LRR for Subspace Segmentation via Tractable Schatten-p Norm Minimization and Factorization

Hengmin Zhang, Jian Yang, *Member, IEEE*, Fanhua Shang, *Member, IEEE*, Chen Gong, *Member, IEEE*, and Zhenyu Zhang

***Abstract*—Recently, nuclear norm-based low rank representation (LRR) methods have been popular in several applications, such as subspace segmentation. However, there exist two limitations: one is that nuclear norm as the relaxation of rank function will lead to the suboptimal solution since nuclear norm-based minimization subproblem tends to the over-relaxations of singular value elements and treats each of them equally; the other is that solving LRR problems may cause more time consumption due to involving singular value decomposition of the large scale matrix at each iteration. To overcome both disadvantages, this paper mainly considers two tractable variants of LRR: one is Schatten-p norm minimization-based LRR (i.e., S_pNM_LRR) and the other is Schatten-p norm factorization-based LRR (i.e., S_pNF_LRR) for $p = 1$, 2/3 and 1/2. By introducing two or more auxiliary variables in the constraints, the alternating direction method of multiplier (ADMM) with multiple updating variables can be devised to solve these variants of LRR. Furthermore, both computational complexity and convergence property are given to evaluate nonconvex multiblocks ADMM algorithms. Several experiments finally validate the efficacy and efficiency of our methods on both synthetic data and real world data.**

***Index Terms*—Alternating direction method of multiplier (ADMM) algorithm, low rank representation (LRR), Schatten-p norm, singular value decomposition, subspace segmentation.**

Manuscript received September 25, 2017; revised January 11, 2018; accepted February 26, 2018. Date of publication March 14, 2018; date of current version March 5, 2019. This work was supported in part by the National Science Fund of China under Grant U1713208 and Grant 61472187, in part by the National Science Fund of Jiangsu Province under Grant BK20171430, in part by the "Six Talent Peak" Project of Jiangsu Province of China under Grant DZXX-027, in part by the 973 Program under Grant 2014CB349303, and in part by the Program for Changjiang Scholars and the Postgraduate Research and Practice Innovation Program of Jiangsu Province under Grant KYCX17_0359. This paper was recommended by Associate Editor S. Cruces. *(Corresponding author: Jian Yang.)*

H. Zhang and J. Yang are 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 210094, China (e-mail: zhanghengmin@126.com; csjyang@njust.edu.cn).

F. Shang is with the School of Artificial Intelligence, Xidian University, Xi'an 710071, China (e-mail: shangfanhua@hotmail.com).

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

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

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

Digital Object Identifier 10.1109/TCYB.2018.2811764

I. Introduction

IT IS well known that the problem of subspace segmentation aims to segment the observed data into multiple subspaces and find a low-dimensional subspace fitting each same subject. This basic strategy has widespread applications in computer vision and machine learning [1]–[4]. Generally speaking, the existing subspace segmentation methods are mainly divided into four classes, including iterative methods [5], [6], algebraic methods [7], [8], statistical methods [9], [10], and spectral clustering methods [11], [12]. Among them, spectral clustering methods are insensitive to the noises and outliers, thus they perform well over other clustering methods as verified in [11]–[13]. The most representative works [e.g., sparse subspace segmentation (SSC) and low rank representation (LRR)] conduct the clustering task to achieve a sparse or low-rank coefficients matrix and then construct a similarity graph. Given a set of sufficiently sampled data vectors $X = [X_1, \ldots, X_k] = [x_1, \ldots, x_n] \in \mathbb{R}^{d\times n}$ generated by a union of k subspaces $\{S_i\}$, then the problem of subspace segmentation can be defined as below.

Definition 1: Let X_i be a collection of n_i data vectors from the subspace S_i with $n = \sum_{i=1}^{k} n_i$. The goal of subspace segmentation is to cluster the given data as different groups according to the underlying subspaces they are drawn from.

It should be noted that the main difference between SSC and LRR methods is the constraints on the representation matrix. Moreover, they usually consider to optimize convex l_1-norm and nuclear norm-based minimization (NNM) problems, respectively. The main reason is due to that l_0-norm and rank function-based optimizations are general NP-hard and difficult to solve directly. Specially, LRR aims to take the correlation relationship of coefficient matrix into consideration, and it ends to find a low rank solution instead of a sparse solution like SSC. To achieve robust subspace recovery like [12], we mainly consider to solve the original LRR problem reformulated by

$$\min_{\{Z,E\}} \text{rank}(Z) + \lambda\|E\|_\ell, \ \text{s.t.}\ AZ + E = X \tag{1}$$

where $\lambda > 0$ is a balanced parameter, and rank(Z) counts the number of nonzero singular values of matrix Z, and $\|E\|_\ell$ represents a certain error measurement, such as the $L_{2,1}$-norm $\|\cdot\|_{2,1}$ used for characterizing the sample specific corruptions and outliers [12], the squared Frobenius norm $\|\cdot\|_F^2$ employed in handling with the Gaussian noise [14], and the L_1-norm $\|\cdot\|_1$ adopted for modeling the random or

sparse corruptions [15]. Moreover, LRR can not only reduce to robust principle components analysis (RPCA) [16], [17] when A is the identity matrix with $\|E\|_\ell = \|E\|_1$ but also reveal the true data structure [18] through learning a proper dictionary A. Theoretical studies [19], [20] have shown that nuclear norm is the tightest convex relaxation of a matrix rank function, which is similar to the relationship between l_1-norm and l_0-norm of a vector. This naturally motivates us to apply nuclear norm to relax the rank function. Fortunately, NNM problems can be solved by numerous of first-order algorithms, such as semi-definite programming [21], accelerated proximal gradient [22]–[24], and alternating direction method of multiplier (ADMM) [25]–[27]. Note that these algorithms usually achieve a globally convergence guarantees.

However, there exist two drawbacks which may limit the applications of nuclear norm-based methods in *efficacy* and *efficiency*: 1) nuclear norm as the convex relaxation of rank function can lead to the biased solution owing to its over-shrinks of rank components of a low rank matrix [14], [28]–[30] and 2) involving the singular value decomposition (SVD) of large scale matrix can lead to the high computational cost, thus not handle large scale problems. To better relax the rank function and mitigate the time consumption, the subsequent Section II states two popular strategies, including *matrix rank minimization methods* and *matrix factorization methods*. The first strategy can motivate us to use nonconvex rank relaxations instead of nuclear norm to obtain a better solution like [28] and [31]. By decomposing a large scale matrix into two much smaller factor matrices, the second strategy can result in lower computational complexity. Since Schatten-p norm is the popular rank relaxed function and has the decomposable formulations, and it also has a more accurate recovery ability for low rank matrix while requiring only a weaker restricted isometry property [32]. Besides, the rigorous theoretical results have verified that Schatten-p norm minimization with small p-values require significantly fewer measurements [33]. Unfortunately, there exists similar limitations with NNM methods for nonconvex Schatten-p norm ones since involving SVD of large scale matrix at each iteration. To improve the *efficacy* and *efficiency*, this paper focuses on presenting tractable Schatten-p norm minimization and factorization approaches for the rank relaxations in (1). The main contributions are summarized as follows.

First, different from convex nuclear norm-based LRR model, we use nonconvex Schatten-p norm and its bounding functions (see Table II) as the rank relaxations. By choosing $p = 1, 2/3$ and $1/2$, we give two tractable LRR models: one is S_pNM_LRR for increasing *efficacy*, while the other is S_pNF_LRR for enhancing *efficiency* though a fast and accurate algorithm should be the interest of researches.

Second, we propose nonconvex multiblocks ADMMs with at least two dual variables to solve our LRR variants via Schatten-p norm. Although such models are both nonconvex and nonsmooth, we establish the convergence guarantees and show their computational complexities (see Table III).

Third, several experiments on both synthesized data and real world data can verify the superiority of the proposed algorithms (see Algorithms 1 and 2) from both clustering accuracy and time consumption (see Tables IV–VII), respectively.

TABLE I
SCHATTEN-p FUNCTIONS WITH $p = 0, 1/2, 2/3, 1$ IN (2) AND THEIR CORRESPONDING THRESHOLDING OPERATORS IN (3)

p		$\mathcal{S}_{\beta,p}(\sigma_i(Y)) = \theta(\sigma_i)\cdot\xi$
0	rank(Z)	$\begin{cases} 0, & \sigma_i < \sqrt{2\beta}, \\ \{0, \mathrm{sgn}(\sigma_i)\sqrt{2\beta}\}, & \sigma_i = \sqrt{2\beta}, \\ \sigma_i, & \sigma_i > \sqrt{2\beta}. \end{cases}$
1/2	$\|Z\|^{1/2}_{S_{1/2}}$ [34]	$\theta(\sigma_i) = \frac{2}{3}\sigma_i(1+\cos(\frac{2\pi}{3}-\frac{2}{3}\varphi))$, $\varphi = \arccos(\frac{\beta}{4}(\frac{\sigma_i}{3})^{-1.5})$, $\xi = \begin{cases} 1, & \sigma_i > \frac{\sqrt[3]{54}}{4}(2\eta)^{2/3}, \\ 0, & \text{otherwise.} \end{cases}$
2/3	$\|Z\|^{2/3}_{S_{2/3}}$ [34]	$\theta(\sigma_i) = ((\varpi + \sqrt{2\sigma_i/\varpi - \varpi^2})/2)^3$, $\varpi = \frac{2}{\sqrt{3}}(2\beta)^{1/4}\cosh(\phi/3)^{1/2}$, $\phi = \mathrm{arccosh}(\frac{27\sigma_i^2}{16}(2\beta)^{-1.5})$, $\xi = \begin{cases} 1, & \sigma_i > \frac{2}{3}(3(2\beta)^3)^{1/4}, \\ 0, & \text{otherwise.} \end{cases}$
1	$\|Z\|_*$ [35]	$\max\{\sigma(Y) - \beta, \mathbf{0}\}$.

TABLE II
USED CONVEX AND NONCONVEX SCHATTEN-p NORM AND THEIR BOUNDING FUNCTIONS $\psi_p(U, V)$ WITH DIFFERENT p-VALUES CHOICE

p-value	$\|Z\|^p_{S_p}$	$\min_{Z=UV^T}\psi_p(U,V)$
$p = 1$	$\|Z\|_*$ [36]	$\min_{Z=UV^T}\frac{1}{2}(\|U\|_F^2 + \|V\|_F^2)$
$p = 2/3$	$\|Z\|^{2/3}_{\mathrm{F/N}}$ [37]	$\min_{Z=UV^T}\frac{1}{3}(2\|U\|_* + \|V\|_F^2)$
$p = 1/2$	$\|Z\|^{1/2}_{\mathrm{BiN}}$ [37]	$\min_{Z=UV^T}\frac{1}{2}(\|U\|_* + \|V\|_*)$

Outline: Section II presents the most related works. Borrowing from Tables I and II, Section III proposes two tractable LRR variants. Two efficient nonconvex multiblocks ADMMs are further provided by introducing multiple auxiliary variables in Section IV. Subsequently, we give the detailed analysis including computational complexity and convergence behavior in Section V. Extensive experiments can show the superiority of our proposed algorithms in Section VI. Finally, we state the conclusions and future works in Section VII.

II. RELATED WORKS

This section first presents two popular strategies for rank relaxations and then introduces Schatten-p norm and its decomposable formulations and thresholding operators.

A. Matrix Rank Minimization Methods

Besides nuclear norm, several nonconvex rank relaxations (e.g., Schatten-p ($0 < p < 1$) norm [34], [38]–[42], weighted nuclear norm [43], and truncated nuclear norm [44]) have been widely applied to substitute the rank function. Additionally, some nonconvex surrogates of l_0-norm (e.g., SCAD [28], MCP [31], and logarithm [45]) listed in [46]–[48] have been successfully extended to the rank relaxations. These nonconvex functions usually perform better than convex nuclear norm since they overcome the imbalanced penalization of different singular values.

B. Matrix Factorization Methods

Matrix factorization is the other popular method for rank relaxations through two ways: one is factorizing the nuclear norm of large scale matrix into the product of two or three small scale matrices [49]–[51], the other is minimizing the sum of squared Frobenius norm of two smaller matrices [36]. In doing so, the computational complexity can be reduced by computing small scale matrices. Note that the aforementioned are mainly based on the approximations of nuclear norm. Especially, nonconvex Schatten-p norm (e.g., $p = 2/3$ and $1/2$) based matrix factorization methods [37], [52]–[54] have been recently proposed to solve matrix completion and RPCA problems, respectively.

Consider that the unbiased property and the decomposable formulation of nonconvex Schatten-p norm, this paper focuses on employing them for rank relaxations of (1). Now, let us define the Schatten-p norm of a matrix $Z \in \mathbb{R}^{m\times n}$ by using the l_p-norm of its singular values as follows:

$$\|Z\|_{S_p} \triangleq \left(\sum_{i=1}^{\min(m,n)} \sigma_i^p(Z)\right)^{1/p}, \quad 0 \le p < +\infty \tag{2}$$

where $Z = U\text{Diag}(\sigma(Z))V^T$ is the economical SVD with two orthogonal matrices $U \in \mathbb{R}^{m\times n}$ and $V \in \mathbb{R}^{n\times n}$ when $n = \min(m, n)$, $\sigma_i(Z)$ is the ith entry of singular values vector denoted by $\sigma(Z) = (\sigma_1(Z), \sigma_2(Z), \ldots, \sigma_{\min(m,n)}(Z))^T$. It follows from (2) that the gap between rank function (i.e., $p = 0$) and nuclear norm (i.e., $p = 1$) can be bridged by setting $0 < p < 1$ [34], [39], [55]. Most especially, when the rank number is relatively larger, nonconvex Schatten-p norm can show its superiority over nuclear norm for relaxing a matrix rank function. Otherwise, the superiority of nonconvex Schatten-p norm over the convex one will become slight or not. To solve S_pNM_LRR problem [34], [42], we do introduce the generalized thresholding operator [47], denoted by

$$\text{Prox}_{\beta,p}(Y) = \text{argmin}_Z \beta\|Z\|_{S_p}^p + \frac{1}{2}\|Z - Y\|_F^2 \tag{3}$$

where $\beta > 0$, $\text{Prox}_{\beta,p}(Y) = U_Y\text{Diag}(\mathcal{S}_{\beta,p}(\sigma(Y)))V_Y^T$ with $Y = U_Y\text{Diag}(\sigma(Y))V_Y^T$. It follows from [34], [35], and [55] that we can obtain the closed-form solutions of (3) for $p = 1$, $2/3$ and $1/2$, respectively. Note that Table I lists Schatten-p norm and its corresponding thresholding operators, and Table II gives the bounding functions of Schatten-p norm, and Fig. 1 indicates the superiority of nonconvex Schatten-p norm with $p = 2/3$, $1/2$ in shrinking singular values over nuclear norm. Different from [56]–[58], which consider convex L2-Graph, F-norm, and nuclear norm-based methods for subspace clustering, we know that nonconvex Schatten-p norm-based methods usually show their superiority for the rank minimization problems. To our best knowledge, all the solutions of (3) for $0 < p < 1$ can be achieved by generalized iterated shrinkage algorithm [42], [59], iteratively reweighted nuclear norm algorithm [46], fixed-point iteration algorithm [47], and generalized iterated matrix/vector soft thresholding algorithms [39], [41], respectively.

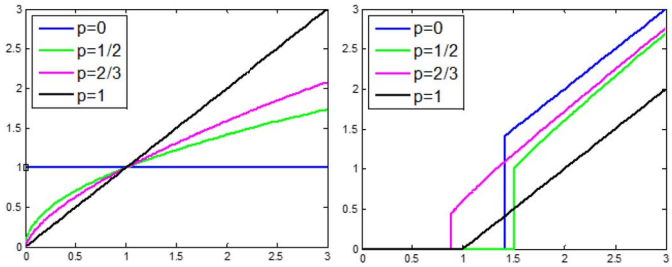

Fig. 1. Plots of the (a) function l_p for various values of $p = 0, 1/2, 2/3, 1$ and (b) their corresponding thresholding operators of (3). Here, $\beta = 1$ and singular values $\sigma(Y) = [0, 3]$ for the Schatten-p norm.

III. Problem Formulations

To better capture the global structure of coefficients matrix, nonconvex Schatten-p norm instead of nuclear norm is used to substitute the rank function of (1) for showing the superiority of representation ability [34], [39], [41], [42]. This can motivate us to keep each subproblem having closed-form solution, and then establish more general S_pNM_LRR model with two constraints as follows:

$$\min_{\{J,Z,E\}} \|J\|_{S_p}^p + \lambda\|E\|_\ell, \quad \text{s.t.} \quad AZ + E = X, J = Z \tag{4}$$

where $\|E\|_\ell$ can measure different styles of noise, and LRR [12] can be regarded as the special case of model (4) when $p = 1$. We here emphasize on two nonconvex cases of model (4) with $p = 2/3, 1/2$ (i.e., Schatten-2/3 and $1/2$ norm) since they can obtain competitive performance against the convex case for $p = 1$ as verified in [42]. However, different from the formulation of model (4), the most related S_pNM clustering method [42] aims to solve the problem with one constraint

$$\min_{\{Z,E\}} \|Z\|_{S_p}^p + \lambda\|E\|_\ell, \quad \text{s.t.} \quad AZ + E = X \tag{5}$$

where we stress out that solving both (4) and (5) usually involves the SVD of large scale matrix at each iteration. Naturally, it will suffer from higher time consumption in order to limit their applicability for large scale optimizations. Inspired by the matrix factorization-based strategies for nuclear norm [36], Schatten-$1/2$ norm and $2/3$ norm [37], [52]–[54] (i.e., bounding them by the Bi-Forbenius norm, Bi-nuclear norm $\|\cdot\|_{\text{BiN}}$, and Forbenius/Nuclear hybrid norm $\|\cdot\|_{F/N}$, respectively), we can give the corresponding alterative formulations listed in Table II. For any coefficients matrix $Z \in \mathbb{R}^{m\times n}$ with $\text{rank}(Z) = r \le d$ [i.e., the upper bounding of the rank(Z)], we can factorize Z into two small scale matrices $U \in \mathbb{R}^{m\times d}$ and $V \in \mathbb{R}^{n\times d}$ such that $Z = UV^T$. Equation (5) can be transformed into the following S_pNF_LRR model:

$$\min_{\{U,V,E\}} \psi_p(U, V) + \lambda\|E\|_\ell, \quad \text{s.t.} \quad AUV^T + E = X \tag{6}$$

where $\psi_p(U, V)$ may be represented for more different p-values according to [53]. However, we here only consider the cases of $p = 1, 2/3$ and $1/2$, respectively.

Compared with (4) and (5), (6) aims to reduce the computational complexity by computing the operators for the small

scale matrices. Furthermore, we can devise the algorithm to show the *efficiency* of S_pNF_LRR model for subspace segmentation. To solve (6) with different $\psi_p(U, V)$ and guarantee that each subproblem has the analytic solution, we need to employ the judicious variable partition technique, which has been used in [12], [54], and [60] due that multiple variables are involved in the objective function. These can motivate us to obtain the special formulations with at least two constraints, denoted by

$$\begin{cases} S_1NF\_LRR: \ \min_{\{U,V,Z,E\}} \frac{1}{2}\left(\|U\|_F^2 + \|V\|_F^2\right) + \lambda\|E\|_\ell \\ \qquad \text{s.t.} \ \ AZ + E = X, \quad UV^T = Z \\ S_{\frac{2}{3}}NF\_LRR: \ \min_{\{U,V,M,Z,E\}} \frac{1}{3}\left(2\|M\|_* + \|V\|_F^2\right) + \lambda\|E\|_\ell \\ \qquad \text{s.t.} \ \ AZ + E = X, \quad UV^T = Z, \quad M = U \\ S_{\frac{1}{2}}NF\_LRR: \ \min_{\{U,V,M,N,Z,E\}} \frac{1}{2}(\|M\|_* + \|N\|_*) + \lambda\|E\|_\ell \\ \qquad \text{s.t.} \ \ AZ + E = X, \quad UV^T = Z, \quad M = U, \quad N = V. \end{cases}$$

This paper will evaluates the subspace segmentation ability of both S_pNM_LRR and S_pNF_LRR models, which are solved by nonconvex ADMM algorithms with multiblock variables. Using the minimization and factorization strategies, the final clustering accuracy can show their efficacy and efficiency, respectively. Note that no additional constraints are restricted on the representation matrix and the noise structure like [12] and [42], thus the results may not be state-of-the-art. Besides, one usually choose the original data (i.e., $A = X$) instead of using XP^*, where P^* can be obtained by orthogonalizing the columns of X^T as the dictionary although the latter can save time consumption, which has been validated in [12]. If we choose $A = X$, low rank matrix Z_* obtained by Algorithms 1 and 2 can be used to construct an affinity matrix W,[1] i.e., $(|Z_*| + |Z_*^T|)/2$, which is both symmetric and entry-wise nonnegative. The spectral clustering method in [64] can be used to obtain the clustering accuracy for showing the efficacy. Meanwhile, the total time computations are also obtained for showing the efficiency.

IV. Optimization Algorithms

To solve S_pNM_LRR (4) and S_pNF_LRR (6), respectively, this section will devises the solutions according to the popular ADMM algorithm [25], [26]. Note that the used optimization algorithms are nonconvex and multiblocks. The basic iterations of ADMM are first to give the augmented Lagrangian function, and then focus on updating primal variables, dual variables, and penalty parameter at each iteration in an alternating way.

A. S_pNM_LRR via ADMM Algorithm

By introducing two dual variables Y_1 and Y_2, we give the augmented Lagrangian function of (4) as follows:

$$\begin{aligned} &\mathcal{L}_\mu(J, Z, E, Y_1, Y_2) \\ &\quad = \|J\|_{S_p}^p + \lambda\|E\|_\ell + \langle Y_1, J - Z\rangle + \langle Y_2, AZ + E - X\rangle \\ &\qquad + \frac{\mu}{2}\left(\|J - Z\|_F^2 + \|AZ + E - X\|_F^2\right) \end{aligned} \tag{7}$$

[1]Note that different constructions of affinity matrix have various effects on the clustering accuracy, which have been verified in [11], [12], and [61]–[63].

Algorithm 1 S_pNM_LRR Optimized by ADMM

Input: X, A, $p = 1, 2/3$ and $1/2$, $\lambda > 0$, $\text{InT} = (I + A^TA)^{-1}$.
Initialization: $k = 0$, $\rho > 1$, $\mu_0 = 10^{-6}$, $Z_0 = \mathbf{0}$, $E_0 = \mathbf{0}$, $Y_{1,0} = \mathbf{0}$, $Y_{2,0} = \mathbf{0}$.

1. **while** not converged **do**
2. Update J_{k+1} by solving
$$J_{k+1} = \text{argmin}_J \|J\|_{S_p}^p + \frac{\mu_k}{2}\left\|J - \left(Z_k - \frac{Y_{1,k}}{\mu_k}\right)\right\|_F^2;$$
3. Update Z_{k+1} by solving
$$Z_{k+1} = \text{InT} \times \left(J_{k+1} + \frac{Y_{1,k}}{\mu_k} - A^T\left(E_k - X + \frac{Y_{2,k}}{\mu_k}\right)\right);$$
4. Update E_{k+1} by computing
$$E_{k+1} = \text{argmin}_E \lambda\|E\|_\ell + \frac{\mu_k}{2}\left\|E - \left(X - AZ_{k+1} - \frac{Y_{2,k}}{\mu_k}\right)\right\|_F^2;$$
5. Update the multipliers $Y_{1,k+1}$ and $Y_{2,k+1}$ by
$$\begin{cases} Y_{1,k+1} = Y_{1,k} + \mu_k(J_{k+1} - Z_{k+1}); \\ Y_{2,k+1} = Y_{2,k} + \mu_k(AZ_{k+1} + E_{k+1} - X); \end{cases}$$
6. Update μ_{k+1} by $\mu_{k+1} = \rho\mu_k$;
7. **end while**

Output: Optimal Representation Matrix $Z_* \leftarrow Z_{k+1}$.

where the penalty parameter $\mu > 0$, the iteration steps for solving (7) are performed by updating one of the primal variables $\{J, Z, E\}$ in sequence while fixed others, and then updating one of the dual and penalty variables $\{Y_1, Y_2, \mu\}$ in sequence while fixed others. Actually, the dual variable is obtained by gradient ascent on the resulting dual problem. The optimization scheme is outlined in Algorithm 1. Note that the symbol "$\times$" represents the multiplier of two variables in the whole paper, updating J_{k+1}, Z_{k+1}, and E_{k+1} are very key steps, and they all have closed-form solutions especially for various choices of $\|J\|_{S_p}^p$ with $p = 1, 2/3$ and $1/2$ [34], [35] and $\|E\|_\ell$ with $\ell = l_2, l_1$ and $l_{2,1}$. Actually, there exist several Schatten-p norm-based minimization problems and optimizations [39], [42], [46], [63]. However, some of them may not have the closed-form solutions [42] and several relaxed strategies can be used to avoid the nonsmooth property [63]. We will give several differences between S_pNM_LRR model (4) and (5) of [42], which are listed as follows.

1) Comparison of Models and Algorithms: To obtain low rank solutions [42, eq. (5)], the additional auxiliary variable (i.e., J) is introduced. Specifically, two dual variables ADMM algorithm is used for model (4), while the linearized ADMM with adaptive penalty (LADMMAP) is used in [42]. Note that it is easy to guarantee that each subproblem has the closed-form solution, while [42] cannot do since the involved subproblem of Schatten-p norm optimization cannot be easily solvable when the data matrix A is not the identity matrix.

2) Comparison of Subproblems: To solve the subproblem of Schatten-p norm efficiently, we know that [42] needs to linearize the quadratic penalty term and add a proximal term, while Algorithm 1 does not. Moreover, both solutions are not the same for computing Schatten-p norm minimization due that [42] mainly considers to use the generalized iterated matrix soft thresholding algorithm for any $0 < p < 1$

through [59], while this paper gives the closed-form solutions only when $p = 1, 2/3$ and $1/2$ according to [34] and [55].

3) Comparison of Convergence Analysis: To our best knowledge, [42] first states that the nonconvex iteration thresholding step can converge to a stationary point for any thresholding function when it satisfies three specific properties. Then together with the updating rules for the convex step and the other variables steps, it can conclude the convergence property of the LADMMAP algorithm. However, the limit point of the variable sequence generated by our algorithm can be proved to satisfy the first-order Karush–Kuhn–Tucker (KKT) condition using several basic assumptions and key lemmas.

B. S_pNF_LRR via ADMM Algorithm

Similar to Algorithm 1 for solving the S_pNM_LRR model and guaranteeing that each subproblem can be easily solvable, one needs the augmented Lagrangian functions of (6) via different $\psi_p(U, V)$ [53] as follows:

$$\begin{aligned}&\mathcal{L}_\mu(U, V, Z, E, Y_1, Y_2)\\&= \frac{1}{2}\Big(\|U\|_F^2 + \|V\|_F^2\Big) + \lambda\|E\|_\ell + \langle Y_1, AZ + E - X\rangle\\&\quad + \big\langle Y_2, UV^T - Z\big\rangle + \frac{\mu}{2}\Big(\big\|UV^T - Z\big\|_F^2 + \|AZ + E - X\|_F^2\Big)\end{aligned} \tag{8}$$

$$\begin{aligned}&\mathcal{L}_\mu(U, V, M, Z, E, Y_1, Y_2, Y_3)\\&= \frac{1}{3}\Big(2\|M\|_* + \|V\|_F^2\Big) + \lambda\|E\|_\ell + \langle Y_1, AZ + E - X\rangle\\&\quad + \big\langle Y_2, UV^T - Z\big\rangle + \langle Y_3, M - U\rangle\\&\quad + \frac{\mu}{2}\Big(\|M - U\|_F^2 + \big\|UV^T - Z\big\|_F^2 + \|AZ + E - X\|_F^2\Big)\end{aligned} \tag{9}$$

$$\begin{aligned}&\mathcal{L}_\mu(U, V, M, N, Z, E, Y_1, Y_2, Y_3)\\&= \frac{1}{2}(\|M\|_* + \|N\|_*) + \lambda\|E\|_\ell + \langle Y_1, AZ + E - X\rangle\\&\quad + \big\langle Y_2, UV^T - Z\big\rangle + \langle Y_3, M - U\rangle + \langle Y_4, N - V\rangle\\&\quad + \frac{\mu}{2}\Big(\|M - U\|_F^2 + \|N - V\|_F^2 + \big\|UV^T - Z\big\|_F^2\\&\qquad + \|AZ + E - X\|_F^2\Big).\end{aligned} \tag{10}$$

Note that (8)–(10) are the corresponding augmented Lagrangian functions of S_pNF_LRR (6) with $p = 1$, $2/3$ and $1/2$, which are much more complicated than the function (7) by respectively replacing Z with UV^T and $\|Z\|_{S_p}^p$ with different substitutes of $\psi_p(U, V)$ and introducing more constraints. Moreover, two factor matrices can be extended to three and more factor matrices for any $0 < p < 1$ according to [53]. The detailed optimization schemes are outlined in Algorithm 2, and the multiplier variables are updated for $\{Y_{1,k+1}, Y_{2,k+1}\}$ with $p = 1$, $\{Y_{1,k+1}, Y_{2,k+1}, Y_{3,k+1}\}$ with $p = 2/3$ and $\{Y_{1,k+1}, Y_{2,k+1}, Y_{3,k+1}, Y_{4,k+1}\}$ with $p = 1/2$. Each involved variables will be updated in sequence by fixing others. Obviously, to solve (8)–(10) efficiently, we need to iteratively update more variables than (7) at each iteration until deducing the whole iterative scheme.

Although Schatten-p norm has been used to substitute the rank function in our models, other nonconvex rank relaxations can also be utilized here. These nonconvex rank relaxations usually perform better than convex nuclear norm, but it is still not clear to claim that which one is the best surrogate choice without additional requirements due that the solutions are the local optimal instead of the global optimal. Specially, fast solution via factorization for LRR is the first work for Schatten-p norm, but they also have other applications in image classification, matrix completion, RPCA, and LRR. Detailed statements can be found in these related references. Apply our methods to the clustering task, the following experiments can demonstrate two conclusions: one is that the proposed Algorithm 1 is effective but slow, while the proposed Algorithm 2 is efficient but of lower performance.

It follows from Algorithms 1 and 2 that both S_pNM_LRR and S_pNF_LRR can be optimized by the variants of ADMM algorithm, which have been used to solve nonconvex and multiblocks problems in [39] and [65]–[67]. However, they may slow down the convergence speed or even lead to the divergence when there are too many variables. Due to the lack of a convergence guarantee, it is necessary for us to propose a reliable algorithm with a convergence guarantee. Different from several existing LADMM, LADMMAP, and LADMMAP(A) optimization algorithms in [27], [42], and [68], they need more iterations to converge so that their convergence rates are sublinear [63]. The reason may be that they do not introduce other additional variables while using the linearized strategy to approximate the quadratic term, which will lead to an inexact solution. However, this paper focuses on nonconvex multiblocks ADMM algorithms by introducing two or more auxiliary variables in the constraints so that each subproblem of Algorithms 1 and 2 can obtain the closed-form solution.

V. Analysis of Proposed Algorithms

Both the computational complexity and convergence property are two basic evaluation criteria to review an optimization algorithm. Next, we will give the detailed analysis of our ADMM algorithms, respectively. Note that the total computational complexity for these iterative algorithms mainly depends on the total number of iterations and the computational cost at each iteration. Moreover, the convergence guarantees of ADMM algorithms for solving nonconvex multiblocks problems are very challenging as stated in [66], [67], and [69].

A. Complexity Analysis

Given a $p \times q$ matrix, the computational complexity of its SVD is $O(\min(pq^2, p^2q))$ at a time, the multiplication for $p \times q$ matrix and $q \times c$ matrix is $O(pqc)$, and the inverse for $r \times r$ matrix is $O(r^3)$. The computations of our algorithms mainly depend on the aforementioned three cases. We here suppose that $A = X$ and set $X, E \in \mathbb{R}^{m\times n}$ and $Z \in \mathbb{R}^{n\times n}$ in Algorithm 1, and $U, V \in \mathbb{R}^{n\times d}$ with compatible variables in Algorithm 2. These can help us obtain the detailed computational complexity of each step in Algorithms 1 and 2, which are listed in Table III, so it is easy to obtain the total complexity of both proposed algorithms at each iteration by

TABLE III
COMPUTATIONAL COMPLEXITY OF SEVERAL UPDATING STEPS IN ALGORITHMS 1 AND 2

Alg./Steps	2	3	4	5	6	7
1	$O(n^3)$	$O(mn^2+n^3)$	$O(mn^2)$	/	/	/
2 (1)	$O(2nd^2+dn^2+d^3)$	$O(2nd^2+dn^2+d^3)$	$O(mn^2+dn^2+n^3)$	$O(mn^2)$	/	/
2(2/3)	$O(2nd^2+dn^2+d^3)$	$O(2nd^2+dn^2+d^3)$	$O(nd^2)$	$O(mn^2+dn^2+n^3)$	$O(mn^2)$	/
2(1/2)	$O(2nd^2+dn^2+d^3)$	$O(2nd^2+dn^2+d^3)$	$O(nd^2)$	$O(nd^2)$	$O(mn^2+dn^2+n^3)$	$O(mn^2)$

Algorithm 2 S_pNF_LRR Optimized by ADMM

Input: $X, A, p, \lambda > 0$, d and $\text{InT} = (I + A^TA)^{-1}$.
Initialization: $k = 0$, $\rho > 1$, $\mu_0 = 10^{-6}$, $V_0 = \mathbf{0}$, $Z_0 = \mathbf{0}$, $M_0 = \mathbf{0}$, $N_0 = \mathbf{0}$, $E_0 = \mathbf{0}$, $Y_{1,0} = \mathbf{0}$, $Y_{2,0} = \mathbf{0}$, $Y_{3,0} = \mathbf{0}$, $Y_{4,0} = \mathbf{0}$.

1. **while** not converged **do**
2. if $p = 1$, update $\{U_{k+1}, V_{k+1}\}$ by solving

$$\begin{cases} U_{k+1} = \text{argmin}_U \frac{1}{2}\|U\|_F^2 + \frac{\mu_k}{2}\left\|UV_k^T - Q_k\right\|_F^2; \\ V_{k+1} = \text{argmin}_V \frac{1}{2}\|V\|_F^2 + \frac{\mu_k}{2}\left\|U_{k+1}V^T - Q_k\right\|_F^2; \end{cases}$$

if $p = 2/3$, update $\{U_{k+1}, V_{k+1}, M_{k+1}\}$ by solving

$$\begin{cases} U_{k+1} = \text{argmin}_U \left\|M_k - U + \frac{Y_{3,k}}{\mu_k}\right\|_F^2 + \left\|UV_k^T - Q_k\right\|_F^2; \\ V_{k+1} = \text{argmin}_V \frac{1}{3}\|V\|_F^2 + \frac{\mu_k}{2}\left\|U_{k+1}V^T - Q_k\right\|_F^2; \\ M_{k+1} = \text{argmin}_M \frac{2}{3}\|M\|_* + \frac{\mu_k}{2}\left\|M - U_{k+1} + \frac{Y_{3,k}}{\mu_k}\right\|_F^2; \end{cases}$$

if $p = 1/2$, update $\{U_{k+1}, V_{k+1}, M_{k+1}, N_{k+1}\}$ by solving

$$\begin{cases} U_{k+1} = \text{argmin}_U \left\|M_k - U + \frac{Y_{3,k}}{\mu_k}\right\|_F^2 + \left\|UV_k^T - Q_k\right\|_F^2; \\ V_{k+1} = \text{argmin}_V \left\|N_k - V + \frac{Y_{4,k}}{\mu_k}\right\|_F^2 + \left\|U_{k+1}V^T - Q_k\right\|_F^2; \\ M_{k+1} = \text{argmin}_M \frac{1}{2}\|M\|_* + \frac{\mu_k}{2}\left\|M - U_{k+1} + \frac{Y_{3,k}}{\mu_k}\right\|_F^2; \\ N_{k+1} = \text{argmin}_N \frac{1}{2}\|N\|_* + \frac{\mu_k}{2}\left\|N - V_{k+1} + \frac{Y_{4,k}}{\mu_k}\right\|_F^2; \end{cases}$$

where $Q_k = Z_k - \frac{Y_{2,k}}{\mu_k}$ for $p = 1, 2/3$ and $1/2$, respectively.

3. Update $\{Z_{k+1}, E_{k+1}\}$ by solving

$$\begin{cases} Z_{k+1} = \text{InT} \times \left(U_{k+1}V_{k+1}^T + \frac{Y_{2,k}}{\mu_k} - A^T\left(E_k - X + \frac{Y_{1,k}}{\mu_k}\right)\right); \\ E_{k+1} = \text{argmin}_E \lambda\|E\|_\ell + \frac{\mu_k}{2}\left\|E - \left(X - AZ_{k+1} - \frac{Y_{1,k}}{\mu_k}\right)\right\|_F^2; \end{cases}$$

4. Update the multipliers by

$$\begin{cases} Y_{1,k+1} = Y_{1,k} + \mu_k(AZ_{k+1} + E_{k+1} - X); \\ Y_{2,k+1} = Y_{2,k} + \mu_k(U_{k+1}V_{k+1}^T - Z_{k+1}); \\ Y_{3,k+1} = Y_{3,k} + \mu_k(M_{k+1} - U_{k+1}); \\ Y_{4,k+1} = Y_{4,k} + \mu_k(N_{k+1} - V_{k+1}); \end{cases}$$

5. Update μ_{k+1} by $\mu_{k+1} = \rho\mu_k$;
6. **end while**

Output: Optimal Representation Matrix $Z_* \leftarrow Z_{k+1}$.

summing all the steps, respectively. If each algorithm needs k iterations to converge, the entire complexity of the proposed algorithms are calculated by $k \times \sharp$, where $\sharp$ represents the sum of complexity from all the updating steps. Note that when $d \ll \min\{m, n\}$, we conclude that Algorithm 2 can efficiently deal with the large scale problems compared with both Algorithm 1 and LADMMAP algorithm [42] since the decomposable strategy can largely reduce the computational complexity through operating small scale matrices. Thus, the total computational complexities of both proposed algorithms at each iteration are ranked by Algorithm 1$\gg$Algorithm 2(1/2)$>$ Algorithm 2(2/3)$>$Algorithm 2(1) in a decreasing order according to the computations in Table III.

B. Convergence Analysis

It is well known that ADMM algorithms are usually adapted to solve the problems with a convex and nonsmooth objective function with two variables and structured linear constraints. However, it is not easy to guarantee the convergence of ADMM algorithm for solving nonconvex multiblock problems in general. The following will show that the sequence generated by Algorithm 1 converges to a KKT point under some milder conditions. Since variables are nonseparable and multiblocks in the constraints of S_pNF_LRR model, there is no yet established convergence proof of Algorithm 2 for solving (8)–(10) as discussed in [70]. Compared to Algorithm 1, the theoretical convergence analysis of Algorithm 2 becomes more challenging. Moreover, there has no theoretical evidence for the convergence of ADMM algorithm for the S_pNF_LRR. Nevertheless, nonconvex ADMM variants have been verified to achieve good numerical performance in subspace clustering with a similar matrix factorization strategy. We leave the convergence analysis of Algorithm 2 to be a future research. Note that the converged solutions of nonconvex problems may be different due to the sensitivity of the initialization. However, we empirically give the initial variables through [12] and [42], and then present some basic preliminaries.

Lemma 1 [26]: Let X be a real Hilbert space endowed with an inner product $\langle\cdot,\cdot\rangle$, a norm $\|\cdot\|$ with its the dual norm $\|\cdot\|^{\text{dual}}$, and $y \in \partial\|x\|$, where $\partial f(\cdot)$ is the subgradient of $f(\cdot)$. Then, we have $\|y\|^{\text{dual}} = 1$ if $x \neq 0$, and $\|y\|^{\text{dual}} \leq 1$ if $x = 0$.

Definition 2 [71]: For any permutation π for $\{1, 2, \ldots, r\}$, if $f(x_1, x_2, \ldots, x_r) = f(|x_{\pi(1)}|, |x_{\pi(2)}|, \ldots, |x_{\pi(r)}|)$ holds, we will call $f(\cdot) : \mathbb{R}^r \to \mathbb{R}$ an absolutely symmetric function. And the matrix function $f \circ \sigma$ is defined by

$$[f \circ \sigma](X) \coloneqq f(\sigma_1(X), \sigma_1(X), \ldots, \sigma_r(X)). \tag{11}$$

Lemma 2 [71]: Let $F(\cdot) : \mathbb{R}^{m\times n} \to \mathbb{R}$ be denoted by $F(X) = f \circ \sigma(X)$ and the function $f(\cdot) : \mathbb{R}^n \to \mathbb{R}$ be absolutely symmetric and differentiable at $\sigma(X)$. And the SVD of matrix X is $X = U\text{diag}(\sigma(X))V^T$, then the subdifferential of $F(X)$ (i.e., $f \circ \sigma$) at a matrix X is given by

$$\frac{\partial F(X)}{\partial X} = \partial(f \circ \sigma)(X) = U\text{diag}(\partial f[\sigma(X)])V^T \tag{12}$$

where $(\partial f(\sigma_i(X))/\partial X) = c_i f'[\sigma_i(X)]$ with $c_i \in \partial|\sigma_i(X)|$.

Next, we prove the boundedness of variable sequences and then obtain the convergence guarantee of Algorithm 1. The detailed proof can be found in the supplementary material.

Theorem 1: Assume that the dictionary A is normalized, let $\{\mathcal{T}_k = (J_k, Z_k, E_k; Y_{1,k}, Y_{2,k})\}_{k=1}^{\infty}$ be the sequence generated

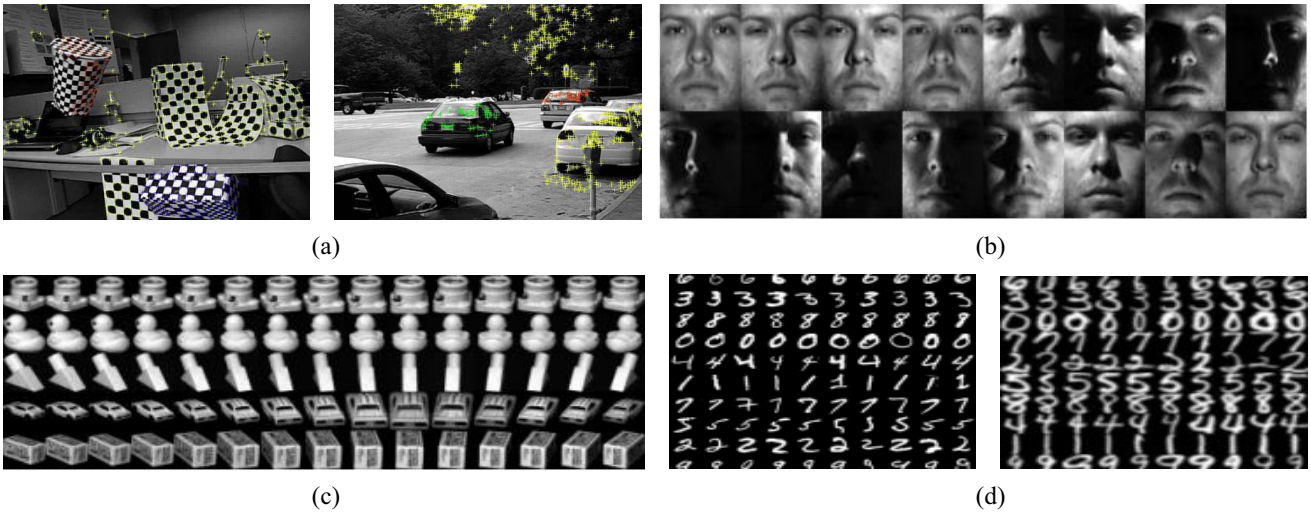

Fig. 2. Some sample images from the real-world databases. (a) Hopkins 155. (b) Extended YaleB. (c) COIL20. (d) MNIST (left) and USPS (right).

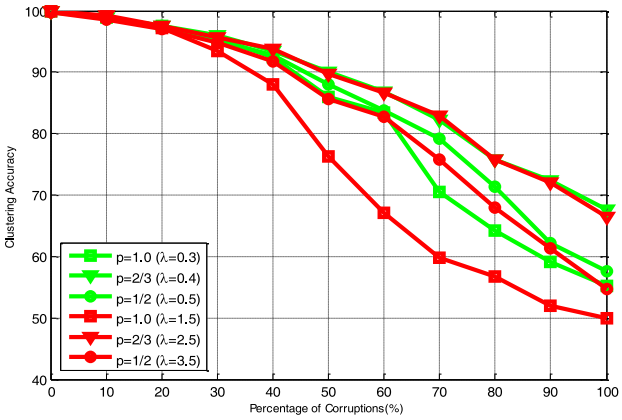

Fig. 3. Comparison of clustering accuracy on the synthetic data with the increasing percentage of corruptions from 0% to 100% for both S_pNM_LRR (green) and S_pNF_LRR (red) with $d = 20$. Note that the parameter λ is tuned to be the best at 50% percentage of corruptions.

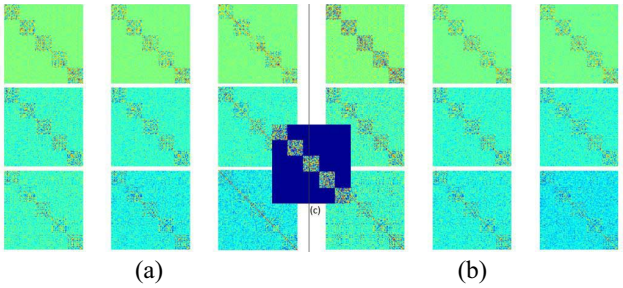

Fig. 4. For the clean data added by the 0%, 30%, and 60% percentage of corruptions from top to bottom, respectively. The comparison of the (c) true representation matrix; both the (a) ones obtained by S_pNM_LRR and (b) ones obtained by S_pNF_LRR with $d = 20$. Note that (a) and (b) are obtained by setting $p = 1, 2/3, 1/2$ from left to right, respectively.

by Algorithm 1. If A^TA is invertible and $\lim_{k\to\infty} \mu_k\{\mathcal{N}_{k+1} - \mathcal{N}_k\} = (\mathbf{0}, \mathbf{0})$ with $\mathcal{N}_k = (Z_k, E_k)$ hold, we have the following.

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

2) Any accumulation point $\mathcal{T}_* = (J_*, Z_*, E_*; Y_{1,*}, Y_{2,*})$ of sequence $\{\mathcal{T}_k\}_{k=1}^{\infty}$ satisfies the first-order KKT conditions. Specially, $\{\mathcal{T}_k\}_{k=1}^{\infty}$ converges to a KKT point, i.e.,

$$\begin{cases} \mathbf{0} = J_* - Z_* & (13) \\ \mathbf{0} = AZ_* + E_* - X & (14) \\ \mathbf{0} \in \partial \|J_*\|_{S_p}^p + Y_{1,*} & (15) \\ \mathbf{0} = A^T Y_{2,*} - Y_{1,*} & (16) \\ \mathbf{0} \in \lambda \partial \|E_*\|_\ell + Y_{2,*}. & (17) \end{cases}$$

VI. Experiments

This section will conduct several experiments on various subspace segmentation tasks to verify the *efficacy* and *efficiency* of Algorithms 1 and 2 for solving the problem

$$\min_{\{Z,E\}} \text{rank}(Z) + \lambda \|E\|_{2,1}, \quad \text{s.t.} \quad X = XZ + E. \qquad (18)$$

Note that (18) is the special example of (1) by setting $\|E\|_\ell = \|E\|_{2,1}$ and $A = X$, we mainly consider to use the Schatten-p norm with various p-values (e.g., $p = 1$, $2/3$ and $1/2$) and their upper bounding functions shown in Table II to replace the rank function in (18), respectively. Each algorithm converges until reaching the same stop condition, i.e., $\|XZ_{k+1} + E_{k+1} - X\|_\infty < 10^{-8}$. The obtained clustering accuracy and execution time will show the efficacy and efficiency on both synthetic data and five real-world databases (see Fig. 2). We here choose S_1NM_LRR[2] (i.e., LRR [12]) as the baseline. Honestly speaking, our methods cannot obtain state-of-the-art performance in clustering tasks since no additional constraints are considered in our models unlike [72] and [73]. The involved parameter(s) of each method are manually and carefully adjusted to achieve the best clustering performance, and we choose the best one in two sets: $d \in \{2, 10, 20, 40, 60, 80, 100, 120, 140, 160\}$ and $\lambda \in [0.01, 10]$. Additionally, besides comparing with both SSC method [11] and graph-based clustering methods[3] [74], [75] we also compare our methodologies with [42] though there is

[2]http://www.cis.pku.edu.cn/faculty/vision/zlin/lrr

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

TABLE IV
CLUSTERING RESULTS (%) AND TIME CONSUMPTION (S) OF THE USED CLUSTERING METHODS ON THE HOPKINS 155 DATA

Methods	p	Mean	StD	Max	Time
SSC [11]	—	4.02	10.04	39.53	1510
[42]	1	7.32	9.42	44.04	1030
	2/3	7.70	9.47	44.04	1196
	1/2	7.77	9.56	44.04	1270
S_pNM_LRR	1	3.39	6.41	36.36	1309
	2/3	3.84	6.75	36.36	1464
	1/2	4.04	7.01	36.36	1572
S_pNF_LRR	1	3.24	6.05	36.36	296
	2/3	5.20	8.34	37.97	321
	1/2	5.84	8.92	37.97	410

no doubt that nonconvex Schatten-p norm can be used to measure the coefficients matrix, and then applied to the subspace segmentation tasks. However, both (4) and (5) are, respectively, solved by Algorithm 1 and LADMMAP [42], which can lead to higher computational complexity since involving the SVD of large scale matrix. Note that [42] needs less timing consumption compared with Algorithm 1. This is because that the former involves less variables but more timing costs than Algorithm 2, which uses the decomposable strategies, while both [42] and Algorithm 1 do not. Both [74] and [75] are relatively faster clustering methods since no SVD computations. All the experiments were implemented by MATLAB2014a onto a PC with 4.0 GB of RAM and Intel Core i3-4158 CPU@3.50 GHZ.

A. Synthetic Data

Similar to the settings in [12], [50], and [51], we first construct five independent subspaces $\{S_i\}_{i=1}^5$, whose bases $\{U_i\}_{i=1}^5$ are generated by updating sequences $U_{i+1} = TU_i$ for $i = 1, 2, 3, 4$. Here, T is a random rotation matrix and U_1 is a random column orthogonal matrix of dimension 100×4, and each subspace has a rank of 4 and an ambient dimension of 100. We construct a 100×200 data matrix $X = [X_1, \ldots, X_5]$ by sampling 40 data vectors from each subspace by $X_i = U_i C_i$ with C_i being a 4×40 independent identically distributed standard Gaussian matrix. Then we can achieve the low-rank representation matrix with the size of 200×200. To further test the performance, we add dense Gaussian noise distributed to zero mean and 0.05^2 deviation on the clean data, and corrupt the data by randomly choosing different percentages of the entries in X and adding them with noises uniformly distributed on $1.21 \times [-0.5, 0.5]$.

The average clustering accuracies are reported in Fig. 3 after ten randomly generated realizations. As we can see, the clustering accuracy gradually decreases as the percentage of corruptions increase from 0% to 100% with the interval 10%. Moreover, both S_pNM_LRR and S_pNF_LRR with $p = 2/3$, $1/2$ obtain superior performance over $p = 1$. The visual comparisons of both the true representation matrix and the obtained ones by S_pNM_LRR and S_pNF_LRR are reported, respectively, in Fig. 4. Moreover, the representation matrices obtained by our methods can better approximate the true one especially for some slight corruptions in the clean data.

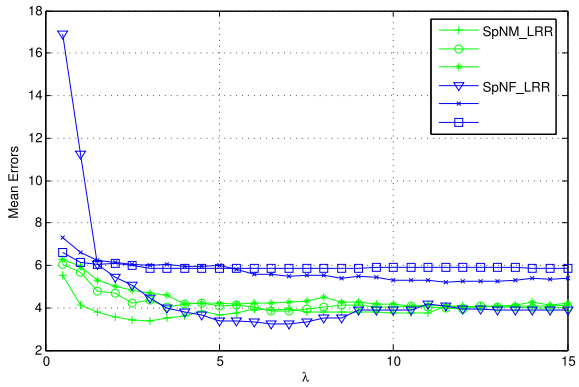

Fig. 5. Mean errors influenced by various choices of parameter λ in both methods for p =1, 2/3, 1/2 (top to bottom) on Hopkins 155, respectively.

B. Motion Segmentation

We here will evaluate the performance of both S_pNM_LRR and S_pNF_LRR on the Hopkins 155[4] database, which consists of 155 video sequences, 120 of which contain two moving objects and 35 of which contain three moving objects with the extracted feature points and their tracks across frames. Each of motion sequence corresponds to a single subspace. Inspired by [76], we use PCA to project the data into a 12-D subspace, and choose the same parameters for all the sequences and algorithms. The graph-based clustering methods are not compared due to their specific spectral embedding formulations in [74] and [75].

The maximum, mean, and standard deviation (StD) of the clustering errors are reported in Table IV, as well as the time consumptions. We observe that S_pNM_LRR can obtain lower mean errors than S_pNF_LRR, which can verify the efficacy of S_pNM_LRR. But S_pNF_LRR needs less time consumption than both S_pNM_LRR and [11], [42], which can verify the efficiency of S_pNF_LRR. And the cases of both S_pNM_LRR and S_pNF_LRR for $p = 2/3, 1/2$ have bigger mean errors than $p = 1$. The reason is that the true representation matrix has very lower rank, it can restrain the superiority of nonconvex Schatten-p norm as verified in [34]. Note that most sequences are easy to be segmented and hence all algorithms can obtain zero errors and medians in the experiments. Additionally, the clustering performance of both methods for all 155 motion sequences will largely depend on the choice of λ when the data are grossly corrupted. Fig. 5 shows the clustering errors for various values of $\lambda \in [0.5{:}0.5{:}15]$ in the abscissa when the rank parameter of S_pNF_LRR is fixed to 10, i.e., $d = 10$. It follows from the numerical results that the mean errors varies [3.39, 5.55], [3.84, 6.04], and [4.04, 6.26] for S_pNM_LRR and [3.24, 16.91], [5.20, 7.29], and [5.84, 6.60] for S_pNF_LRR with $p = 1$, $2/3$ and $1/2$ in sequence, respectively.

C. Face Images Clustering

This section will perform two experiments on the Extended Yale B[5] database, which contains face images of 38 subjects

[4]http://www.vision.jhu.edu/data
[5]http://vision.ucsd.edu/leekc/ExtYaleDatabase/ExtYale

TABLE V
CLUSTERING ACCURACY (ACC, %) AND TIME CONSUMPTION (S) OF ALL USED METHODS ON THE EXTENDED YALE B DATABASE

Methods	p	Raw		200D	
		ACC	Time	ACC	Time
SSC [11]	—	70.78	97	72.97	35
[42]	1	65.72	50	66.96	34
	2/3	69.78	55	71.34	38
	1/2	67.51	63	72.28	43
S_pNM_LRR	1	64.22	62	67.03	38
	2/3	72.34	71	73.12	43
	1/2	72.03	74	73.44	46
S_pNF_LRR	1	72.50	20	72.81	10
	2/3	72.81	22	73.12	12
	1/2	72.34	22	73.28	12
L_p_un [74]	1	37.71	5	39.53	5
CLR_L_p [75]	1	51.41	2	55.78	1
	2	51.41	2	55.78	1

(a)

(b)

Fig. 6. Partial shadows of the original images are removed to obtain the clean images in both (a) and (b). Each part (left to right) of them include the original\reconstructed\residual images for $p = 1, 2/3$ and $1/2$ from top row to bottom row, respectively.

with various illumination conditions. We here resize each original image into 48×42 to conduct two experiments, in which one uses the first ten subjects to form the 2016×640 raw matrix X, the other projects it onto a 200-D subspace by applying PCA for removing partial noises.

The results are shown in both Fig. 6 and Table V, we observe that the face images reconstructed by both S_pNM_LRR and S_pNF_LRR are slightly brightness compared with the original ones. The reason is since our methods are able to remove the shadows of the corrupted face images. In addition, it follows from Table V that S_pNF_LRR requires less time consumption than S_pNM_LRR by using the factor strategies. Both L_p_un [74] and CLR_L_p [75] as the compared graph-based clustering methods can obtain lower performance in efficacy though they consumes less timing in efficiency. Compared with the raw data, the time consumptions also become less and the clustering results are improved after the projecting procedure. This confirms that using PCA can not only reduce the dimension to save time consumption but also weaken the influence of noises to improve clustering accuracy. Meanwhile, we can see that nonconvex cases can obtain higher clustering results than convex case as listed in Table V. Especially, we choose $\lambda \in \{1.5, 2\}$ for S_pNM_LRR and $\lambda \in \{2, 3, 6, 8\}$ with $d = 100$ for S_pNF_LRR to obtain the best performance.

D. COIL20 Images Clustering

This section will experiments on the COIL20[6] database which has 32×32 gray scale images of 20 objects viewed from different angles. The experiments are carried out for the original images without any processing. We choose the number of subjects $K = 15$ and 20, and select the best parameters of all methods to achieve the best clustering results. Moreover, the convergence behaviors of both S_pNM_LRR and S_pNF_LRR methods are shown in Fig. 7, in which the ordinate denotes the stopping criteria by $\|XZ_{k+1} + E_{k+1} - X\|_\infty$, and the abscissa is the number of iterations. The stoping error values of our approaches all drop with the increasing iterations number as demonstrated in Fig. 7. Especially for both S_pNF_LRR and S_pNM_LRR with $p = 1$, they drop very quickly and converge much fast. Since Algorithm 1$\gg$Algorithm 2(1/2)>Algorithm 2(2/3)> Algorithm 2(1) at each iteration, and both methods can run about 200 iterations to stop. Thus, it is easy to explain that S_pNF_LRR need less timing consumption than S_pNM_LRR. For $K = 15$ and 20 subjects, the clustering accuracy and the timing consumption obtained by both S_pNM_LRR and S_pNF_LRR methods together with the compared methods [11], [42], [74], [75] are shown in Table VI. Obviously, our methods can show their superiority in both efficacy and efficiency compared with other methods in general. Moreover, we also have a similar observation with the face images clustering, the computational efficiency for solving S_pNF_LRR are significantly improved over S_pNM_LRR. When the number of subjects are ranged from 15 to 20, the clustering results become lower and the time consumption becomes higher. Especially, the clustering accuracy of S_pNM_LRR can be achieved by setting $\lambda \in \{0.2, 0.5, 0.7, 0.8\}$, while the ones of S_pNF_LRR are obtained by letting $d = 140$ and $\lambda \in \{0.3, 1.5, 3.0\}$.

E. Digital Images Clustering

MNIST[7] is a handwritten digit database with ten subjects. We select a subset (i.e., the first 100 images of each digit) with each image having 28×28 pixels. USPS[8] is also a handwritten digit database of 9298 images with each image having 16×16 pixels. We also select the first 100 images of each digit.

[6]http://www.cs.columbia.edu/CAVE/software/softlib/coil-20
[7]http://yann.lecun.com/exdb/mnist
[8]http://www.gaussianprocess.org/gpml/data

TABLE VI
CLUSTERING ACCURACY (%) AND TIME CONSUMPTION (S) OF ALL USED METHODS ON THE COIL20 DATABASE WITH 15 AND 20 SUBJECTS

Methods	p	15 subjects		20 subjects	
		ACC	Time	ACC	Time
SSC [11]	—	68.70	222	65.63	342
[42]	1	65.09	145	64.38	229
	2/3	67.31	177	65.06	285
	1/2	66.46	215	64.93	352
S_pNM_LRR	1	67.50	197	64.93	450
	2/3	69.54	242	66.94	553
	1/2	68.80	298	65.76	611
S_pNF_LRR	1	64.07	31	63.13	48
	2/3	67.13	31	66.74	48
	1/2	66.67	32	65.76	50
L_p_un [74]	1	60.31	22	55.42	55
CLR_L_p [75]	1	66.11	4	65.42	10
	2	64.72	4	64.51	9

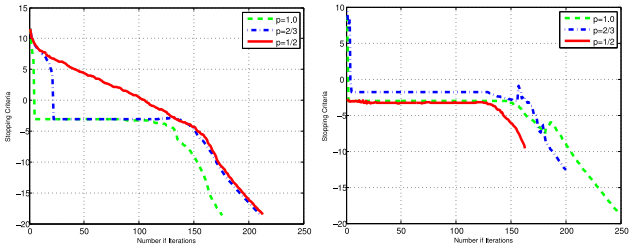

Fig. 7. Convergence of the stopping criteria (log) versus number of iterations obtained by both methods on 15 subjects of COIL20 database, respectively. (a) S_pNM_LRR. (b) S_pNF_LRR.

TABLE VII
CLUSTERING ACCURACY (%) AND TIME CONSUMPTION (S) OF ALL USED METHODS ON THE TWO HANDWRITTEN DIGITS DATABASES

Methods	p	MNIST		USPS	
		ACC	Time	ACC	Time
SSC [11]	—	68.0	213	74.3	120
[42]	1	67.5	119	76.9	106
	2/3	68.9	136	78.3	123
	1/2	68.1	156	77.4	126
S_pNM_LRR	1	68.5	147	78.0	140
	2/3	69.9	183	78.7	169
	1/2	69.2	209	78.3	181
S_pNF_LRR	1	66.4	20	70.9	18
	2/3	68.9	25	78.2	21
	1/2	67.5	28	76.5	23
L_p_un [74]	1	45.0	21	65.7	19
CLR_L_p [75]	1	55.7	5	76.7	5
	2	55.7	5	76.7	5

Here, each related digit coming from both MNIST and USPS will be represented by 784- and 256-D vector, respectively. Fig. 8 and Table VII show the constructed affinity matrices and the best clustering accuracies of all involved methods on both MNIST and USPS, respectively. Besides, our Schatten-p norm with $p = 1, 2/3$ and $1/2$ can produce the block-diagonal coefficient matrix. Except the compared graph-based clustering methods [74], [75], S_pNF_LRR method is still the faster method among these methods, which can further show its superiority in efficiency. It follows from Fig. 8 that the affinity matrices of Fig. 8(b) reveal clear block-diagonal and blurred non block-diagonal structures, which can further explain that

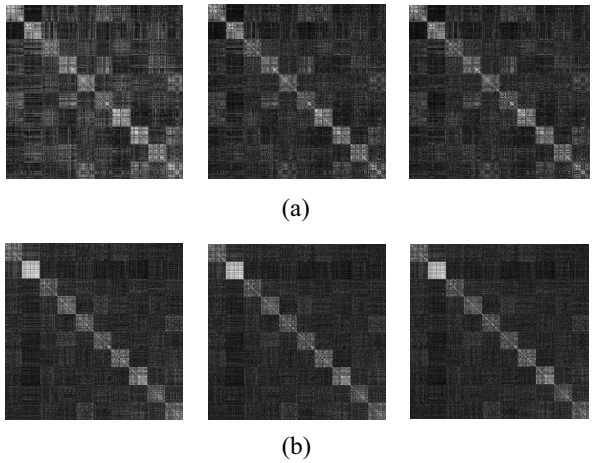

Fig. 8. Comparison of the affinity matrices produced by S_pNM_LRR with $p = 1, 2/3$ and $1/2$ (left to right) on both (a) MNIST and (b) USPS.

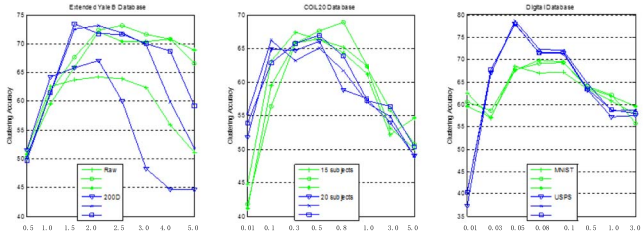

Fig. 9. Clustering accuracies obtained by S_pNM_LRR with $p = 1, 2/3$ and $1/2$ (top to bottom) on four databases under various parameter choices.

each subject becomes highly compact and different subjects become far apart. We observe that Fig. 8(a) do not show these obvious differences of the block structures no matter how the involved subjects are same or not. As a whole, Fig. 8(b) can better capture the block-diagonal structure than Fig. 8(a). This visual results can reflect the fact that the clustering results obtained by S_pNM_LRR on USPS are higher than MNIST in Table VII. Similar trends have been verified in [72]. We choose $\lambda \in \{0.05, 0.06, 0.08\}$ for S_pNM_LRR and S_pNF_LRR with $\lambda \in \{0.5, 2\}$ and $d = 120$.

F. Parameters Discussion

This section will give the optimal choice of parameters. Besides different p-values (e.g., $p = 1, 2/3$ and $1/2$) for the Schatten-p norm in both S_pNM_LRR and S_pNF_LRR, there exist the common parameter λ, and an additional ingredient for the expected rank number d of the coefficient matrix in S_pNF_LRR. Thus, the discussions of λ for S_pNM_LRR and (λ, d) for S_pNF_LRR are given independently as follows.

For S_pNM_LRR, the choice of the parameter λ is not unique on the above each of experiments. There exists a recent study in [77], which states that estimating λ needs to normalize the column vectors of the data matrix X and manually tune it around $1/\sqrt{\log(n)}$, i.e., there exists a range for $\lambda \in [\lambda_{\min}, \lambda_{\max}]$, where both $\lambda_{\min}$ and $\lambda_{\max}$ will decrease as the total number of the samples, i.e., n. However, the suggested parameter $1/\sqrt{\log(n)}$ is purely a moderately "good" choice, but not the best parameter configuration for S_pNM_LRR with

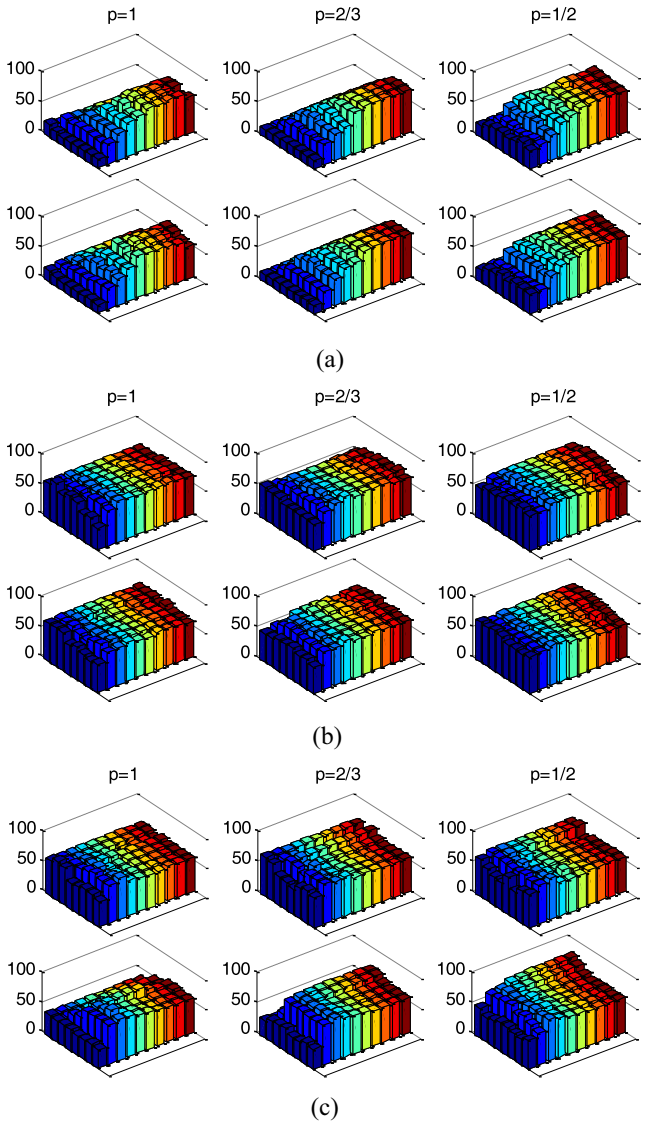

Fig. 10. Clustering accuracies are obtained by the S_pNF_LRR on three databases under various choices of parameters. Note that X-label represents the values of given d, and Y-label represents the parameter λ values, in which $\lambda \in \{0.1, 0.5, 1, 2, 4, 6, 8, 10\}$ for (a), $\lambda \in \{0.05, 0.1, 0.3, 0.5, 0.8, 1, 2, 3\}$ for (b), and $\lambda \in \{0.01, 0.1, 0.5, 1, 2, 4, 6, 8\}$ for (c), respectively. (a) Extended Yale B Database. (b) COIL20 Database. (c) Digital Database.

$p = 1$. The clustering accuracy obtained in our experimental settings are shown in Fig. 9 under different values of λ. The results are insensitive to λ in a small range such as $[1.0, 3.0]$ for the Extended Yale B, $[0.1, 0.8]$ for the COIL20 and $[0.03, 0.1]$ for both USPS and MNIST digital databases.

For S_pNF_LRR, we analyze the affects of the parameter pairs (λ, d) for the clustering accuracy. The results are reported in Fig. 10(a)–(c) for various (λ, d). Since different clustering data accords to different rank number of representation matrix, which has various effect on the clustering accuracy even for the same λ. Specially, S_pNF_LRR can obtain lower clustering results when d is smaller than the true rank. We know that the determination of reducing rank is an open problem, several works [36], [54], [78] have provided some rank estimation strategies to achieve a good d. Inspired by them, we set d sufficiently large (e.g., at least larger than the true rank) to obtain a higher clustering accuracy. This agrees with [79] which says as long as d is large enough, any local minima can be a global optima. Thus, we pick a relatively larger d ($\geq$80), S_pNF_LRR can achieve higher performance.

VII. Conclusion

In this paper, we present Schatten-p norm with differen p-values for measuring low rank coefficients matrix and propose two nonconvex LRR models including S_pNM_LRR and S_pNF_LRR. The main merits of both models are their accurate approximation and decomposition strategy for their nearly unbiased relaxations of the rank function. They can help us to achieve the higher clustering accuracy and save the time consumption in order to have a wide range of applications. Meanwhile, nonconvex multiblock ADMM algorithms have been devised to solve the proposed Schatten-p norm-based methods, and then we give the algorithmic analysis from both computational complexity and convergence guarantees under some mild assumptions. Experiments on both synthesized and real world data can demonstrate the superiority of our proposed methods for the subspace clustering tasks.

Consider that both S_pNM_LRR and S_pNF_LRR are optimized by the nonconvex ADMMs with multiblocks, one can prove their convergence properties under relatively weaker conditions, e.g., [34], [49], [66], and [69]. Furthermore, the Schatten-p norm factorization strategy can also be used to solve more complicated LRR variants and other low rank matrix recovery models, e.g., [70], [75], [80], and [81]. Besides, we can extend the proposed both methods for dealing with the large scale and deep subspace clustering problems, e.g., [82] and [83].

Acknowledgment

The authors would like to thank the Editor-in-Chief, the Associate Editor, and the anonymous reviewers for their critical and constructive comments and suggestions.

References

[1] T. Bouwmans, "Traditional and recent approaches in background modeling for foreground detection: An overview," *Comput. Sci. Rev.*, vols. 11–12, pp. 31–66, May 2014.
[2] J. Ho, M.-H. Yang, J. Lim, K.-C. Lee, and D. Kriegman, "Clustering appearances of objects under varying illumination conditions," in *Proc. IEEE Conf. Comput. Vis. Pattern. Recognit. (CVPR)*, vol. 1. Madison, WI, USA, 2003, pp. 11–18.
[3] R. Vidal, R. Tron, and R. Hartley, "Multiframe motion segmentation with missing data using powerfactorization and GPCA," *Int. J. Comput. Vis.*, vol. 79, no. 1, pp. 85–105, 2008.
[4] A. Y. Yang, J. Wright, Y. Ma, and S. S. Sastry, "Unsupervised segmentation of natural images via lossy data compression," *Comput. Vis. Image. Understand.*, vol. 110, no. 2, pp. 212–225, 2008.
[5] P. S. Bradley and O. L. Mangasarian, "k-plane clustering," *J. Glob. Optim.*, vol. 16, no. 1, pp. 23–32, 2000.
[6] L. Lu and R. Vidal, "Combined central and subspace clustering for computer vision applications," in *Proc. Int. Conf. Mach. Learn. (ICML)*, Pittsburgh, PA, USA, 2006, pp. 593–600.
[7] J. P. Costeira and T. Kanade, "A multibody factorization method for independently moving objects," *Int. J. Comput. Vis.*, vol. 29, no. 3, pp. 159–179, 1998.
[8] R. Vidal, Y. Ma, and S. Sastry, "Generalized principal component analysis (GPCA)," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 12, pp. 1945–1959, Dec. 2005.

[9] M. E. Tipping and C. M. Bishop, "Mixtures of probabilistic principal component analyzers," *Neural. Comput.*, vol. 11, no. 2, pp. 443–482, 1999.

[10] S. Rao, R. Tron, R. Vidal, and Y. Ma, "Motion segmentation in the presence of outlying, incomplete, or corrupted trajectories," *IEEE Trans. Pattern. Anal. Mach. Intell.*, vol. 32, no. 10, pp. 1832–1845, Oct. 2010.

[11] E. Elhamifar and R. Vidal, "Sparse subspace clustering: Algorithm, theory, and applications," *IEEE Trans. Pattern. Anal. Mach. Intell.*, vol. 35, no. 11, pp. 2765–2781, Nov. 2013.

[12] G. Liu *et al.*, "Robust recovery of subspace structures by low-rank representation," *IEEE Trans. Pattern. Anal. Mach. Intell.*, vol. 35, no. 1, pp. 171–184, Jan. 2013.

[13] R. Vidal and P. Favaro, "Low rank subspace clustering (LRSC)," *Pattern. Recognit. Lett.*, vol. 43, pp. 47–61, Jul. 2014.

[14] G. Lerman and T. Zhang, "Robust recovery of multiple subspaces by geometric l_p minimization," *Ann. Stat.*, vol. 39, no. 5, pp. 2686–2715, 2011.

[15] E. J. Candes and Y. Plan, "Matrix completion with noise," *Proc. IEEE*, vol. 98, no. 6, pp. 925–936, Jun. 2010.

[16] 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.

[17] E. J. Candes, X. Li, Y. Ma, and J. Wright, "Robust principal component analysis?" *J. ACM.*, vol. 58, no. 3, pp. 1–37, 2011.

[18] G. Liu, Q. Liu, and P. Li, "Blessing of dimensionality: Recovering mixture data via dictionary pursuit," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 1, pp. 47–60, Jan. 2017.

[19] M. Fazel, H. Hindi, and S. P. Boyd, "A rank minimization heuristic with application to minimum order system approximation," in *Proc. Amer. Control. Conf (ACC)*, vol. 6. Arlington, VA, USA, 2001, pp. 4734–4739.

[20] 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.

[21] M. Jaggi and M. Sulovský, "A simple algorithm for nuclear norm regularized problems," in *Proc. Int. Conf. Mach. Learn. (ICML)*, Haifa, Israel, 2011, pp. 471–478.

[22] K.-C. Toh and S. Yun, "An accelerated proximal gradient algorithm for nuclear norm regularized least squares problems," *Pac. J. Optim.*, vol. 6, no. 3, pp. 615–640, 2010.

[23] 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.

[24] 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.

[25] 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, 2011.

[26] Z. Lin, M. Chen, L. Wu, and Y. Ma, "The augmented lagrange multiplier method for exact recovery of corrupted low-rank matrices," *arXiv preprint arXiv:1009.5055*, 2010.

[27] 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.

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

[29] 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.

[30] 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.

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

[32] M. Malek-Mohammadi, M. Babaie-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.

[33] M. Zhang, Z.-H. 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.

[34] L. Luo, J. Yang, J. Qian, Y. Tai, and G.-F. 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.

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

[36] N. Srebro, J. D. M. Rennie, and T. S. Jaakkola, "Maximum-margin matrix factorization," in *Proc. Adv. Neural Inf. Process. Syst. (NIPS)*, 2004, pp. 1329–1336.

[37] F. Shang, Y. Liu, and J. Cheng, "Scalable algorithms for tractable Schatten quasi-norm minimization," in *Proc. Assoc. Adv. Artif. Intell. (AAAI)*, Phoenix, AZ, USA, 2016, pp. 2016–2022.

[38] F. Nie and H. Huang, "Subspace clustering via new low-rank model with discrete group structure constraint," in *Proc. Int. Joint. Conf. Artif. Intell. (IJCAI)*, New York, NY, USA, 2016, pp. 1874–1880.

[39] F. Nie, H. Wang, X. Cai, H. Huang, and C. Ding, "Robust matrix completion via joint Schatten p-norm and l_p-norm minimization," in *Proc. IEEE Int. Conf. Data. Mining (ICDM)*, Brussels, Belgium, 2012, pp. 566–574.

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

[41] 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.

[42] X. Zhang, C. Xu, X. Sun, and G. Baciu, "Schatten-q regularizer constrained low rank subspace clustering model," *Neurocomputing*, vol. 182, pp. 36–47, Mar. 2016.

[43] S. Gu *et al.*, "Weighted nuclear norm minimization and its applications to low level vision," *Int. J. Comput. Vis.*, vol. 121, no. 2, pp. 183–208, 2017.

[44] 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.

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

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

[47] C. Lu, C. Zhu, C. Xu, S. Yan, and Z. Lin, "Generalized singular value thresholding," in *Proc. Assoc. Adv. Artif. Intell. (AAAI)*, Austin, TX, USA, 2015, pp. 1805–1811.

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

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

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

[51] Y. Liu, L. C. Jiao, and F. Shang, "A fast tri-factorization method for low-rank matrix recovery and completion," *Pattern Recognit.*, vol. 46, no. 1, pp. 163–173, 2013.

[52] F. Shang, Y. Liu, and J. Cheng, "Tractable and scalable schatten quasi-norm approximations for rank minimization," in *Proc. Int. Conf. Artif. Intell. Stat. (AISTATS)*, 2016, pp. 620–629.

[53] F. Shang, Y. Liu, and J. Cheng, "Unified scalable equivalent formulations for Schatten quasi-norms," *arXiv preprint arXiv:1606.00668*, 2016.

[54] F. Shang, J. Cheng, Y. Liu, Z.-Q. Luo, and Z. Lin, "Bilinear factor matrix norm minimization for robust PCA: Algorithms and applications," *IEEE Trans. Pattern. Anal. Mach. Intell.*, to be published, doi: 10.1109/TPAMI.2017.2748590.

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

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

[57] X. Peng, J. Lu, Z. Yi, and R. Yan, "Automatic subspace learning via principal coefficients embedding," *IEEE Trans. Cybern.*, vol. 47, no. 11, pp. 3583–3596, Nov. 2017.

[58] X. Peng, Z. Yu, Z. Yi, and H. Tang, "Constructing the L2-graph for robust subspace learning and subspace clustering," *IEEE Trans. Cybern.*, vol. 47, no. 4, pp. 1053–1066, Apr. 2017.

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

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

[61] J. Wang, X. Wang, F. Tian, C. H. Liu, and H. Yu, "Constrained low-rank representation for robust subspace clustering," *IEEE Trans. Cybern.*, vol. 47, no. 12, pp. 4534–4546, Dec. 2017.

[62] F. Nie, X. Wang, and H. Huang, "Clustering and projected clustering with adaptive neighbors," in *Proc. KDD*, New York, NY, USA, 2014, pp. 977–986.

[63] 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.
[64] J. Shi and J. Malik, "Normalized cuts and image segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 8, pp. 888–905, Aug. 2000.
[65] 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.
[66] T. Lin, S. Ma, and S. Zhang, "On the global linear convergence of the ADMM with multiblock variables," *SIAM J. Optim.*, vol. 25, no. 3, pp. 1478–1497, 2015.
[67] 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, nos. 1–2, pp. 57–79, 2016.
[68] J. Yang and X. Yuan, "Linearized augmented Lagrangian and alternating direction methods for nuclear norm minimization," *Math. Comput.*, vol. 82, no. 281, pp. 301–329, 2013.
[69] S. Magnússon, P. C. Weeraddana, M. G. Rabbat, and C. Fischione, "On the convergence of alternating direction Lagrangian methods for nonconvex structured optimization problems," *IEEE Trans. Control Netw. Syst.*, vol. 3, no. 3, pp. 296–309, Sep. 2016.
[70] C. Georgakis, Y. Panagakis, and M. Pantic, "Dynamic behavior analysis via structured rank minimization," *Int. J. Comput. Vis.*, vol. 126, nos. 2–4, pp. 333–357, 2018.
[71] A. S. Lewis and H. S. Sendov, "Nonsmooth analysis of singular values. Part I: Theory," *Set Valued Anal.*, vol. 13, no. 3, pp. 213–241, 2005.
[72] J. Liu, Y. Chen, J. Zhang, and Z. Xu, "Enhancing low-rank subspace clustering by manifold regularization," *IEEE Trans. Image Process.*, vol. 23, no. 9, pp. 4022–4030, Sep. 2014.
[73] H. Gao, F. Nie, X. Li, and H. Huang, "Multi-view subspace clustering," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Santiago, Chile, 2015, pp. 4238–4246.
[74] F. Nie, H. Wang, H. Huang, and C. Ding, "Unsupervised and semi-supervised learning via l_1-norm graph," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Barcelona, Spain, 2011, pp. 2268–2273.
[75] F. Nie, X. Wang, M. I. Jordan, and H. Huang, "The constrained Laplacian rank algorithm for graph-based clustering," in *Proc. Assoc. Adv. Artif. Intell. (AAAI)*, Phoenix, AZ, USA, 2016, pp. 1969–1976.
[76] C. Lu *et al.*, "Robust and efficient subspace segmentation via least squares regression," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, Florence, Italy, 2012, pp. 347–360.
[77] G. Liu, H. Xu, J. Tang, Q. Liu, and S. Yan, "A deterministic analysis for LRR," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 38, no. 3, pp. 417–430, Mar. 2016.
[78] 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.
[79] S. Burer and R. D. C. Monteiro, "Local minima and convergence in low-rank semidefinite programming," *Math. Program.*, vol. 103, no. 3, pp. 427–444, 2005.
[80] D. Luo, F. Nie, C. Ding, and H. Huang, "Multi-subspace representation and discovery," in *Proc. ECML PKDD*, 2011, pp. 405–420.
[81] S. Chen *et al.*, "Low-rank latent pattern approximation with applications to robust image classification," *IEEE Trans. Image Process.*, vol. 26, no. 11, pp. 5519–5530, Nov. 2017.
[82] X. Peng, H. Tang, L. Zhang, Z. Yi, and S. Xiao, "A unified framework for representation-based subspace clustering of out-of-sample and large-scale data," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 27, no. 12, pp. 2499–2512, Dec. 2016.
[83] X. Peng, S. Xiao, J. Feng, W.-Y. Yau, and Z. Yi, "Deep subspace clustering with sparsity prior," in *Proc. Int. Joint Conf. Artif. Intell. (IJCAI)*, New York, NY, USA, 2016, pp. 1925–1931.

Hengmin Zhang received the B.S. degree from the School of Mathematics and Statistics, Heze University, Heze, China, in 2010 and the M.S. degree from the College of Science, China University of Mining and Technology University, 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, Bayesian learning, and nonconvex optimizations.

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 Biometrics Centre, 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 over 100 scientific papers in pattern recognition and computer vision. His journal papers have been cited over 4000 times in the ISI Web of Science and 9000 times in the Web of Scholar Google. His current research interests include pattern recognition, computer vision, and machine learning. He is/was currently an Associate Editor of *Pattern Recognition Letters*, the IEEE TRANSACTIONS ON NEURAL NETWORKS AND LEARNING SYSTEMS, and *Neurocomputing*. He is a fellow of IAPR.

Fanhua Shang (M'14) received the Ph.D. degree in circuits and systems from Xidian University, Xi'an, China, in 2012.

He was a Research Associate with the Department of Computer Science and Engineering, Chinese University of Hong Kong, Hong Kong. He is currently a Professor with the School of Artificial Intelligence, Xidian University. From 2013 to 2015, he was a Post-Doctoral Research Fellow with the Department of Computer Science and Engineering, Chinese University of Hong Kong. From 2012 to 2013, he was a Post-Doctoral Research Associate with the Department of Electrical and Computer Engineering, Duke University, Durham, NC, USA. His current research interests include machine learning, data mining, pattern recognition, and computer vision.

Chen Gong (M'16) received the dual Ph.D. degrees from Shanghai Jiao Tong University (SJTU), Shanghai, China, and the University of Technology Sydney, Ultimo, NSW, Australia, in 2016, under the supervision of Prof. J. Yang and Prof. D. Tao, respectively.

He is currently a Professor with the School of Computer Science and Engineering, Nanjing University of Science and Technology, Nanjing, China. He has published over 30 technical papers at prominent journals and conferences, such as the the IEEE TRANSACTIONS ON NEURAL NETWORKS AND LEARNING SYSTEMS, the IEEE TRANSACTIONS ON IMAGE PROCESSING, the IEEE TRANSACTIONS ON CYBERNETICS, CVPR, AAAI, IJCAI, and ICDM. His current research interests include machine learning and data mining.

Dr. Gong was a recipient of the Excellent Doctorial Dissertation Award by SJTU and the Chinese Association for Artificial Intelligence. He was also enrolled by the Summit of the Six Top Talents Program of Jiangsu Province, China.

Zhenyu Zhang received the B.S. degree from the School of Science, Nanjing University of Science and Technology, Nanjing, China, in 2015, where he is currently pursuing the Ph.D. degree with the Jiangsu Key Laboratory of Image and Video Understanding for Social Security, School of Computer Science and Engineering.

His current research interests include depth estimation, semantic segmentation, video processing, and deep neural networks.