular values of a matrix) is nonconvex and discrete [15], which makes the minimization problem be NP-hard and not be optimized easily. To solve these issues, some efforts have been made by using convex and nonconvex substitutes along with devising first-order optimization algorithms. Subsequently, we provide some statements from two aspects.

1) *To relax the rank function*, nuclear norm (i.e., the sum of all singular values of a matrix) and nonconvex rank substitutes (e.g., Schatten-p norm [2] and truncated/weighted nuclear norm [25], [26]) were studied for the low-rank matrix recovery problems. As we know, some nonconvex counterparts of L_0-norm[1] (e.g., L_p-norm [28], [29], minimax concave penalty (MCP) [30], and smoothly clipped absolute deviation (SCAD) [31]) can be extended to substitute the rank function by acting on the singular value elements, respectively. The derivations rely on the intimate relationship between

[1]These nonconvex functions [18], [27] satisfy the Kurdyka–Łojasiewicz (KŁ) inequality, which plays the key role for the global convergence property of nonconvex first-order optimization algorithms.

TABLE I

SEVERAL RELATED OPTIMIZATION ALGORITHMS FOR LOW-RANK MATRIX PROBLEM FORMULATIONS WITH SINGLE/MULTIPLE VARIABLES, WITH/WITHOUT CONSTRAINTS, LOCAL/GLOBAL CONVERGENCE GUARANTEES AND NONLINEARIZATION

Algorithms	Problem Formulations	Variables	Constraints	Convergence	Non-linearization
GSVT [16]	$\min_{\mathbf{Z}} f_\lambda(\mathbf{Z}) + \|P_\Omega(\mathbf{D}-\mathbf{Z})\|_L$	single	$\times$	local	$\times$
NLRR [17]	$\min_{\mathbf{Z}} f_\lambda(\mathbf{Z}) + \|\mathbf{D}-\mathbf{AZ}\|_L$	single	$\times$	local	$\times$
PIRNN [18]	$\min_{\mathbf{Z}} f_\lambda(\mathbf{Z}) + g(\mathbf{D}-\mathcal{A}(\mathbf{Z}))$	single	$\times$	local & global	$\times$
ADMc [19]	$\min_{\mathbf{X},\mathbf{Z}} f_\lambda(\mathbf{X}) + g(\mathbf{D}-\mathcal{A}(\mathbf{Z}))$	multiple	$\mathbf{Z}=\mathbf{X}$	local	✓
ADML [20]	$\min_{\mathbf{Z},\mathbf{E}} f_\lambda(\mathbf{Z}) + g(\mathbf{E})$	multiple	$\mathcal{A}(\mathbf{Z})+\mathbf{E}=\mathbf{D}$	local & global	✓
ADM^{2}B [21]	$\min_{\mathbf{x}_i} \sum_{i=1}^{m} f_i(\mathbf{x}_i)$	multiple	$\sum_{i=1}^{m-1} \mathbf{A}_i\mathbf{x}_i + \mathbf{x}_m = \mathbf{D}$	local & global	✓
ADM^{2}F [22]	$\min_{\mathbf{x}_i} \sum_{i=1}^{K} f_i(\mathbf{x}_i) + \iota(\mathbf{x}_1,...,\mathbf{x}_K)$	multiple	$\sum_{i=1}^{K} \mathbf{A}_i\mathbf{x}_i = \mathbf{D}$	local	✓
PALM [23] SPJIM [24]	$\min_{\mathbf{Z},\mathbf{E}} f_\lambda(\mathbf{Z}) + g(\mathbf{E}) + \frac{\gamma}{2}\|\mathcal{A}(\mathbf{Z})+\mathcal{B}(\mathbf{E})-\mathbf{D}\|_F^2$	multiple	$\times$	local & global	$\times$
JS$_p$L$_p$NM [2]	$\min_{\mathbf{Z},\mathbf{E},\mathbf{X}} \lambda\|\mathbf{Z}\|_{S_p}^p + \|P_\Omega(\mathbf{E})\|_L$	multiple	$\mathbf{Z}=\mathbf{X}$, $P_\Omega(\mathbf{E})=P_\Omega(\mathbf{D}-\mathbf{Z})$	local	✓
ADM^{2}GS [4]	$\min_{\mathbf{Z},\mathbf{E},\mathbf{X}} f_\lambda(\mathbf{Z}) + g(\mathbf{E}) + \frac{\gamma}{2}\|\mathcal{C}(\mathbf{X})-\mathbf{D}\|_F^2$	multiple	$\mathcal{A}(\mathbf{Z})+\mathcal{B}(\mathbf{E})=\mathbf{X}$	local & global	✓
this work	$\min_{\mathbf{X},\mathbf{Y},\mathbf{Z},\mathbf{E}} f_\lambda(\mathbf{X}) + g(\mathbf{Y}) + \frac{\gamma}{2}\|\mathcal{A}(\mathbf{Z})+\mathcal{B}(\mathbf{E})-\mathbf{D}\|_F^2$	multiple	$\mathbf{Z}=\mathbf{X}$, $\mathbf{E}=\mathbf{Y}$	local & global	✓

L_0-norm and rank function, as well as between L_1-norm and nuclear norm [24]. The empirical results have shown that nonconvex relaxations can obtain better performance than convex cases, where the former can avoid the imbalanced penalization of different singular values, whereas the latter cannot.

2) *To design the optimization algorithms*, two popular first-order solutions, i.e., alternating direction method of multiplier (ADMM) [32]–[35] and accelerated proximal gradient (APG) algorithm [36]–[38], have been employed to address the constrained and unconstrained minimization problems. The constrained problems can be transformed into unconstrained ones, such as joint Schatten-p norm and L_p-norm matrix (JS$_p$L$_p$NM) [2], and the alternating direction method with continuation (ADMc) [19]. Several nonconvex algorithms listed in Table I (e.g., nonconvex ADMM with a general step-size (ADM^{2}GS) [4], proximal iteratively reweighted nuclear norm (PIRNN) [18], general singular value thresholding (GSVT) [16], nonconvex LRR (NLRR) [17], proximal alternating linearized minimization (PALM) [23], and scalable proximal Jacobian iteration method (SPJIM) [24]) were devised to achieve a low-rank solver, respectively. In addition, the alternating direction method with linear (ADML) constraints [20] and with multiblock variables (ADM^{2}B) [21] were also considered for nonconvex and separable problems. The convergence property of ADMM was analyzed for solving a family (ADM^{2}F) of certain nonconvex consensus and sharing problems in [22]. To make the comparisons, we conclude from [2], [16], [17], [19], [22], and [39] that they can be guaranteed to have the local properties, i.e., the proof process usually includes that the values of the objective function are monotonically decreasing over the iterations and then conclude that there exists a subsequence converging to a stationary point if the sequence is bounded. Relying on the popular KŁ inequality, the convergence analysis in [4], [18], [20], [21], [23], and [24] mainly include the global convergence properties, i.e., the generated sequences are Cauchy ones. However, none of the above considers the local and global convergence guarantees of nonconvex ADMM with two dual variables.

The convergence analysis should be provided for the further statements and comparisons; some first-order algorithms of Table I are briefly discussed as follows.

1) The proof routines of this work and some existing works (e.g., [4], [18], and [23]) mainly provide the global convergence guarantees. In the local convergence analysis of several first-order optimization algorithms (e.g., [16], [17], and [19]), they focus on guaranteeing the limiting point of variable subsequence satisfying the Karush–Kuhn–Tucher (KKT) conditions.
2) To optimize nonsmooth, convex/nonconvex, and multiple variable situations in the objective function, we can observe that it needs to extend some existing results in [2], [4], [19], [23], and [40] through the proper arrangements.
3) To achieve the closed-form solvers of all the subproblems, various strategies could be used, such as introducing some auxiliary variables, such as [2], [4], and [19], or exploiting the linearized techniques, such as [16]–[18], [23], and [24].

It follows from the survey works [14], [41] that the traditional MC, RPCA, LRR, and RMC models involve low-rank matrix plus additive one, in which the additive measurements may represent convex/nonconvex relaxations of L_0-norm, $L_{2,0}$-norm, and rank function. They mainly measure the residual matrix or the coefficient one, respectively. Then, we study the rank minimization problems

$$\min_{\mathbf{Z}} \lambda \text{rank}(\mathbf{Z}) + \|P_\Omega(\mathbf{D}) - P_\Omega(\mathbf{Z})\|_p^q \tag{1}$$

$$\min_{\mathbf{Z}} \lambda \text{rank}(\mathbf{Z}) + \|\mathbf{D}-\mathbf{AZ}\|_\ell \tag{2}$$

TABLE II
SEVERAL NONCONVEX RELAXATION FORMULAS OF L_0-NORM AND THEIR PROXIMAL OPERATORS

Names	Function formulas of $h_\lambda(t)$	Proximal operators of $s^* = \operatorname{argmin}_s h_\lambda(s) + \frac{1}{2}\|s-t\|_2^2$
L_p-norm [2]	$\lambda\|t\|^p$.	$\begin{cases} 0, & \|t\| \le \nu_1, \\ \operatorname{argmin}_{t\in\{0,t_1\}} h(t), & t > \nu_1, \\ \operatorname{argmin}_{t\in\{0,t_2\}} h(t), & \text{otherwise.} \end{cases}$
MCP [30]	$\begin{cases} \lambda\|t\| - \frac{t^2}{2p}, & \|t\| \le p\lambda, \\ \frac{1}{2}p\lambda^2, & \text{otherwise.} \end{cases}$	$\begin{cases} 0, & \|t\| \le \lambda, \\ t, & \|t\| > p\lambda, \\ \operatorname{sign}(t)\frac{p(\|t\|-\lambda)}{p-1}, & \text{otherwise.} \end{cases}$
SCAD [31]	$\begin{cases} \lambda\|t\|, & \|t\| \le \lambda, \\ \frac{1}{2}\lambda^2(p+1), & \|t\| > p\lambda, \\ \frac{\lambda p\|t\| - \frac{1}{2}(t^2+\lambda^2)}{p-1}, & \text{otherwise.} \end{cases}$	$\begin{cases} \operatorname{sign}(t)\max(\|t\|-\lambda, 0), & \|t\| \le 2\lambda, \\ \frac{(p-1)t - \operatorname{sign}(t)\lambda p}{p-2}, & 2\lambda < \|t\| \le p\lambda, \\ t, & \text{otherwise.} \end{cases}$

where the parameter $\lambda > 0$, $\mathbf{D} \in \mathbb{R}^{m\times n}$ is a data matrix and various choices of $\|\cdot\|_\ell$ and $\|\cdot\|_p^q$ in both problems (1) and (2) show their generality for the residual descriptions. In detail, when $0 < q \le 1$, (1) becomes the RMC problem [2], [12]. Here, the identity matrix is denoted as $\mathbf{I}$. When $q = 2$, it is the MC problem [42]. When $\Omega = \mathbf{I}$ and $q = 1$, it is the RPCA problem [43]. Besides this, (2) is the LRR problem [11], [13], [44]. When $\mathbf{A} = \mathbf{I}$, it becomes the RPCA problem [10], [43].

To solve problems (1) and (2), the existing works usually choose the substitutes of rank function and loss function and devise the efficient algorithms [14]. In this work, we mainly consider a class of general problem formulation, denoted as

$$\min_{\mathbf{Z}} \ f_\lambda(\mathbf{Z}) + g(\mathbf{D} - \mathcal{A}(\mathbf{Z})) \tag{3}$$

where $f_\lambda(\mathbf{Z})$ is the substitute of rank($\mathbf{Z}$), $g(\mathbf{D} - \mathcal{A}(\mathbf{Z}))$ measures the loss function, and $\mathcal{A}(\bullet)$ can generalize $P_\Omega(\mathbf{Z})$ and $\mathbf{AZ}$ for problems (1) and (2), respectively. Moreover, problem (3) can be regarded as the unified minimization framework of some existing low-rank matrix recovery models (e.g., joint Schatten-p norm and L_p-norm case [2], and smoothed LRR [45]) when the objective function involves various substitutes. In most cases, problem (3) can be optimized by iteratively reweighted algorithms, such as [3], [18], and [46]. In addition, one may transform it into the equivalent problems step in step as follows:

$$\min_{\mathbf{Z},\mathbf{E}} \ f_\lambda(\mathbf{Z}) + g(\mathbf{E}), \quad \text{s.t. } \mathbf{E} = \mathbf{D} - \mathcal{A}(\mathbf{Z}) \tag{4}$$

$$\min_{\mathbf{Z},\mathbf{E}} \ f_\lambda(\mathbf{Z}) + g(\mathbf{E}) + \frac{\gamma}{2}\|\mathcal{A}(\mathbf{Z}) + \mathbf{E} - \mathbf{D}\|_F^2 \tag{5}$$

where $\gamma > 0$ is the penalty parameter, problem (4) is the constrained version of problem (3), and its penalty version is problem (5). Here, when $\gamma \to +\infty$, problems (3)–(5) have equivalent relations, and when the subproblems do not guarantee achieving the closed-form solvers, we need to introduce the auxiliary variables, such as [4] and [47].

It can be concluded from the existing works that most of the nonconvex optimization algorithms can be designed to achieve a low-rank solver in general but lack the convergence analysis, especially for global situations. This is due to the fact that providing the global convergence guarantees is very challenging for nonconvex, nonsmooth, and even multiblock variables ADMM, as explained in [45] and [48]. However, it will be very critical not only for theoretical support but also for practical applications.

A. Notations

In this work, we employ capital letters to denote a matrix, e.g., $\mathbf{X} \in \mathbb{R}^{m\times n}$, and the lowercase letters to denote a vector, e.g., $\mathbf{x} \in \mathbb{R}^n$, where $\mathbb{R}^{m\times n}$ and $\mathbb{R}^n$ are the sets of all $m \times n$ matrices and all n-dimensional vectors, respectively. Let SVD be the simplified symbol of singular value decomposition, and σ_i is the ith element of a singular value vector. Let $\mathbf{D}_{i,j}$ be the (i, j)th entry and $\|\cdot\|_L$ be the convex norms for L_1, L_2, and $L_{2,1}$; $P_\Omega(\mathbf{D}_{i,j})$ equals to $\mathbf{D}_{i,j}$ if $(i, j) \in \Omega$; otherwise, $P_\Omega(\mathbf{D}_{i,j}) = 0$. $\mathcal{A}(\cdot)$ is a linear mapping, and $\mathcal{A}^*(\cdot)$ is its adjoint. Similar definitions also hold for $\mathcal{B}(\cdot)$ and $\mathcal{C}(\cdot)$ in this work. The Frobenius norm, nuclear norm, and Schatten-p norm are computed by $\|\mathbf{X}\|_F^2 = \sum_{i,j}\mathbf{X}_{i,j}^2 = \sum_i \sigma_i^2$, $\|\mathbf{X}\|_* = \sum_i \sigma_i$, and $\|\mathbf{X}\|_{S_p}^p = \sum_i \sigma_i^p$ ($0 < p < 1$), respectively. The "×" and "✓" signify "without" and "with" in two columns of constraints and nonlinearization, respectively.

B. Contributions

Based on the above statements, the main contributions are summarized from the following three aspects.

1) We extend nonconvex low-rank matrix recovery problem (4) to a more general and complex model compared with the studied mathematical model and other problem formulations listed in Table I, denoted as

$$\min_{\mathbf{Z},\mathbf{E}} \ f_\lambda(\mathbf{Z}) + g(\mathbf{E}), \quad \text{s.t. } \mathcal{A}(\mathbf{Z}) + \mathcal{B}(\mathbf{E}) = \mathbf{D} \tag{6}$$

where a series of equivalent transforms need to be done step by step through the penalty technique or introducing the auxiliary variables in the objective function. Hence, the whole iteration programming without inner loops can be made easily to optimize the problem (6).

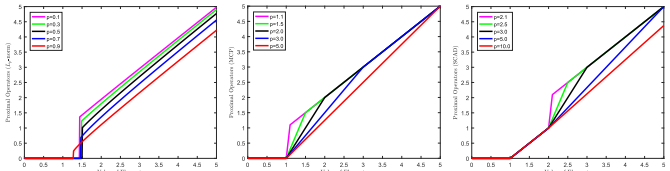


Fig. 1. Plotted curves of nonconvex proximal operators for (a) L_p-norm, (b) MCP, and (c) SCAD with different p values, respectively.

2) We develop nonconvex ADMM with two dual variables, which can guarantee the closed-form solvers in updating the primal variables. By analyzing the exact and inexact iteration procedures of multiblock nonconvex ADMM, the local and global convergence properties are further proved by constructing the potential objective function.
3) The numerical and theoretical results can show the nonincreasing property of a given potential objective function. Extensive experiments on the tasks of image inpainting and subspace clustering further verify that the developed methodology is capable of discovering superior performance over several mostly related works.

Outline: Section II presents the generalized model formulation of some nonconvex low-rank matrix recovery problems and its optimization algorithm. Section III provides the main results for the local and global convergence analyses. Section IV conducts the numerical experiments to validate the convergence properties and the evaluation performance by applying two modified models to the synthesis and real-world data. Finally, we conclude this work in Section V.

II. Problem Formulation

To get the closed-form solvers, it follows from (4) and (5) that problem (6) can be converted into a generalized nonconvex nonsmooth low-rank matrix recovery problem

$$\begin{aligned}&\min_{\mathbf{X},\mathbf{Y},\mathbf{Z},\mathbf{E}} f_\lambda(\mathbf{X})+g(\mathbf{Y})+\frac{\gamma}{2}\|\mathcal{A}(\mathbf{Z})+\mathcal{B}(\mathbf{E})-\mathbf{D}\|_F^2\\&\text{s.t. } \mathbf{Z}=\mathbf{X},\ \mathbf{E}=\mathbf{Y}\end{aligned}\tag{7}$$

where both $\mathbf{X}$ and $\mathbf{Y}$ are the introduced auxiliary variables. This can avoid the usage of linearization techniques, such as [24] and [32]. Several low-rank matrix recovery problems can be thought as the concrete examples of problem (7). The optimization algorithms of problem (7) and [22] and [49] with the constraints are different from [24], [50], and [51] due to the model formulations without the constrained equalities. Here, the optimization algorithms with local and global convergence guarantees are very important before giving the main results. The basic assumptions for the objective function in minimization problem (7) are given as follows.

A1: Both $f_\lambda(\mathbf{X})$ and $g(\mathbf{Y})$ are proper, lower semicontinuous on their domain of definition, respectively.

A2: The objective function is coercive, i.e., $f_\lambda(\mathbf{X})+g(\mathbf{Y})$ is bounded from below, and $\lim_{\|\cdot\|\to+\infty} f_\lambda(\mathbf{X})+g(\mathbf{Y})=+\infty$.

Before providing the convergence analysis, we need to write the augmented Lagrange function of problem (7) at first, where the representation formula is given as

$$\begin{aligned}&\mathcal{L}_\mu(\mathbf{X},\mathbf{Y},\mathbf{Z},\mathbf{E},\Lambda_1,\Lambda_2)\\&\quad=f_\lambda(\mathbf{X})+g(\mathbf{Y})\\&\qquad+\frac{\gamma}{2}\|\mathcal{A}(\mathbf{Z})+\mathcal{B}(\mathbf{E})-\mathbf{D}\|_F^2-\frac{1}{2\mu}\left(\|\Lambda_1\|_F^2+\|\Lambda_2\|_F^2\right)\\&\qquad+\frac{\mu}{2}\left(\left\|\mathbf{Z}-\mathbf{X}+\frac{\Lambda_1}{\mu}\right\|_F^2+\left\|\mathbf{E}-\mathbf{Y}+\frac{\Lambda_2}{\mu}\right\|_F^2\right)\end{aligned}\tag{8}$$

where Λ_1 and Λ_2 are the dual variables, and $\mu>0$ is the penalty parameter. It can be observed from function (8) that it involves nonconvex nonsmooth terms and multiple variables, especially for two dual variables. To the best of our knowledge, few studies have been done on such generalized problem formulation, while it has some wide practices in pattern recognition and computer vision. This further motivates us to develop the designed algorithm involving two equality constraints, which can be regarded as an extension of convex ADMM [44], [52] to nonconvex [2], [47].

Inspired by the iteration procedures of ADMM, we need to update primal variables, dual variables, and penalty parameter in sequence. Given $(\mathbf{X}_k,\mathbf{Y}_k,\mathbf{Z}_k,\mathbf{E}_k,\Lambda_{1,k},\Lambda_{2,k},\mu_k)$ in the kth iteration step, then the $(k+1)$th iteration variables are

$$\begin{cases}\mathbf{X}_{k+1}\in\underset{\mathbf{X}}{\operatorname{argmin}}\ f_\lambda(\mathbf{X})+\dfrac{\mu_k}{2}\left\|\mathbf{X}-\left(\mathbf{Z}_k+\dfrac{\Lambda_{1,k}}{\mu_k}\right)\right\|_F^2 & (9)\\ \mathbf{Y}_{k+1}\in\underset{\mathbf{Y}}{\operatorname{argmin}}\ g(\mathbf{Y})+\dfrac{\mu_k}{2}\left\|\mathbf{Y}-\left(\mathbf{E}_k+\dfrac{\Lambda_{2,k}}{\mu_k}\right)\right\|_F^2 & (10)\\ \mathbf{Z}_{k+1}=\underset{\mathbf{Z}}{\operatorname{argmin}}\ \dfrac{\mu_k}{2}\left\|\mathbf{Z}-\widehat{\mathbf{X}}_{\Lambda_{1,k}}^{k+1}\right\|_F^2+\dfrac{\gamma}{2}\left\|\mathcal{A}(\mathbf{Z})-\widehat{\mathbf{E}}_{\mathbf{D}}^k\right\|_F^2\\ \text{with } \widehat{\mathbf{X}}_{\Lambda_{1,k}}^{k+1}=\mathbf{X}_{k+1}-\dfrac{\Lambda_{1,k}}{\mu_k},\quad \widehat{\mathbf{E}}_{\mathbf{D}}^k=\mathbf{D}-\mathcal{B}(\mathbf{E}_k) & (11)\\ \mathbf{E}_{k+1}=\underset{\mathbf{E}}{\operatorname{argmin}}\ \dfrac{\mu_k}{2}\left\|\mathbf{E}-\widehat{\mathbf{Y}}_{\Lambda_{2,k}}^{k+1}\right\|_F^2+\dfrac{\gamma}{2}\left\|\mathcal{B}(\mathbf{E})-\widehat{\mathbf{Z}}_{\mathbf{D}}^{k+1}\right\|_F^2\\ \text{with } \widehat{\mathbf{Y}}_{\Lambda_{2,k}}^{k+1}=\mathbf{Y}_{k+1}-\dfrac{\Lambda_{2,k}}{\mu_k},\quad \widehat{\mathbf{Z}}_{\mathbf{D}}^{k+1}=\mathbf{D}-\mathcal{A}(\mathbf{Z}_{k+1}) & (12)\\ \Lambda_{1,k+1}=\Lambda_{1,k}+\mu_k(\mathbf{Z}_{k+1}-\mathbf{X}_{k+1}) & (13)\\ \Lambda_{2,k+1}=\Lambda_{2,k}+\mu_k(\mathbf{E}_{k+1}-\mathbf{Y}_{k+1}) & (14)\\ \mu_{k+1}=\min\left(\alpha\mu_k,10^6\right),\quad \alpha>1 & (15)\end{cases}$$

where α determines the convergence speed and the quality of solvers, i.e., when the value of α is larger, the convergence rate becomes faster but generates a worse efficacy and vice versa. Meanwhile, with the increasing number of iterations, the values of μ_k become more larger from (15). Besides this, the resulting subproblems (11) and (12) need to minimize the quadratic problems, and the subproblems (9) and (10) involve to compute the proximal operators by Propositions 1 and 2 given later. As a result, the closed-form solvers can be achieved easily. Without any inner loop, the programming is very easy to be done, and the complexity mainly depends on the computations of SVD and matrix inverse at each of iterations. Moreover, computing low-rank matrix involves the economic SVD, while it has higher computational complexity, especially for the large-scale matrix. However, this issue is out of the research scope of this article because the decomposable strategy leads to too much variables. We focus

on nonconvex ADMM with two dual variables, which is very challenging for proving its convergence property [45], [48].

In the iteration procedures of (group) sparse vector coding, low-rank matrix recovery, and tensor decomposition problems, the proximal operators [41] played the key role in developing efficient first-order algorithms. Those nonconvex relaxations of L_0-norm[2] in Table II can be extended to relax rank function and $L_{2,0}$-norm by acting on the specific entries, respectively. Before computing the primal variables for achieving the closed-form solvers, Propositions 1 and 2 introduce several proximity operators for nonconvex relaxations of L_0-norm, $L_{2,0}$-norm, and rank function, respectively. Here, different nonconvex functions correspond to various p values, e.g., $p \in (0, 1.0)$ for the ℓ_p-norm, $p \in (1.1, +\infty)$ for the MCP function, and $p \in (0, 1.0)$ for the SCAD function. For showing the proximal operators of L_p-norm [2], MCP [30], and SCAD [31], we choose $t \in [0, 5]$ and set $\lambda = 1$; the plotted curves given in Fig. 1 have similar changes when giving the proper choices of the p values at its domain.

Proposition 1: For a proper and lower semicontinuous function $h_\lambda(\cdot)$, it may be nonconvex relaxations of L_0-norm, $L_{2,0}$-norm, or rank function. Then, the corresponding proximal operator $\mathrm{Prox}_\lambda^h(\cdot)$ [41] can be computed by

$$\begin{cases} \underset{\mathbf{s}\in\mathbb{R}^n}{\operatorname{argmin}}\ h_\lambda(\mathbf{s}) + \dfrac{1}{2}\|\mathbf{s}-\mathbf{t}\|_2^2, & (16) \\ \underset{\mathbf{S}\in\mathbb{R}^{p\times q}}{\operatorname{argmin}}\ \displaystyle\sum_i h_\lambda(\|\mathbf{S}_{:,i}|\|_2) + \frac{1}{2}\|\mathbf{S}-\mathbf{T}\|_F^2 & (17) \end{cases}$$

where $h_\lambda(\cdot)$ is defined on the entries (e.g., $\mathbf{s}_i$) of the vector, the column vectors (e.g., $\mathbf{S}_{:,i}$) of the matrix, respectively. The proximal operators of both (16) and (17) are denoted as

$$\mathrm{Prox}_\lambda^h(\mathbf{t}) = \left[\mathrm{Prox}_\lambda^h(\mathbf{t}_1), \mathrm{Prox}_\lambda^h(\mathbf{t}_2), \ldots, \mathrm{Prox}_\lambda^h(\mathbf{t}_n)\right] \quad (18)$$

and

$$\mathrm{Prox}_\lambda^h(\mathbf{T}) = \left[\mathrm{Prox}_\lambda^h(\mathbf{T}_{:,1}), \mathrm{Prox}_\lambda^h(\mathbf{T}_{:,2}), \ldots, \mathrm{Prox}_\lambda^h(\mathbf{T}_{:,q})\right] \quad (19)$$

where

$$\mathrm{Prox}_\lambda^h(\mathbf{T}_{:,i}) = \begin{cases} \mathbf{0}, & \mathbf{T}_{:,i} = \mathbf{0} \\ \dfrac{\mathrm{Prox}_\lambda^h(\|\mathbf{T}_{:,i}|\|_2)}{\|\mathbf{T}_{:,i}|\|_2}\mathbf{T}_{:,i}, & \text{otherwise.} \end{cases}$$

Proposition 2: Suppose that $h_\lambda(\cdot) : \mathbb{R}^+ \to \mathbb{R}^+$ and $\mathrm{Prox}_\lambda^h(\cdot)$ are both monotone, and let $\mathbf{T} = \mathbf{U}\mathrm{Diag}(\sigma(\mathbf{T}))\mathbf{V}^T$ be the SVD of $\mathbf{T} \in \mathbb{R}^{p\times q}$; then, with the usage of some a nonconvex function, such as L_p-norm, MCP, and SCAD, for acting on the singular values of a low-rank matrix in problem (7), then the GSVT operator [16], [53] can be computed by

$$\underset{\mathbf{S}\in\mathbb{R}^{p\times q}}{\operatorname{argmin}} \sum_i h_\lambda(\sigma_i(\mathbf{S})) + \frac{1}{2}\|\mathbf{S}-\mathbf{T}\|_F^2 \quad (20)$$

where the optimal solution of (20) can be represented as $\mathbf{S}^* = \mathbf{U}\mathrm{Diag}(\varrho^*(\mathbf{T}))\mathbf{V}^T$ with $\varrho^*(\mathbf{T}) = [\varrho_1^*(\mathbf{T}), \varrho_2^*(\mathbf{T}), \ldots, \varrho_r^*(\mathbf{T})]$, $\mathbf{U}\mathbf{U}^T = \mathbf{I}$, and $\mathbf{V}\mathbf{V}^T = \mathbf{I}$, in which $\varrho^*(\mathbf{T})$ satisfies $\varrho_1^*(\mathbf{T}) \geq \varrho_2^*(\mathbf{T}) \geq \cdots \geq \varrho_r^*(\mathbf{T})$ for $i = 1, 2, \ldots, r = \min(p, q)$, and we define $\varrho_i^*(\mathbf{T}) \in \mathrm{Prox}_\lambda^h(\sigma_i(\mathbf{T}))$ and set

$$\mathrm{Prox}_\lambda^h(\sigma_i(\mathbf{T})) = \underset{\varrho_i}{\operatorname{argmin}}\ h_\lambda(\varrho_i) + \frac{1}{2}\|\varrho_i - \sigma_i(\mathbf{T})\|_2^2. \quad (21)$$

It is obvious that Propositions 1 and 2 provide the generalized proximity operators instead of the concrete ones. In sequence, it is not difficult to update the primal variables, i.e., $\mathbf{X}^{k+1}$, $\mathbf{Y}^{k+1}$, $\mathbf{Z}^{k+1}$, and $\mathbf{E}^{k+1}$, through solving (9)–(12), respectively. In detail, by computing the derivative with $\mathbf{X}$ of the objective function in (11) and setting it to be $\mathbf{0}$, we can obtain the updated variable, denoted as

$$\mathbf{Z}_{k+1} = (\mu_k\mathbf{I} + \gamma\,\mathcal{A}^*\mathcal{A})^{-1}\Big[\mu_k\widehat{\mathbf{X}}_{\Lambda_{1,k}}^{k+1} + \gamma\,\mathcal{A}^*\big(\widehat{\mathbf{E}}_{\mathbf{D}}^k\big)\Big]. \quad (22)$$

Besides this, it follows from (12) that we have

$$\mathbf{E}_{k+1} = (\mu_k\mathbf{I} + \gamma\,\mathcal{B}^*\mathcal{B})^{-1}\Big[\mu_k\widehat{\mathbf{Y}}_{\Lambda_{2,k}}^{k+1} + \gamma\,\mathcal{B}^*\big(\widehat{\mathbf{Z}}_{\mathbf{D}}^{k+1}\big)\Big]. \quad (23)$$

However, when $\mathcal{A}$ and $\mathcal{B}$ are not the identity operators, and the dimensionality is higher, the optimization solvers of both (11) and (12) may be computational heavily. This is due to the fact that the computations of matrix inversion and multiplication are necessary in both (22) and (23) for updating $\mathbf{Z}_{k+1}$ and $\mathbf{E}_{k+1}$ at each of iterations. Besides this, the involved subproblems are least-squares problems, which can lead to achieve the analytic solvers easily.

In a different way, one can resort to using the linearization technique and then optimizing the surrogate minimization problem of (11), approximated by

$$\begin{aligned}\mathbf{Z}_{k+1} = \underset{\mathbf{Z}}{\operatorname{argmin}}\ &\frac{\mu_k}{2}\Big\|\mathbf{Z} - \widehat{\mathbf{X}}_{\Lambda_{1,k}}^{k+1}\Big\|_F^2 + \left(\frac{\gamma}{2}\Big\|\mathcal{A}(\mathbf{Z}_k) - \widehat{\mathbf{E}}_{\mathbf{D}}^k\Big\|_F^2\right. \\ &\left. + \langle\gamma\,\mathcal{A}^*\Big(\mathcal{A}(\mathbf{Z}_k) - \widehat{\mathbf{E}}_{\mathbf{D}}^k\Big), \mathbf{Z} - \mathbf{Z}_k\rangle + \frac{\eta_{\mathbf{Z}}}{2}\|\mathbf{Z} - \mathbf{Z}_k\|_F^2\right). \end{aligned} \quad (24)$$

Furthermore, we can obtain the approximation problem of (12). As a result, by computing the derivatives with $\mathbf{Z}$ and $\mathbf{E}$ accordingly, and setting both to be $\mathbf{0}$, both $\mathbf{Z}_{k+1}$ and $\mathbf{E}_{k+1}$ can be updated, respectively, by

$$\mathbf{Z}_{k+1} = \frac{1}{\mu_k + \eta_{\mathbf{Z}}}\Big(\mu_k\widehat{\mathbf{X}}_{\Lambda_{1,k}}^{k+1} + \eta_{\mathbf{Z}}\widetilde{\mathbf{Z}}_k\Big) \quad (25)$$

$$\mathbf{E}_{k+1} = \frac{1}{\mu_k + \eta_{\mathbf{E}}}\Big(\mu_k\widehat{\mathbf{Y}}_{\Lambda_{2,k}}^{k+1} + \eta_{\mathbf{E}}\widetilde{\mathbf{E}}_k\Big) \quad (26)$$

where $\widetilde{\mathbf{Z}}_k = \mathbf{Z}_k - (\gamma/\eta_{\mathbf{Z}})\mathcal{A}^*(\mathcal{A}(\mathbf{Z}_k) - \widehat{\mathbf{E}}_{\mathbf{D}}^k)$ and $\widetilde{\mathbf{E}}_k = \mathbf{E}_k - (\gamma/\eta_{\mathbf{E}})\mathcal{B}^*(\mathcal{B}(\mathbf{E}_k) - \widehat{\mathbf{Z}}_{\mathbf{D}}^{k+1})$. As suggested in [16] and [32], both $\eta_{\mathbf{Z}} \geq (\gamma\,\delta_{\mathcal{A}}^{\max}/2)$ and $\eta_{\mathbf{E}} \geq (\gamma\,\delta_{\mathcal{B}}^{\max}/2)$ would determine the step size in the iteration procedures. Then we assume that

$$\delta_{\mathcal{A}}^{\min}\mathcal{I}_{\mathcal{A}} \preceq \mathcal{A}^*\mathcal{A} \preceq \delta_{\mathcal{A}}^{\max}\mathcal{I}_{\mathcal{A}}, \quad \delta_{\mathcal{B}}^{\min}\mathcal{I}_{\mathcal{B}} \preceq \mathcal{B}^*\mathcal{B} \preceq \delta_{\mathcal{B}}^{\max}\mathcal{I}_{\mathcal{B}} \quad (27)$$

where $\delta_{\mathcal{A}}^{\min}$ and $\delta_{\mathcal{A}}^{\max}$ (resp. $\delta_{\mathcal{B}}^{\min}$ and $\delta_{\mathcal{B}}^{\max}$) are the smallest and largest eigenvalues of the linear map $\mathcal{A}^*\mathcal{A}$ (resp. $\mathcal{B}^*\mathcal{B}$), respectively. $\mathcal{I}_{\mathcal{A}}$ and $\mathcal{I}_{\mathcal{B}}$ are both the identity maps.

Let $f_\lambda(\mathbf{X})$ be the nonconvex relaxation of rank function; it follows from both (20) and (21) that we can obtain

$$\mathbf{X}_{k+1} = \widehat{\mathbf{U}}\mathrm{Diag}\Big(\varrho\Big(\widehat{\mathbf{Z}}_{\Lambda_{1,k}}^k\Big)\Big)\widehat{\mathbf{V}}^T \quad (28)$$

[2]Note the settings of parameters for L_p-norm with $0 < p < 1$, MCP with $p > 1$, and SCAD with $p > 2$, as well as the involved variables for the proximal operators of L_p-norm with $\nu_1 = \nu + \lambda p|\nu|^{p-1}$ for $\nu = [\lambda p(1-p)]^{1/(2-p)}$; t_1 and t_2 are the roots of $h(s) = (s-t) + \lambda p|s|^{p-1}\mathrm{sign}(s) = 0$ at $\nu < s < t$ and $t < s < -\nu$, respectively.

Algorithm 1 Optimization Scheme for Problem (7)

Input: $k=0$, μ_0, γ_0, $\alpha>1$, $\mu_{\max}$, $\gamma_{\max}$.
Initialization: $\mathbf{Z}_0$, $\mathbf{E}_0$, $\Lambda_{1,0}$, $\Lambda_{2,0}$.
Output: $(\mathbf{X}_*, \mathbf{Z}_*, \mathbf{E}_*, \Lambda_{1,*}, \Lambda_{2,*})$.

While not converged do
update primal variables
compute $\mathbf{X}_{k+1}$ by (28),
compute $\mathbf{Y}_{k+1}$ by (29) or (30),
compute $\mathbf{Z}_{k+1}$ by (22) or (25),
compute $\mathbf{E}_{k+1}$ by (23) or (26),
update dual variables
compute $\Lambda_{1,k+1}$ by (13),
compute $\Lambda_{2,k+1}$ by (14),
update penalty parameters
Update μ_{k+1} by (15),
until convergence when satisfying

$$\min\left(\frac{\|\mathbf{Z}_{k+1}-\mathbf{X}_{k+1}\|_2}{\sqrt{n}}, \frac{\|\mathbf{E}_{k+1}-\mathbf{Y}_{k+1}\|_2}{\sqrt{m}}\right)<\epsilon.$$

where $\widehat{\mathbf{Z}}^k_{\Lambda_{1,k}}=\widehat{\mathbf{U}}\mathrm{Diag}(\sigma(\widehat{\mathbf{Z}}^k_{\Lambda_{1,k}}))\widehat{\mathbf{V}}^T$ is the SVD of the matrix $\widehat{\mathbf{Z}}^k_{\Lambda_{1,k}}=\mathbf{Z}_k+(\Lambda_{1,k}/\mu_k)$ with $\widehat{\mathbf{U}}\widehat{\mathbf{U}}^T=\widehat{\mathbf{I}}$ and $\widehat{\mathbf{V}}\widehat{\mathbf{V}}^T=\widehat{\mathbf{I}}$.

On the one hand, when $g(\mathbf{Y})$ is a nonconvex relaxation of L_0-norm, it follows from (18) that we can achieve the optimal solver of subproblem (10), i.e., $\mathbf{Y}_{k+1}$ equals to

$$\begin{bmatrix} \mathrm{Prox}^g_{\mu_k}\left(\left[\widehat{\mathbf{E}}^k_{\Lambda_{1,k}}\right]_{11}\right) & \cdots & \mathrm{Prox}^g_{\mu_k}\left(\left[\widehat{\mathbf{E}}^k_{\Lambda_{1,k}}\right]_{1q}\right) \\ \mathrm{Prox}^g_{\mu_k}\left(\left[\widehat{\mathbf{E}}^k_{\Lambda_{1,k}}\right]_{21}\right) & \cdots & \mathrm{Prox}^g_{\mu_k}\left(\left[\widehat{\mathbf{E}}^k_{\Lambda_{1,k}}\right]_{2q}\right) \\ \vdots & \ddots & \vdots \\ \mathrm{Prox}^g_{\mu_k}\left(\left[\widehat{\mathbf{E}}^k_{\Lambda_{1,k}}\right]_{p1}\right) & \cdots & \mathrm{Prox}^g_{\mu_k}\left(\left[\widehat{\mathbf{E}}^k_{\Lambda_{1,k}}\right]_{pq}\right) \end{bmatrix} \tag{29}$$

where $\widehat{\mathbf{E}}^k_{\Lambda_{1,k}}=\mathbf{E}_k+(\Lambda_{2,k}/\mu_k)$. On the other hand, when $g(\mathbf{Y})$ is a nonconvex relaxation of $L_{2,0}$ norm, it follows from Proposition 1 that we have the optimal solver, denoted as

$$\mathbf{Y}_{k+1}=\mathrm{Prox}^g_{\mu_k}\left(\widehat{\mathbf{E}}^k_{\Lambda_{1,k}}\right). \tag{30}$$

In a summary, after updating the primal variables along with (13)–(15), the whole iteration procedures are completed in Algorithm 1. Here, the computational complexity depends on the matrix inverse and multiplier computations in (22) and (23) and the SVD computations in (28) at each iteration. Moreover, we check the convergence condition for stopping the iterations or reaching the maximal number of iterations.

III. Main Convergence Results

The convergence analysis of Algorithm 1 mainly replies to the updating rule of each variable. Before giving the main convergence results, we provide the basic proof routines in five technical parts in the following.

1) In the first part, we prove the boundedness of the dual variables by the primal variables (see Lemma 1).
2) In the second part, we derive the sufficient descent condition of a potential function and prove the boundedness of the variable sequence (see Lemma 2).
3) In the third part, we provide the local guarantees for the global convergence of the subsequence (see Theorem 1).
4) In the fourth part, we prove the supergradient of the potential function to be upper bounded (see Proposition 3).
5) Finally, we present the global guarantees for the global convergence of the generated sequence (see Theorem 2).

Subsequently, we only provide the convergence results due to the limited spaces, while several preliminaries [23], [54]–[56] and detailed proofs are given in the Supplementary Material.

First, we present several first-order optimality conditions of each subproblem for (9)–(12) at the $(k+1)$th iteration, which are denoted by

$$\begin{cases} \mathbf{0}\in\partial f_\lambda(\mathbf{X}_{k+1})+\mu_k\left[\mathbf{X}_{k+1}-\left(\mathbf{Z}_k+\dfrac{\Lambda_{1,k}}{\mu_k}\right)\right] \\ \quad =\partial f_\lambda(\mathbf{X}_{k+1})-\Lambda_{1,k+1}+\mu_k(\mathbf{Z}_{k+1}-\mathbf{Z}_k) & (31) \\ \mathbf{0}\in\partial g(\mathbf{Y}_{k+1})+\mu_k\left[\mathbf{Y}_{k+1}-\left(\mathbf{E}_k+\dfrac{\Lambda_{2,k}}{\mu_k}\right)\right] \\ \quad =\partial g(\mathbf{Y}_{k+1})-\Lambda_{2,k+1}+\mu_k(\mathbf{E}_{k+1}-\mathbf{E}_k) & (32) \\ \mathbf{0}=\Lambda_{1,k+1}+\gamma\,\mathcal{A}^*\big(\mathcal{A}(\mathbf{Z}_{k+1})-\widehat{\mathbf{E}}^k_{\mathbf{D}}\big) & (33) \\ \mathbf{0}=\Lambda_{2,k+1}+\gamma\,\mathcal{B}^*\big(\mathcal{B}(\mathbf{E}_{k+1})-\widehat{\mathbf{Z}}^{k+1}_{\mathbf{D}}\big) & (34) \end{cases}$$

where (31)–(34) will be used repeatedly in providing the convergence analysis. In fact, they are derived by the updating rules of $\Lambda_{1,k+1}$ and $\Lambda_{2,k+1}$, respectively.

Borrowing from (24) to (26), we can compute the optimality conditions of function (8), which is related with $\mathbf{Z}_{k+1}$ and $\mathbf{E}_{k+1}$ for the inexact situations, i.e.,

$$\begin{cases} \mathbf{0}=\Lambda_{1,k+1}+\gamma\,\mathcal{A}^*(\mathcal{A}_{\mathbf{ZE}})+\eta_{\mathbf{Z}}(\mathbf{Z}_{k+1}-\mathbf{Z}_k) & (35) \\ \mathbf{0}=\Lambda_{2,k+1}+\gamma\,\mathcal{B}^*(\mathcal{B}_{\mathbf{ZE}})+\eta_{\mathbf{E}}(\mathbf{E}_{k+1}-\mathbf{E}_k) & (36) \end{cases}$$

where $\mathcal{A}_{\mathbf{ZE}}=\mathcal{A}(\mathbf{Z}_k)-\widehat{\mathbf{E}}^k_{\mathbf{D}}$ and $\mathcal{B}_{\mathbf{ZE}}=\mathcal{B}(\mathbf{E}_{k+1})-\widehat{\mathbf{Z}}^{k+1}_{\mathbf{D}}$. Both (35) and (36) are the revisions of (33) and (34), where the differences are actually derived from the usage of linearization strategy for the quadratic components.

Lemma 1: Let $\{\mathcal{T}_k\}_{k=1}^{+\infty}$ be generated by Algorithm 1; there exist the constants $c_1, c_2, c_3>0$ such that

$$\begin{aligned} &\|\Lambda_{1,k+1}-\Lambda_{1,k}\|_F^2 \\ &\quad\le c_1\|\mathbf{Z}_{k+1}-\mathbf{Z}_k\|_F^2+c_2\|\mathbf{E}_k-\mathbf{E}_{k-1}\|_F^2 & (37) \\ &\|\Lambda_{2,k+1}-\Lambda_{2,k}\|_F^2 \\ &\quad\le c_2\|\mathbf{Z}_{k+1}-\mathbf{Z}_k\|_F^2+c_3\|\mathbf{E}_{k+1}-\mathbf{E}_k\|_F^2. & (38) \end{aligned}$$

Define $\mathcal{T}_k=(\mathbf{X}_k,\mathbf{Y}_k,\mathbf{Z}_k,\mathbf{E}_k,\Lambda_{1,k},\Lambda_{2,k})$, it follows from (8) that the supergradient or gradient [57] of the function

$\mathcal{L}_{\mu_{k+1}}(\mathcal{T}_{k+1})$ at each of variables can be given as

$$\begin{cases} \partial_{\mathbf{X}}\mathcal{L}_{\mu_{k+1}}(\mathcal{T}_{k+1})|_{\mathbf{X}=\mathbf{X}_{k+1}} \\ \quad = \partial f_\lambda(\mathbf{X}_{k+1}) - \Lambda_{1,k+1} - \dfrac{\mu_{k+1}}{\mu_k}\big(\Lambda_{1,k+1} - \Lambda_{1,k}\big) & (39) \\ \partial_{\mathbf{Y}}\mathcal{L}_{\mu_{k+1}}(\mathcal{T}_{k+1})|_{\mathbf{Y}=\mathbf{Y}_{k+1}} \\ \quad = \partial g(\mathbf{Y}_{k+1}) - \Lambda_{2,k+1} - \dfrac{\mu_{k+1}}{\mu_k}\big(\Lambda_{2,k+1} - \Lambda_{2,k}\big) & (40) \\ \partial_{\mathbf{Z}}\mathcal{L}_{\mu_{k+1}}(\mathcal{T}_{k+1})|_{\mathbf{Z}=\mathbf{Z}_{k+1}} \\ \quad = \gamma\,\mathcal{A}^*(\mathcal{B}(\mathbf{E}_{k+1}) - \mathcal{B}(\mathbf{E}_k)) + \dfrac{\mu_{k+1}}{\mu_k}\big(\Lambda_{1,k+1} - \Lambda_{1,k}\big) & (41) \\ \partial_{\mathbf{E}}\mathcal{L}_{\mu_{k+1}}(\mathcal{T}_{k+1})|_{\mathbf{E}=\mathbf{E}_{k+1}} = \dfrac{\mu_{k+1}}{\mu_k}\big(\Lambda_{2,k+1} - \Lambda_{2,k}\big) & (42) \\ \partial_{\Lambda_1}\mathcal{L}_{\mu_{k+1}}(\mathcal{T}_{k+1})|_{\Lambda_1=\Lambda_{1,k+1}} = \dfrac{\Lambda_{1,k+1} - \Lambda_{1,k}}{\mu_k} & (43) \\ \partial_{\Lambda_2}\mathcal{L}_{\mu_{k+1}}(\mathcal{T}_{k+1})|_{\Lambda_2=\Lambda_{2,k+1}} = \dfrac{\Lambda_{2,k+1} - \Lambda_{2,k}}{\mu_k} & (44) \end{cases}$$

where both (39) and (40) can be achieved by computing the supergradient of $\mathcal{L}_{\mu_{k+1}}(\mathcal{T}_{k+1})$ with $\mathbf{X}$ and $\mathbf{Y}$ through (13) and (14), respectively. Similarly, it follows from (8) that both (41) and (42) can be obtained by (33) and (34) accordingly, as well as both (43) and (44) can be get easily. In the proof process of convergence analysis, the above given (31)–(34) and (39)–(44) will be used repeatedly.

Instead of using the function (8) directly, we next need to construct the potential function, represented as

$$\widehat{\mathcal{L}}_{\mu_k}(\mathcal{T}_k, \widetilde{\mathbf{E}}) = \mathcal{L}_{\mu_k}(\mathcal{T}_k) + \frac{\mu_{k+1} + \mu_k}{2\mu_k^2} c_2 \|\mathbf{E}_k - \widetilde{\mathbf{E}}\|_F^2. \quad (45)$$

Then, we will prove the nonincreasing property of (45) and show the boundedness of the variable sequences.

Lemma 2: Let $\{\mathcal{T}_k\}_{k=1}^{+\infty}$ be the variable sequence generated by Algorithm 1; then, we have the following.

1) The potential function is nonincreasing, i.e., there exist the constants $c_4, c_5 > 0$ such that

$$\widehat{\mathcal{L}}_{\mu_{k+1}}(\mathcal{T}_{k+1}, \mathbf{E}_k) + c_4\|\mathbf{Z}_{k+1} - \mathbf{Z}_k\|_F^2 + c_5\|\mathbf{E}_{k+1} - \mathbf{E}_k\|_F^2 \le \widehat{\mathcal{L}}_{\mu_k}(\mathcal{T}_k, \mathbf{E}_{k-1}). \quad (46)$$

2) The sequence $\{\mathcal{T}_k\}_{k=1}^{\infty}$ is bounded.

Remark 1: The proofs of Lemmas 1 and 2 hold for the exact cases, where the solvers of (22) and (23) are analytic. Similar proofs also hold for the inexact situations, i.e., approximately updating the variables $\mathbf{Z}_{k+1}$ and $\mathbf{E}_{k+1}$ via (24)–(26). It should be noted that there exist some revisions stated as follows.

1) In Lemma 1, the conclusions also hold for (24)–(26) but need to add the quadratic terms, i.e., $\|\mathbf{Z}_k - \mathbf{Z}_{k-1}\|_F^2$ and $\|\mathbf{E}_k - \mathbf{E}_{k-1}\|_F^2$, in the right-hand sides of both (37) and (38), respectively. Naturally, they determine the formulation of the potential function for defining $c_1' = c_1 + 2\eta_{\mathbf{Z}}^2$ and $c_2' = c_2 + 2\eta_{\mathbf{E}}^2$, represented by

$$\widehat{\mathcal{L}}_{\mu_k}(\mathcal{T}_k, \widetilde{\mathbf{Z}}, \widetilde{\mathbf{E}}) = \mathcal{L}_{\mu_k}(\mathcal{T}_k) + \frac{\mu_{k+1} + \mu_k}{2\mu_k^2} \times \big(c_1'\|\mathbf{Z}_k - \widetilde{\mathbf{Z}}\|_F^2 + c_2'\|\mathbf{E}_k - \widetilde{\mathbf{E}}\|_F^2\big) \quad (47)$$

which can be regarded as the modified (45), and it played a key role for the local and global convergence analyses.

2) In Lemma 2, we further derive the revisions through (24)–(26), which can lead to some changeable details when we conduct the proofs for the boundedness of variable sequence under the inexact case. Meanwhile, both $\Lambda_{1,k+1}$ and $\Lambda_{2,k+1}$ will be used repeatedly for producing the corresponding revisions to the convergence analysis.

TABLE III

MODEL NAMES AND OBJECTIVE FORMULATIONS FOR BOTH LRR AND RMC MINIMIZATION PROBLEMS AND THEIR VARIANTS

Names	Problem Formulations
RMC	$\min_{\mathbf{Z},\mathbf{E}} \sum_{i=1}^{r} f_\lambda(\sigma_i(\mathbf{Z})) + \sum_{i=1}^{m}\sum_{j=1}^{n} g(P_\Omega(\mathbf{E}_{i,j}))$ $s.t.,\ P_\Omega(\mathbf{Z}) + P_\Omega(\mathbf{E}) = P_\Omega(\mathbf{D}).$
mRMC	$\min_{\mathbf{X},\mathbf{Y},\mathbf{Z},\mathbf{E}} \sum_{i=1}^{r} f_\lambda(\sigma_i(\mathbf{X})) + \sum_{i=1}^{m}\sum_{j=1}^{n} g(P_\Omega(\mathbf{Y}_{i,j})) + \frac{\gamma}{2}\|P_\Omega(\mathbf{Z}) + P_\Omega(\mathbf{E}) - P_\Omega(\mathbf{D})\|_F^2$ $s.t.,\ \mathbf{Z} = \mathbf{X},\quad \mathbf{E} = \mathbf{Y}.$
LRR	$\min_{\mathbf{Z},\mathbf{E}} \sum_{i=1}^{r} f_\lambda(\sigma_i(\mathbf{Z})) + \sum_{j=1}^{n} g(\|\mathbf{E}_j\|_2)$ $s.t.,\ \mathbf{AZ} + \mathbf{E} = \mathbf{D}.$
mLRR	$\min_{\mathbf{X},\mathbf{Y},\mathbf{Z},\mathbf{E}} \sum_{i=1}^{r} f_\lambda(\sigma_i(\mathbf{X})) + \sum_{j=1}^{n} g(\|\mathbf{Y}_j\|_2) + \frac{\gamma}{2}\|\mathbf{AZ} + \mathbf{E} - \mathbf{D}\|_F^2$ $s.t.,\ \mathbf{Z} = \mathbf{X},\quad \mathbf{E} = \mathbf{Y}.$

Theorem 1 (Global Convergence Property of the Subsequence): Suppose that $\{\mathcal{T}_k\}_{k=1}^{+\infty}$ is the sequence generated by Algorithm 1; the objective function satisfies Assumptions A1 and A2. Then, we can achieve the following.

1) $\lim_{k\to+\infty} \|\mathcal{T}_{k+1} - \mathcal{T}_k\|_F^2 = 0$.
2) Any cluster point $\mathcal{T}_* = (\mathbf{X}_*, \mathbf{Y}_*, \mathbf{Z}_*, \mathbf{E}_*, \Lambda_{1,*}, \Lambda_{2,k})$ of the generated sequence is a stationary point of (8), and it also satisfies the KKT conditions.

Proposition 3 Suppose that the same assumptions with Theorem 1 satisfy; there exists a constant $c_8 > 0$ such that

$$\begin{aligned} &\operatorname{dist}(\mathbf{0}, \partial\widehat{\mathcal{L}}_{\mu_{k+1}}(\mathcal{T}_{k+1}, \mathbf{E}_k)) \\ &\quad \le c_8(\|\mathbf{Z}_{k+1} - \mathbf{Z}_k\|_F + \|\mathbf{E}_{k+1} - \mathbf{E}_k\|_F \\ &\qquad + \|\Lambda_{1,k+1} - \Lambda_{1,k}\|_F + \|\Lambda_{2,k+1} - \Lambda_{2,k}\|_F). \end{aligned} \quad (48)$$

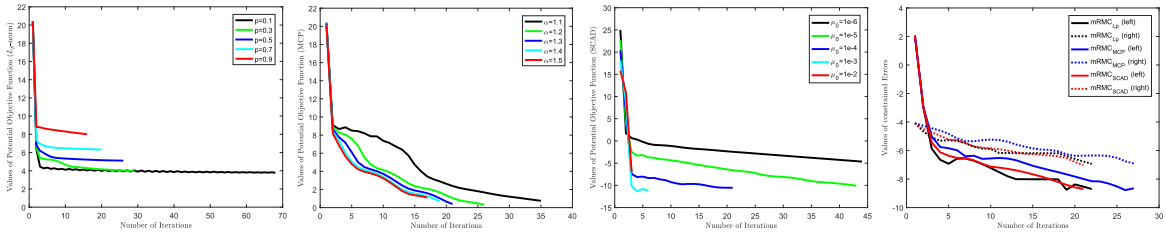
Based on the above conclusions, we subsequently prove the key result for the global convergence property.

Theorem 2 (Global Convergence Property of the Generated Sequence): Suppose that Assumptions A1 and A2 hold for the potential function $\widehat{\mathcal{L}}_{\mu_k}(\mathcal{T}_k, \mathbf{E}_{k-1})$ (45), which is a KŁ function; then, the sequence $\{\mathcal{T}_k\}_{k=1}^{+\infty}$ generated by Algorithm 1 is a Cauchy one, which has a finite length, i.e., $\sum_{k=1}^{+\infty}\|\mathcal{T}_{k+1} - \mathcal{T}_k\|_F < +\infty$. Moreover, the sequence $\{\mathcal{T}_k\}_{k=1}^{+\infty}$ converges to $\mathcal{T}_*$, which is a critical point of problem (8).

Remark 2: For the convergence analysis of Proposition 3 and Theorem 2 under the inexact case, the main revisions are located at computing the gradients of (45) with $\mathbf{Z}$ and $\mathbf{E}$, as well as involving the terms related with both $\Lambda_{1,k+1}$ and $\Lambda_{2,k+1}$. Based on these correlations, it needs several revisions through (24)–(26), which is useful for guaranteeing the local and global convergence properties.

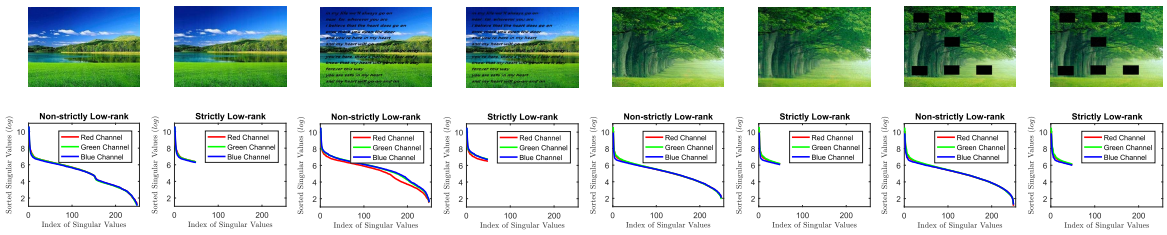
IV. NUMERICAL EXPERIMENTS

In this section, we conduct numerical experiments on both synthesis data and real-world data. For the synthesis data, we discuss the influences of several parameters for the convergence curves derived by mRMC and mLRR provided in



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

Fig. 2. Plotted curves of the potential objective function in (a)–(c) and the constrained errors in (d) derived by mRMC for three nonconvex functions.



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

Fig. 3. Sorted singular values of three channels in the second row of (a)–(d) for nonstrictly and strictly low-rank of the original images and the masked images in the first row of (a)–(d).

Table III. Both of them can be regarded as the modified variants of RMC and LRR, respectively. For the real-world data, we mainly test the recovery ability for the tasks of image inpainting and subspace clustering. The inpainting results are evaluated by the peak signal-to-noise ratio (PSNR) and the relative error (RelErr), while the clustering results are evaluated by segmentation accuracy. The timing costs are provided for the comparisons of computational efficiency. All the comparison methods are implemented by the released codes. Furthermore, the best results are reported by tuning the involved parameters suggested by the published papers. For the penalty parameter γ of problem (7), the continuity strategy will be used by choosing an initial value γ_0 and increasing it by multiplying $\alpha > 1$ until reaching the target one, i.e., $\gamma_{\max} = 1e + 6$. For the initial variables, we set $\mathbf{Z}_0 = \mathbf{0}$, $\mathbf{E}_0 = \Lambda_{2,0} = \mathbf{0}$, and $\Lambda_{1,0} = \mathbf{I}$. All of the numerical experiments are performed by the MATLAB R2016b on a PC with Intel[3] Core[4] i7-4770 CPU @ 3.40 GHz and 16.0 GB of RAM. Subsequently, we mainly introduce the implementation details and make qualitative and quantitative comparisons.

1) *Experiments of mRMC:* It follows from [2], [16], and [26] that the first synthesis experiments construct the matrix $\mathbf{X} = \mathbf{U}_1\mathbf{U}_2$ with rank $\mathbf{r} = 80$, which is generated by virtue of *randn* for $\mathbf{U}_1 = randn(400, \mathbf{r})$ and $\mathbf{U}_2 = randn(\mathbf{r}, 400)$, respectively. 20% of elements in $\mathbf{X}$ are selected randomly as $\mathbf{\Omega}$ to be missing, where the observed matrix is denoted as $\mathcal{P}_{\mathbf{\Omega}}(\mathbf{X})$. The second real-world experiments test the incomplete mask with text and block missing on the natural images, where the sorted singular values are illustrated for nonstrictly low-rank and strictly one. Each of color images has red, green, and blue channels; we need to rehabilitate the missing pixels on each channel independently and then combine them to get the final results. The comparison methods are related to nuclear norm and its weighted/truncated variants, as well as the extended Schatten-p norm, while this work focuses on employing nonconvex ADMM to optimize the mRMC problems derived from L_p-norm with $p \in (0, 1)$ [2], MCP [30] with $p \in (1.1, +\infty)$, and SCAD [31] with $p \in (2.1, +\infty)$, respectively.

[3]Registered trademark.
[4]Trademarked.

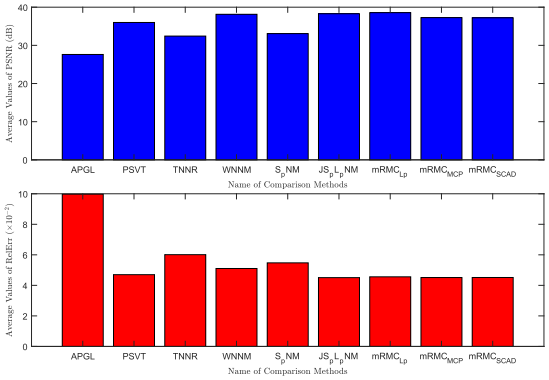


Fig. 4. Comparisons of average values of (a) PSNR and (b) RelErr for all the involved methods to the nonstrictly and strictly low-rank images.

a) Fig. 2(a)–(c) illustrates the plotted convergence curves (log) of potential objective functions derived by the mRMC with L_p-norm, MCP, and SCAD for the proper choices of parameters, i.e., $p \in \{0.1, 0.3, 0.5, 0.7, 0.9\}$, $\alpha \in \{1.1, 1.2, 1.3, 1.4, 1.5\}$, and $\mu_0 \in \{1e-6, 1e-5, 1e-4, 1e-3, 1e-2\}$, respectively. The nonincreasing properties are validated from the numerical settings and the visual results. Besides this, the stopping conditions provided in Fig. 2(d), where $((\|\mathbf{Z}_{k+1} - \mathbf{X}_{k+1}\|_2)/\sqrt{n})$ and $((\|\mathbf{E}_{k+1} - \mathbf{Y}_{k+1}\|_2)/\sqrt{m})$, are substituted by "left" and "right" for the representation simplicity. For an accurate solution, the above parameters should

TABLE IV

COMPARISONS OF INPAINTING RESULTS ON NONSTRICTLY AND STRICTLY LOW-RANK IMAGES WITH TEXT AND RANDOM MISSING PIXELS. NOTE THAT PSNR IS MEASURED BY DB, AND RELERR IS SCALED BY $\times 10^{-2}$, AND TIME IS IN SECONDS

Methods	(b) Non-strictly			(b) Strictly			(d) Non-strictly			(d) Strictly		
	PSNR	RelErr	Time	PSNR	RelErr	Time	PSNR	RelErr	Time	PSNR	RelErr	Time
APGL	25.432	11.116	**0.92**	29.387	7.749	**0.65**	26.226	12.115	**1.51**	29.399	8.931	**1.13**
PSVT	**31.381**	**5.604**	*3.32*	46.796	1.044	*2.81*	*29.908*	*7.929*	*3.03*	35.896	4.227	*2.42*
TNNR	29.694	6.806	18.61	37.028	3.215	15.43	29.086	8.717	19.09	33.917	5.308	14.19
WNNM	28.590	7.728	12.45	**50.646**	**0.671**	3.01	27.302	10.703	11.89	**46.019**	**1.318**	3.97
S_pNM	*31.231*	*5.702*	12.28	37.155	3.168	10.87	**30.513**	**7.396**	12.88	33.402	5.633	12.18
JS$_p L_p$NM	29.969	6.594	17.71	*50.504*	*0.681*	8.99	28.730	9.081	17.11	43.946	1.673	9.02
mRMC$_{L_p}$	29.749	6.763	23.88	50.019	0.721	13.25	28.460	9.367	21.96	*45.554*	*1.390*	17.36
mRMC$_{\text{MCP}}$	30.940	5.897	32.73	48.860	0.823	27.29	29.014	8.788	32.82	40.310	2.543	26.01
mRMC$_{\text{SCAD}}$	31.163	5.747	51.46	48.791	0.830	48.15	28.956	8.847	40.21	39.987	2.639	34.93

TABLE V

CLUSTERING ACCURACY (%) OF DIFFERENT SUBSPACE SEGMENTATION METHODS ON THE EXTENDED YALE B DATABASE

Methods	2 Subjects		3 Subjects		5 Subjects		8 Subjects		10 Subjects	
	Mean	Std	Mean	Std	Mean	Std	Mean	Std	Mean	Std
LRR	94.609	14.496	93.958	12.336	91.891	9.587	90.078	10.712	86.688	6.249
LADM	94.492	14.467	91.510	13.358	90.859	9.649	89.795	7.954	82.672	7.176
IRLS	94.609	14.496	95.651	10.583	91.906	9.617	90.117	10.715	85.500	6.195
$S_{2/3}$NM	94.297	14.415	95.287	10.562	92.828	8.409	92.793	5.568	*90.563*	5.980
$S_{1/2}$NM	94.219	14.495	*96.458*	*8.135*	92.859	8.030	91.221	5.419	88.539	*5.835*
mLRR$_{L_p}$	*96.289*	**10.116**	95.547	9.528	92.563	7.848	*93.565*	*4.077*	**90.961**	**4.976**
mLRR$_{\text{MCP}}$	*96.289*	*10.145*	**96.589**	**7.885**	*93.875*	*6.356*	92.881	4.582	89.375	6.568
mLRR$_{\text{SCAD}}$	**96.484**	10.147	95.469	9.492	**94.453**	**6.234**	**94.434**	**4.104**	90.063	6.022

not be too small and too large. We can see that $p = 0.5$, $\alpha = 1.1$, and $\mu_0 = 1e-3$ work well.

b) Fig. 3(a)–(d) presents two natural images with/without text and block missing elements (top), where nonstrictly low-rank and strictly low-rank are considered for the sorted singular values of the three channels. The empirical distributions (bottom) have shown that there exist some obvious differences, where the nonstrictly low-rank reflects all numbers of nonzero singular values of the original data matrix, while the strictly low-rank retains the largest 20% singular values.

c) Fig. 4 shows the average values of PSNR and RelErr of all the methods in both (a) and (b), where the proposed methods achieve the nearly best recovery performance. In detail, Table IV shows the values of PSNR, RelErr, and time, respectively, where noncovex methods can achieve better performance than convex APGL, and strictly low-rank cases have the values of higher PSNR and lower RelErr than nonstrictly ones. Moreover, we presented black marks for the best results and used italics for the second best results. The optimal parameters are set $p \in \{0.5, 1.5, 2.5, 5.0, 10.0\}$ and $\lambda \in \{0.5, 1.0, 1.5, 2.0, 2.5\}$ along with $\gamma_0 = p/\text{norm}(\mathbf{D})$ for the superior performance. We can see that both WNNM and JS$_p L_p$M have obvious superiority, especially in strictly low-rank cases. Besides this, our methods have much more timing costs than other comparison methods because of updating too many variables. Hence, due to the heavier computational loads for large-scale problems, developing lower complexity algorithms will be very challenging.

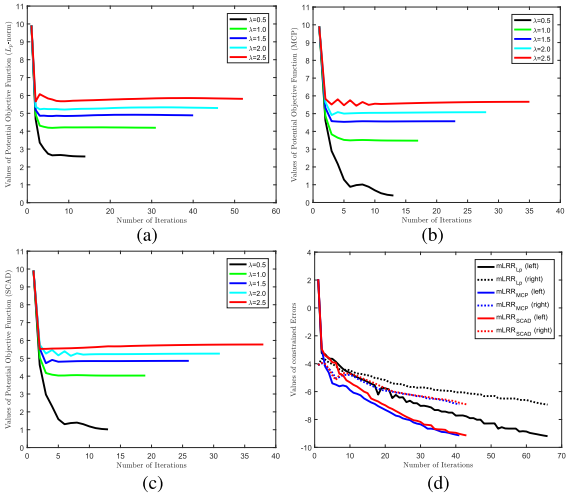


Fig. 5. Plotted curves of the potential objective function in (a)–(c) and the constrained errors in (d) derived by mLRR for three nonconvex functions.

2) *Experiments of mLRR:* It follows from [44], [45], and [58] that the first synthesis experiments construct 15 independent subspaces $\{\mathbf{S}_i\}_{i=1}^k$, whose bases $\{\mathbf{U}_i\}_{i=1}^k$ are generated by updating the sequences $\mathbf{U}_{i+1} = \mathbf{T}\mathbf{U}_i$ for $1 \le i \le 15$, where $\mathbf{T}$ is a random rotation matrix, $\mathbf{U}_1$ is a random column orthogonal matrix of dimension 200×5, and each subspace has a rank of 5 and an ambient dimension of 200. We sample 20 data vectors from each subspace by $\mathbf{X}_i = \mathbf{U}_i\mathbf{C}_i$, $1 \le i \le 15$, with $\mathbf{C}_i$ being a 5×40 i.i.d. standard Gaussian matrix and then randomly choose 20% samples to be corrupted by adding Gaussian noise with zero mean and standard deviation $0.5\|\mathbf{X}\|_F^2$. The second experiments are performed on the Extended Yale B database, which contains 38 individuals for 64 frontal face images. We downsample all the

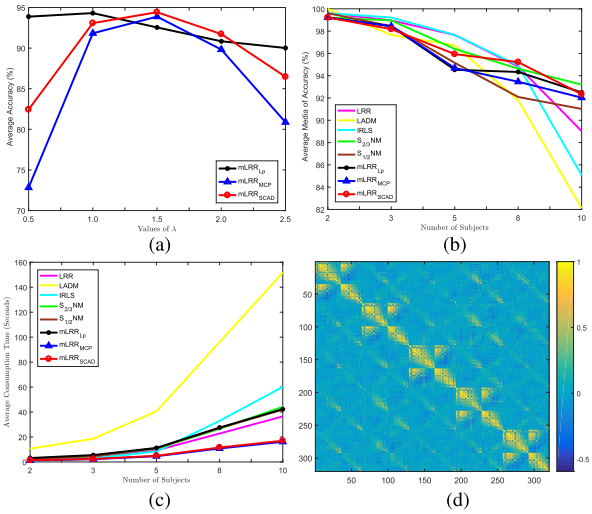


Fig. 6. Numerical and visual comparisons of mLRR for the performance analysis and computation time in (a)–(c) and the block-diagonal affinity matrix in (d) on face clustering.

original images into 48 × 42 pixels and vectorize each of them as a data point, which is normalized to have a unit length. Following the similar settings in [33] and [59], this task randomly samples k subjects face images from all 38 subjects to generate the data matrix $\mathbf{D} \in \mathbb{R}^{d\times 64k}$, where $d = 2016$ and $k \in \{2, 3, 5, 8, 10\}$. The experiments for all these sets are performed to run 20 trials, and then, the mean, media, and std of clustering accuracy and the mean of timing costs are reported for the resulting evaluations. For the comparison results, we have the following observations.

a) Fig. 5(a)–(c) presents the convergence curves (log) of objective function for mLRR with three nonconvex relaxations, i.e., L_p-norm, MCP, and SCAD. It can be seen that larger values of $\lambda \in \{0.5, 1.0, 1.5, 2.0, 2.5\}$ correspond to larger values of potential objective function along with too much number of iterations. Along with other tunable parameters, e.g., p, μ_0 and $\gamma_0 = 0.1p \times \text{norm}(\mathbf{D})$, the optimal choices follow from the settings in the experiments of mRMC. Based on the stopping conditions, we also provide the comparisons of constrained errors in (d). It can be seen that all of the plotted convergence curves have the nonincreasing property, which accords to the theoretical results.

b) Table V shows the clustering results evaluated by the mean values of accuracy and Std, in which noncovex methods can achieve better segmentation performance than convex ones [32], [44]. With the increasing number of subjects, the values of accuracy and Std become smaller. It follows from [45] that LRR, LADM, and IRLS were studied for solving nuclear norm minimization problems, while the proposed algorithms focus on optimizing the mLRR problems derived by $L_{0.5}$-norm, MCP, and SCAD, respectively. The results show that the proposed methods obtain a relatively higher clustering efficacy by choosing $\lambda = 1.5$.

c) Fig. 6(a) provides how the average accuracy changes with various λ values. The plotted curves of the average media of accuracy and computation time are given for the number of subjects in both (b) and (c). Moreover, the block-diagonal affinity matrix for 5 subjects is illustrated in (d). It is evident that the proposed mLRR solutions have satisfying performance and running speed than other related methods.

Based on these numerical results, it is obvious that the qualitative and quantitative comparisons can present the superiority of proposed algorithms over the compared ones on the real-world databases. Moreover, the theoretical results are verified by the plotted convergence curves on the synthesis data.

V. Conclusion

In this work, we focus on providing the local and global convergence properties of nonconvex ADMM with two dual variables, which can be used to solve a class of nonconvex low-rank matrix minimization problems. By constructing a potential function and relying on the iteration rules, we prove the boundedness of dual variables and the sufficient descent condition along with the boundedness of primal variables and also provide the global convergence of the subsequence. Furthermore, we prove the supergradient of the potential function to be upper bounded and the global convergence property of the generated variable sequence. In the experiments of synthetic and real-world data, the numerical and theoretic results maintain consistency. Besides this, the developed solutions can keep the well superiority over several mostly related optimization approaches when applying to the tasks of image inpainting and subspace clustering.

Acknowledgment

The authors would like to thank the editors and the anonymous reviewers for their constructive comments in helping them to improve this article.

References

[1] Q. Yao and J. T. Kwok, "Efficient learning with a family of nonconvex regularizers by redistributing nonconvexity," *J. Mach. Learn. Res.*, vol. 18, pp. 1–52, Jan. 2018.

[2] F. Nie, H. Wang, H. Huang, and C. Ding, "Joint Schatten p-norm and ℓ_p-norm robust matrix completion for missing value recovery," *Knowl. Inf. Syst.*, vol. 42, no. 3, pp. 525–544, 2015.

[3] 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.

[4] L. Yang, T. Pong, and X. Chen, "Alternating direction method of multipliers for nonconvex background/foreground extraction," *SIAM J. Imag. Sci.*, vol. 10, no. 1, pp. 74–110, 2017.

[5] X. Peng, Z. Yu, H. Tang, and Z. Yi, "Constructing the L_2-graph for robust subspace learning and subspace clustering," *IEEE Trans. Cybern.*, vol. 47, no. 4, pp. 1053–1066, Apr. 2017.

[6] J. Chen and J. Yang, "Robust subspace segmentation via low-rank representation," *IEEE Trans. Cybern.*, vol. 44, no. 8, pp. 1432–1445, Aug. 2014.

[7] 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.

[8] W. Dong, G. Shi, X. Li, Y. Ma, and F. Huang, "Compressive sensing via nonlocal low-rank regularization," *IEEE Trans. Image Process.*, vol. 23, no. 8, pp. 3618–3632, Aug. 2014.

[9] 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.

[10] J. Wright, A. Ganesh, S. Rao, Y. Peng, and Y. Ma, "Robust principal component analysis: Exact recovery of corrupted low-rank matrices via convex optimization," in *Proc. Adv. Neural Inf. Process. Syst. (NIPS)*, 2009, pp. 2080–2088.

[11] 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.

[12] Y. Chen, A. Jalali, S. Sanghavi, and C. Caramanis, "Low-rank matrix recovery from errors and erasures," *IEEE Trans. Inf. Theory*, vol. 59, no. 7, pp. 4324–4337, Jul. 2013.

[13] X. Peng, C. Lu, and Z. Yi, "Connections between nuclear-norm and frobenius-norm-based representations," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 29, no. 1, pp. 218–224, Jan. 2017.

[14] T. Bouwmans, A. Sobral, S. Javed, S. K. Jung, and E.-H. Zahzah, "Decomposition into low-rank plus additive matrices for background/foreground separation: A review for a comparative evaluation with a large-scale dataset," *Comput. Sci. Rev.*, vol. 23, pp. 1–71, Feb. 2017.

[15] 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.

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

[17] W. Jiang, J. Liu, H. Qi, and Q. Dai, "Robust subspace segmentation via nonconvex low rank representation," *Inf. Sci.*, vol. 340, pp. 144–158, May 2016.

[18] 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.

[19] Z.-F. Jin *et al.*, "An alternating direction method with continuation for nonconvex low rank minimization," *J. Sci. Comput.*, vol. 66, no. 2, pp. 849–869, Feb. 2016.

[20] K. Guo, D. R. Han, and T. T. Wu, "Convergence of alternating direction method for minimizing sum of two nonconvex functions with linear constraints," *Int. J. Comput. Math.*, vol. 94, no. 8, pp. 1653–1669, 2017.

[21] K. Guo, D. Han, D. Z. W. Wang, and T. Wu, "Convergence of ADMM for multi-block nonconvex separable optimization models," *Frontiers Math. China*, vol. 12, no. 5, pp. 1139–1162, 2017.

[22] M. Hong, Z.-Q. Luo, and M. Razaviyayn, "Convergence analysis of alternating direction method of multipliers for a family of nonconvex problems," *SIAM J. Optim.*, vol. 26, no. 1, pp. 337–364, Jan. 2016.

[23] 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.

[24] 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.*, vol. 30, no. 9, pp. 2825–2839, Sep. 2019.

[25] 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.

[26] 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.

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

[28] R. Chartrand and V. Staneva, "Restricted isometry properties and nonconvex compressive sensing," *Inverse Problems*, vol. 24, no. 3, pp. 657–682, 2008.

[29] 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.*, Dec. 2013, pp. 217–224.

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

[31] 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.

[32] Z. Lin, R. Liu, and Z. Su, "Linearized alternating direction method with adaptive penalty for low-rank representation," in *Proc. Adv. Neural Inf. Process. Syst. (NIPS)*, 2011, pp. 612–620.

[33] J. Xu *et al.*, "Scaled simplex representation for subspace clustering," *IEEE Trans. Cybern.*, vol. 51, no. 3, pp. 1493–1505, Mar. 2021.

[34] 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.

[35] X. Wang, J. Yan, B. Jin, and W. Li, "Distributed and parallel ADMM for structured nonconvex optimization problem," *IEEE Trans. Cybern.*, vol. 51, no. 9, pp. 4540–4552, Sep. 2021.

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

[37] 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.

[38] 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.

[39] C. Chen, J. Zhang, L. Shen, P. Zhao, and Z. Luo, "Communication efficient primal-dual algorithm for nonconvex nonsmooth distributed optimization," in *Proc. Int. Conf. Artif. Intell. Statist. (ICAIS)*, 2021, pp. 1594–1602.

[40] Z. Jia, X. Gao, X. Cai, and D. Han, "Local linear convergence of the alternating direction method of multipliers for nonconvex separable optimization problems," *J. Optim. Theory Appl.*, vol. 188, no. 1, pp. 1–25, Jan. 2021.

[41] F. Wen, L. Chu, P. Liu, and R. C. Qiu, "A survey on nonconvex regularization-based sparse and low-rank recovery in signal processing, statistics, and machine learning," *IEEE Access*, vol. 6, pp. 69883–69906, 2018.

[42] E. J. Candès and Y. Plan, "Matrix completion with noise," *Proc. IEEE*, vol. 98, no. 6, pp. 925–936, Jun. 2009.

[43] E. J. Candès, X. Li, Y. Ma, and J. Wright, "Robust principal component analysis?" *J. ACM*, vol. 58, no. 1, pp. 1–37, 2009.

[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. 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.

[46] 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.

[47] F. Wen, R. Ying, P. Liu, and R. C. Qiu, "Robust PCA using generalized nonconvex regularization," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 30, no. 6, pp. 1497–1510, Jun. 2020.

[48] C. Chen, B. He, Y. Ye, and X. Yuan, "The direct extension of ADMM for multi-block convex minimization problems is not necessarily convergent," *Math. Program.*, vol. 155, no. 1, pp. 57–79, 2014.

[49] R. I. Boţ and E. R. Csetnek, "ADMM for monotone operators: Convergence analysis and rates," *Adv. Comput. Math.*, vol. 45, no. 1, pp. 327–359, Feb. 2019.

[50] W. Dai, L. Zhang, J. Fu, T. Chai, and X. Ma, "Dual-rate adaptive optimal tracking control for dense medium separation process using neural networks," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 32, no. 9, pp. 4202–4216, Sep. 2021.

[51] W. Dai, T. Li, L. Zhang, Y. Jia, and H. Yan, "Multi-rate layered operational optimal control for large-scale industrial processes," *IEEE Trans. Ind. Informat.*, vol. 18, no. 7, pp. 4749–4761, Jul. 2022.

[52] Y. Jiao, Q. Jin, X. Lu, and W. Wang, "Alternating direction method of multipliers for linear inverse problems," *SIAM J. Numer. Anal.*, vol. 54, no. 4, pp. 2114–2137, Jan. 2016.

[53] 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.

[54] R. G. Bartle and D. R. Sherbert, *Introduction to Real Analysis*. New York, NY, USA: Wiley, 2000.

[55] 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.

[56] R. T. Rockafellar and R. J. Wets, *Variational Analysis*, vol. 317. Springer, 2009.

[57] K. Border. (2001). *The Supergradient of a Concave Function*. [Online]. Available: http://www.hss.caltech.edu/kcb/Notes/Supergrad.pdf
[58] C. Lu, H. Min, Z. Zhao, L. Zhu, D. Huang, and S. Yan, "Robust and efficient subspace segmentation via least squares regression," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2012, pp. 347–360.
[59] C. Lu, J. Feng, Z. Lin, T. Mei, and S. Yan, "Subspace clustering by block diagonal representation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 41, no. 2, pp. 487–501, Feb. 2019.

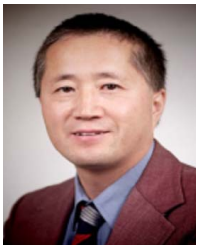
Hengmin Zhang received the B.S. degree from the School of Mathematics and Statistics, Heze University, Heze, China, in 2010, the M.S. degree from the College of Science, China University of Mining and Technology (CUMT), Xuzhou, China, in 2013, and the Ph.D. degree from the School of Computer Science and Engineering, Nanjing University of Science and Technology (NUST), Nanjing, China, in 2019.

He was a Post-Doctoral Fellow with the School of Information Science and Engineering, East China University of Science and Technology (ECUST), Shanghai, China, from 2019 to 2021. He is currently a Post-Doctoral Fellow with the PAMI Research Group, Department of Computer and Information Science, University of Macau (UM), Macau. His research interests include pattern recognition, nonconvex optimization algorithms, and large-scale representation learning methods.

Feng Qian (Senior Member, IEEE) received the Ph.D. degree in automation from the East China Institute of Chemical Technology, Shanghai, China, in 1995.

He was the Vice-President of the East China University of Science and Technology, Shanghai, where he was the Director of the Key Laboratory of Smart Manufacturing in Energy Chemical Process, Ministry of Education, and the Process System Engineering Research Center, Ministry of Education. His current research interests include modeling, control, optimization, and integration of petrochemical complex industrial processes and their industrial applications.

Dr. Qian is a member of the China Instrument and Control Society, the Chinese Association of Higher Education, China's PTA Industry Association, and the Chinese Academy of Engineering.



Peng Shi (Fellow, IEEE) received the Ph.D. degree in electrical engineering from The University of Newcastle, Callaghan, NSW, Australia, in 1994, the Ph.D. degree in mathematics from the University of South Australia, Adelaide, SA, Australia, in 1998, the D.Sc. degree from the University of Glamorgan, Pontypridd, U.K., in 2006, and the D.E. degree from The University of Adelaide, Adelaide, in 2015.

He is currently a Professor with The University of Adelaide. His research interests include systems, control theory and applications to autonomous and robotic systems, intelligence systems, network systems, and cyber–physical systems.

Dr. Shi is also a member of the Academy of Europe and a fellow of IET, IEAust, and CAA. He has been on the editorial board of many journals. He has been the President of the International Academy for Systems and Cybernetic Science and the Distinguished Lecturer of the IEEE SMC Society.

Wenli Du (Member, IEEE) received the B.Sc. and M.Sc. degrees in chemical process automation from the Dalian University of Technology, Dalian, China, in 1997 and 2000, respectively, and the Ph.D. degree in control theory and control engineering from the East China University of Science and Technology, Shanghai, China, in 2005.

She is currently the Dean of the Graduate School and the Vice-Dean of the Key Laboratory of Advanced Smart Manufacturing in Energy Chemical Process, Ministry of Education in China, Beijing, China. Her research interests include modeling, control, optimization, and fault diagnosis of chemical processes. The research results have been widely applied to industrial ethylene plants and PTA plants, which brings significant economic and social benefits.

Yang Tang (Senior Member, IEEE) received the B.S. and Ph.D. degrees in electrical engineering from Donghua University, Shanghai, China, in 2006 and 2010, respectively.

From 2008 to 2010, he was a Research Associate with The Hong Kong Polytechnic University, Hong Kong. From 2011 to 2015, he was a Post-Doctoral Researcher with the Humboldt University of Berlin, Berlin, Germany, and the Potsdam Institute for Climate Impact Research, Potsdam, Germany. He is currently a Professor with the East China University of Science and Technology, Shanghai. His current research interests include distributed estimation/control/optimization, cyber–physical systems, hybrid dynamical systems, computer vision, reinforcement learning, and their applications.

Prof. Tang was a recipient of the Alexander von Humboldt Fellowship. He has been a recipient of the ISI Highly Cited Researchers Award by Clarivate Analytics since 2017. He is also an associate editor of several IEEE TRANSACTIONS and IFAC journals.

Jianjun Qian (Member, IEEE) received the Ph.D. degree in pattern recognition and intelligence systems from the Nanjing University of Science and Technology (NUST), Nanjing, China, in 2014.

Since 2019, he has been a Post-Doctoral Researcher with the Institute of Textiles and Clothing (ITC), The Hong Kong Polytechnic University, Hong Kong. He is currently an Associate Processor 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.

Dr. Qian was selected as a Hong Kong Scholar in China in 2018.

Chen Gong (Member, IEEE) received the Ph.D. degree from the University of Technology Sydney, Sydney, NSW, Australia, in 2017.

He is currently a Professor with the Nanjing University of Science and Technology, Nanjing, China. He has published more than 100 technical papers in prominent journals and conferences. His research interests mainly include machine learning and data mining.

Dr. Gong won the Excellent Doctorial Dissertation Award of the Chinese Association for Artificial Intelligence, the Young Elite Scientists Sponsorship Program of the China Association for Science and Technology, the Wu Wen-Jun AI Excellent Youth Scholar Award, and the Science Fund for Distinguished Young Scholars of Jiangsu Province. He was also selected as the Global Top Chinese Young Scholars in AI released by Baidu.

Jian Yang (Member, IEEE) 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. From 2006 to 2007, he is a Chang-Jiang Professor with the School of Computer Science and Engineering, NUST. He is the author of more than 200 scientific papers in pattern recognition and computer vision. His papers have been cited over 30 000 times in Google Scholar. His research interests include pattern recognition, computer vision, and machine learning.

Dr. Yang is also a fellow of IAPR. He is/was an Associate Editor of *Pattern Recognition*, *Pattern Recognition Letters*, IEEE TRANSACTIONS ON NEURAL NETWORKS AND LEARNING SYSTEMS, and *Neurocomputing*.