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., $\SpNM$ LRR) and the other is Schatten-$p$ norm factorization-based LRR (i.e., $\SpNF$ 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.

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 $\ell_1$-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

$$\min_{Z, E} \text{rank}(Z) + \lambda \|E\|_1, \quad \text{s.t. } AZ + E = X$$

where $\lambda > 0$ is a balanced parameter, and $\text{rank}(Z)$ counts the number of nonzero singular values of matrix $Z$, and $\|E\|_1$ 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 $||E||_F = ||E||_1$ 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 the decomposable formulations and thresholding operators. 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 rigourous 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 $S_p$NM_LRR for increasing efficacy, while the other is $S_p$NFG_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.

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.

<table>
<thead>
<tr>
<th>$p$</th>
<th>$\psi_p(U, V)$</th>
<th>$\min_{U, V} \psi_p(U, V)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>$\frac{1}{2} |U|_F^2 + |V|_F^2$</td>
<td>$\frac{1}{2} |U|_F^2 + |V|_F^2$</td>
</tr>
<tr>
<td>2/3</td>
<td>$\frac{1}{3} |U|_F^2 + |V|_F^2$</td>
<td>$\frac{1}{3} |U|_F^2 + |V|_F^2$</td>
</tr>
<tr>
<td>1/2</td>
<td>$\frac{1}{2} |U|_F + |V|_F$</td>
<td>$\frac{1}{2} |U|_F + |V|_F$</td>
</tr>
</tbody>
</table>

**TABLE II**

Used convex and nonconvex Schatten-$p$ norm and their bounding functions $\psi_p(U, V)$ with different $p$-values choice.

<table>
<thead>
<tr>
<th>$p$-value</th>
<th>$|Z|_{S_p}$</th>
<th>$\min_{u,v} \psi_{S_p}(U, V)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>$\frac{1}{2} |U|_F^2 + |V|_F^2$</td>
<td>$\frac{1}{2} |U|_F^2 + |V|_F^2$</td>
</tr>
<tr>
<td>2/3</td>
<td>$\frac{1}{3} |U|_F^2 + |V|_F^2$</td>
<td>$\frac{1}{3} |U|_F^2 + |V|_F^2$</td>
</tr>
<tr>
<td>1/2</td>
<td>$\frac{1}{2} |U|_F + |V|_F$</td>
<td>$\frac{1}{2} |U|_F + |V|_F$</td>
</tr>
</tbody>
</table>

**TABLE I**

Schatten-$p$ Functions with $p = 0, 1/2, 2/3, 1$ in (2) and their corresponding thresholding operators in (3).
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-\(p\) formulation of nonconvex Schatten-\(p\) norm can show its superiority over nuclear norm for relaxing the rank number is relatively larger, nonconvex Schatten-\(p\) norm with \(\sigma(Y) = [0, 3]\) 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 \(S_p\)NM_LRR model with two constraints as follows:

\[
\min_{[J,Z,E]} \frac{1}{2} \|J\|_{Sp}^p + \lambda \|E\|_\ell, \quad \text{s.t.} \quad AZ + E = X, J = Z
\]

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 \(S_p\)NM clustering method [42] aims to solve the problem with one constraint

\[
\min_{[J,Z,E]} \frac{1}{2} \|J\|_{Sp}^p + \lambda \|E\|_\ell, \quad \text{s.t.} \quad AZ + E = X
\]

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 \(\|\cdot\|_{BIN}\), and Forbenius/Nuclear hybrid norm \(\|\cdot\|_{F/N}\), respectively), we can give the corresponding alternative formulations listed in Table II. For any coefficients matrix \(Z \in \mathbb{R}^{m \times n}\) with rank\(Z = r \leq d\) [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 \(S_p\)NF_LRR model:

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

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 $S_p$NM_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 to multiple variables are involved in the objective function. These can motivate us to obtain the special formulations with at least two constraints, denoted by

$$S_1\text{NF}_LRR : \min_{(U, V, Z)} \frac{1}{2} \left( \| U \|_F^2 + \| V \|_F^2 \right) + \lambda \| E \|_\ell$$

s.t. $AZ + E = X$, $UV^T = Z$

$$S_2\text{NF}_LRR : \min_{(U, V, M, Z)} \frac{1}{2} \left( \| M \|_* + \| V \|_F^2 \right) + \lambda \| E \|_\ell$$

s.t. $AZ + E = X$, $UV^T = Z$, $M = U$

$$S_3\text{NF}_LRR : \min_{(U, V, M, N, E)} \frac{1}{2} \left( \| M \|_* + \| N \|_* + \| E \|_\ell \right)$$

s.t. $AZ + E = X$, $UV^T = Z$, $M = U$, $N = V$.

This paper will evaluates the subspace segmentation ability of both $S_p$NM_LRR and $S_p$NF_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 efficiency and efficacy, 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 $Z_a$ obtained by Algorithms 1 and 2 can be used to construct an affinity matrix $W^1$, i.e., $|Z_a^1 + |Z_a^2^T|)/2$, 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 $S_p$NM_LRR (4) and $S_p$NF_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. $S_p$NM_LRR via ADMM Algorithm

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

$$L_\mu(J, Z, E, Y_1, Y_2) = \|J\|_{Sp}^p + \lambda \| E \|_\ell + \langle Y_1, J - Z \rangle + \langle Y_2, AZ + E - X \rangle$$

$$+ \frac{\mu}{2} \left( \| J - Z \|_F^2 + \| AZ + E - X \|_F^2 \right)$$

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

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 “$x$” 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 $\|J\|_S^p$ with $p = 1, 2/3$ and 1/2 [34], [35] and $\|E\|_\ell$ with $\ell = l_2, l_1$ and $l_1$. 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 $S_p$NM_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 $0 < p < 1$.

\begin{itemize}
  \item [1.] While not converged do
  \item [2.] Update $J_{k+1}$ by solving
    $$J_{k+1} = \arg\min_J \|J\|_{Sp}^p + \frac{\mu}{2} \| J - (Z_k - Y_1k) \|_F^2.$$ 
  \item [3.] Update $Z_{k+1}$ by solving
    $$Z_{k+1} = \text{InT} \times \left( J_{k+1} + \frac{Y_1k}{\mu} - A^T (E_k - X + \frac{Y_2k}{\mu} j) \right);$$
  \item [4.] Update $E_{k+1}$ by computing
    $$E_{k+1} = \arg\min_E \| E \|_\ell + \frac{\mu}{2} \| (X - AZ_{k+1}) - Y_2k \|_F^2;$$
  \item [5.] Update the multipliers $Y_{1,k+1} = Y_{1,k} + \mu_k (J_{k+1} - Z_{k+1});$
    $$Y_{2,k+1} = Y_{2,k} + \mu_k (AZ_{k+1} + E_{k+1} - X);$$
  \item [6.] Update $\mu_{k+1} = \mu_{k+1} = \rho \mu_k;.$
  \item [7.] End while
\end{itemize}

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

---

Authorized licensed use limited to: NANJING UNIVERSITY OF SCIENCE AND TECHNOLOGY. Downloaded on July 21, 2020 at 08:49:55 UTC from IEEE Xplore. Restrictions apply.
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. \( S_p \)NF_LRR via ADMM Algorithm

Similar to Algorithm 1 for solving the \( S_p \)NM_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:

\[
\mathcal{L}_\mu(U, V, Z, E, Y_1, Y_2) = \frac{1}{2} \left( \|U\|_F^2 + \|V\|_F^2 \right) + \lambda \|E\|_\ell + \langle Y_1, AZ + E - X \rangle \\
+ \langle Y_2, UV^T - Z \rangle + \frac{\mu}{2} \left( \|UV^T - Z\|_F^2 + \|AZ + E - X\|_F^2 \right)
\]

(8)

\[
\mathcal{L}_\mu(U, V, M, Z, E, Y_1, Y_2, Y_3) = \frac{1}{2} \left( 2\|M\|_\ast + \|V\|_F^2 \right) + \lambda \|E\|_\ell + \langle Y_1, AZ + E - X \rangle \\
+ \langle Y_2, UV^T - Z \rangle + \langle Y_3, M - U \rangle \\
+ \frac{\mu}{2} \left( \|M - U\|_F^2 + \|UV^T - Z\|_F^2 + \|AZ + E - X\|_F^2 \right)
\]

(9)

\[
\mathcal{L}_\mu(U, V, M, N, Z, E, Y_1, Y_2, Y_3) = \frac{1}{2} \left( \|M\|_\ast + \|N\|_\ast \right) + \lambda \|E\|_\ell + \langle Y_1, AZ + E - X \rangle \\
+ \langle Y_2, UV^T - Z \rangle + \langle Y_3, M - U \rangle + \langle Y_4, N - V \rangle \\
+ \frac{\mu}{2} \left( \|M - U\|_F^2 + \|N - V\|_F^2 + \|UV^T - Z\|_F^2 + \|AZ + E - X\|_F^2 \right)
\]

(10)

Note that (8)–(10) are the corresponding augmented Lagrangian functions of \( S_p \)NF_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 \( \|Z\|_F^p \) 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 to 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 \( S_p \)NM_LRR and \( S_p \)NF_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 \( O(\min(p^2, p^2q)) \) 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
### Algorithm 2 \( S_pNF_{LRR} \) Optimized by ADMM

**Input:** \( X, A, p, \lambda > 0, d \) and \( \text{int} = (I + A^TA)^{-1} \).

**Initialization:** \( k = 0, \rho > 1, \mu_0 = 10^{-6}, V_0 = 0, Z_0 = 0, M_0 = 0, N_0 = 0, E_0 = 0, Y_{1,0} = 0, Y_{2,0} = 0, Y_{3,0} = 0, Y_{4,0} = 0 \).

1. **while** not converged do
2. if \( p = 1 \), update \( \{U_{k+1}, V_{k+1}\} \) by solving
   \[
   \begin{align*}
   U_{k+1} &= \arg\min_{U_k} \frac{1}{2} \|U_k\|_F^2 + \frac{\mu_k}{2} \|U_kV_k^T - Q_k\|_F^2; \\
   V_{k+1} &= \arg\min_{V_k} \frac{1}{2} \|V_k\|_F^2 + \frac{\mu_k}{2} \|U_{k+1}V_{k+1}^T - Q_k\|_F^2; \\
   if \quad p = 2/3, \quad update \quad \{U_{k+1}, V_{k+1}, M_{k+1}, N_{k+1}\} \quad by \quad solving \\
   U_{k+1} &= \arg\min_{U_k} \frac{1}{2} \|M_k - U + \frac{\tau_k}{\mu_k} M_k \|_F^2 + \frac{\mu_k}{2} \|U_{k+1}V_{k+1}^T - Q_k\|_F^2; \\
   V_{k+1} &= \arg\min_{V_k} \frac{1}{2} \|N_k - V + \frac{\tau_k}{\mu_k} V_k \|_F^2 + \frac{\mu_k}{2} \|M_{k+1}V_{k+1}^T - Q_k\|_F^2; \\
   M_{k+1} &= \arg\min_{M_k} \frac{1}{2} \|M_k - U_{k+1} + \frac{\tau_k}{\mu_k} U_{k+1} \|_F^2; \\
   N_{k+1} &= \arg\min_{N_k} \frac{1}{2} \|N_k - V_{k+1} + \frac{\tau_k}{\mu_k} V_{k+1} \|_F^2; \\
   \end{align*}
   \]
3. **update** \( \{Z_{k+1}, E_{k+1}\} \) by solving
   \[
   \begin{align*}
   Z_{k+1} &= \text{int} \times (U_{k+1}V_{k+1}^T + \frac{\tau_k}{\mu_k} U_{k+1}V_{k+1}^T - X); \\
   E_{k+1} &= \arg\min_{E_k} \frac{1}{2} \|E_k\|_F^2 + \frac{\mu_k}{2} \|E_k - \frac{1}{2}(X - AZ_{k+1} + \frac{\tau_k}{\mu_k} X)\|_F^2; \\
   \end{align*}
   \]
4. **update the multipliers by**
   \[
   \begin{align*}
   Y_{1,k+1} &= Y_{1,k} + \mu_k (AZ_{k+1} + E_{k+1} - X); \\
   Y_{2,k+1} &= Y_{2,k} + \mu_k (U_{k+1}V_{k+1}^T - Z_{k+1}); \\
   Y_{3,k+1} &= Y_{3,k} + \mu_k (M_{k+1} - U_{k+1}); \\
   Y_{4,k+1} &= Y_{4,k} + \mu_k (N_{k+1} - V_{k+1}); \\
   \end{align*}
   \]
5. **update** \( \mu_{k+1} \) by \( \mu_{k+1} = \rho \mu_k; \)
6. **end while**

**Output:** Optimal Representation Matrix \( Z_k \leftarrow Z_{k+1} \).

---

### Table III

**Computational Complexity of Several Updating Steps in Algorithms 1 and 2**

<table>
<thead>
<tr>
<th>Alg/Steps</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>#1 ( \mathcal{O}(n^2) )</td>
<td>( \mathcal{O}(n^2 + n_2^2 + d^2) )</td>
<td>( \mathcal{O}(n_2^2 + n^2 + n_2^2) )</td>
<td>( \mathcal{O}(n_2^2 + n^2) )</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>#2(2/3) ( \mathcal{O}(n_2^2 + n^2 + d^2) )</td>
<td>( \mathcal{O}(n_2^2 + n^2 + d^2) )</td>
<td>( \mathcal{O}(n^2) )</td>
<td>( \mathcal{O}(n_2^2 + n^2 + d^2) )</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>#2(1/2) ( \mathcal{O}(n_2^2 + n^2 + d^2) )</td>
<td>( \mathcal{O}(n_2^2 + n^2 + d^2) )</td>
<td>( \mathcal{O}(n_2^2 + n^2 + d^2) )</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

---

**Algorithm 2** \( S_pNF_{LRR} \) Optimized by ADMM in a decreasing order according to the computations in Table III.
by Algorithm 1. If $A^TA$ is invertible and $\lim_{k \to \infty} \mu_k \{N_{k+1} - N_k\} = (0, 0)$ with $N_k = (Z_k, E_k)$ hold, we have the following.

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

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

$$\begin{align*}
0 &= J_\ast - Z_\ast \\
0 &= AZ_\ast + E_\ast - X \\
0 &\in \partial \|J_\ast\|_p^p + Y_{1, \ast} \\
0 &= A^TY_{2, \ast} - Y_{1, \ast} \\
0 &\in \lambda \partial \|E_\ast\|_\ell + Y_{2, \ast}.
\end{align*}$$

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

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

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 $S_{1\text{NM-LRR}}^2$ (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

---

Footnotes:

$^2$http://www.cis.pku.edu.cn/faculty/vision/zlin/lrr

$^3$http://www.escience.cn/people/fpnie/papers.html
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 both [42] and Algorithm 1 do not. Both [74] and [75] are related to Algorithm 2, which uses the decomposable strategies, while 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 $100 \times 4$, and each subspace has a rank of 4 and an ambient dimension of 100. We construct a $100 \times 200$ data matrix $X = [X_1, \ldots, X_5]$ by sampling 40 data vectors from each subspace by $X_i = U_iC_i$ with $C_i$ being a $4 \times 40$ independent identically distributed standard Gaussian matrix. Then we can achieve the low-rank representation matrix with the size of $200 \times 200$. To further test the performance, we add dense Gaussian noise distributed to zero mean and 0.05 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 $Sp_{NM\_LRR}$ and $Sp_{NF\_LRR}$ 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 $Sp_{NM\_LRR}$ and $Sp_{NF\_LRR}$ 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.

<table>
<thead>
<tr>
<th>Methods</th>
<th>$p$</th>
<th>Mean</th>
<th>Std</th>
<th>Max</th>
<th>Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>$[42]$</td>
<td>1</td>
<td>7.32</td>
<td>9.42</td>
<td>44.04</td>
<td>1030</td>
</tr>
<tr>
<td>$Sp_{NM_LRR}$</td>
<td>1</td>
<td>3.39</td>
<td>6.41</td>
<td>36.36</td>
<td>1309</td>
</tr>
<tr>
<td>$Sp_{NF_LRR}$</td>
<td>2/3</td>
<td>3.84</td>
<td>7.65</td>
<td>36.36</td>
<td>1464</td>
</tr>
<tr>
<td></td>
<td>1/2</td>
<td>4.04</td>
<td>7.01</td>
<td>36.36</td>
<td>1572</td>
</tr>
</tbody>
</table>

B. Motion Segmentation

We here will evaluate the performance of both $Sp_{NM\_LRR}$ and $Sp_{NF\_LRR}$ 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 $Sp_{NM\_LRR}$ can obtain lower mean errors than $Sp_{NF\_LRR}$, which can verify the efficacy of $Sp_{NM\_LRR}$. But $Sp_{NF\_LRR}$ needs less time consumption than $Sp_{NM\_LRR}$ and [11], [42], which can verify the efficiency of $Sp_{NF\_LRR}$. And the cases of both $Sp_{NM\_LRR}$ and $Sp_{NF\_LRR}$ 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 $\lambda$ 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 $Sp_{NF\_LRR}$ 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 $Sp_{NM\_LRR}$ and $[3.24, 16.91], [5.20, 7.29]$, and $[5.84, 6.60]$ for $Sp_{NF\_LRR}$ 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

\footnotesize

\begin{table}[h]
\centering
\caption{Clustering Results (%) and Time Consumption (s) of the Used Clustering Methods on the Hopkins 155 Data}
\begin{tabular}{|c|c|c|c|c|c|}
\hline
Methods       & $p$  & Mean & Std  & Max  & Time  \\
\hline
$[42]$        & 1    & 7.32 | 9.42 | 44.04| 1030  \\
$Sp_{NM\_LRR}$& 1/2  & 3.39 | 6.41 | 36.36| 1309  \\
$Sp_{NF\_LRR}$& 2/3  & 3.84 | 7.65 | 36.36| 1464  \\
               & 1/2  & 4.04 | 7.01 | 36.36| 1572  \\
\hline
\end{tabular}
\end{table}

\footnotesize

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{fig5}
\caption{Mean errors influenced by various choices of parameter $\lambda$ in both methods for $p = 1, 2/3, 1/2$ (top to bottom) on Hopkins 155, respectively.}
\end{figure}

http://www.vision.jhu.edu/data

http://vision.ucsd.edu/leekc/ExtYaleDatabase/ExtYale
Authorized licensed use limited to: NANJING UNIVERSITY OF SCIENCE AND TECHNOLOGY. Downloaded on July 21,2020 at 08:49:55 UTC from IEEE Xplore. Restrictions apply.
TABLE VI
CLUSTERING ACCURACY (%) AND TIME CONSUMPTION (s) OF ALL USED METHODS ON THE COIL20 DATABASE WITH 15 AND 20 SUBJECTS

<table>
<thead>
<tr>
<th>Methods</th>
<th>p</th>
<th>15 subjects</th>
<th></th>
<th>15 subjects</th>
<th></th>
<th>20 subjects</th>
<th></th>
<th>20 subjects</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>ACC</td>
<td>Time</td>
<td>ACC</td>
<td>Time</td>
<td>ACC</td>
<td>Time</td>
<td>ACC</td>
</tr>
<tr>
<td>SSC [11]</td>
<td>1</td>
<td>68.70</td>
<td>222</td>
<td>65.63</td>
<td>342</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>2/3</td>
<td>65.09</td>
<td>145</td>
<td>64.58</td>
<td>229</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>1/2</td>
<td>66.66</td>
<td>215</td>
<td>64.93</td>
<td>352</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>S\textsubscript{p}NM_LRR</td>
<td>1</td>
<td>67.50</td>
<td>197</td>
<td>64.93</td>
<td>450</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>2/3</td>
<td>69.54</td>
<td>242</td>
<td>66.94</td>
<td>553</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>1/2</td>
<td>68.80</td>
<td>298</td>
<td>65.76</td>
<td>611</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>S\textsubscript{p}NF_LRR</td>
<td>1</td>
<td>64.07</td>
<td>31</td>
<td>63.13</td>
<td>48</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>2/3</td>
<td>67.13</td>
<td>31</td>
<td>66.74</td>
<td>48</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>1/2</td>
<td>66.69</td>
<td>32</td>
<td>65.76</td>
<td>50</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>L\textsubscript{p}	extsubscript{un} [74]</td>
<td>1</td>
<td>60.31</td>
<td>22</td>
<td>55.42</td>
<td>55</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CLR\textsubscript{L}	extsubscript{p} [75]</td>
<td>2</td>
<td>64.72</td>
<td>4</td>
<td>64.51</td>
<td>9</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th>Methods</th>
<th>p</th>
<th>MNIST ACC</th>
<th>Time</th>
<th>USPS ACC</th>
<th>Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>SSC [11]</td>
<td>1</td>
<td>68.00</td>
<td>213</td>
<td>74.30</td>
<td>120</td>
</tr>
<tr>
<td></td>
<td>2/3</td>
<td>68.90</td>
<td>136</td>
<td>78.30</td>
<td>123</td>
</tr>
<tr>
<td></td>
<td>1/2</td>
<td>68.10</td>
<td>156</td>
<td>78.90</td>
<td>126</td>
</tr>
<tr>
<td>S\textsubscript{p}NM_LRR</td>
<td>1</td>
<td>68.50</td>
<td>147</td>
<td>78.00</td>
<td>140</td>
</tr>
<tr>
<td></td>
<td>2/3</td>
<td>69.90</td>
<td>183</td>
<td>78.70</td>
<td>169</td>
</tr>
<tr>
<td></td>
<td>1/2</td>
<td>69.20</td>
<td>209</td>
<td>78.30</td>
<td>181</td>
</tr>
<tr>
<td>S\textsubscript{p}NF_LRR</td>
<td>1</td>
<td>66.40</td>
<td>20</td>
<td>70.90</td>
<td>18</td>
</tr>
<tr>
<td></td>
<td>2/3</td>
<td>68.90</td>
<td>25</td>
<td>78.20</td>
<td>21</td>
</tr>
<tr>
<td></td>
<td>1/2</td>
<td>67.50</td>
<td>28</td>
<td>76.50</td>
<td>23</td>
</tr>
<tr>
<td>L\textsubscript{p}	extsubscript{un} [74]</td>
<td>1</td>
<td>45.00</td>
<td>21</td>
<td>63.70</td>
<td>19</td>
</tr>
<tr>
<td>CLR\textsubscript{L}	extsubscript{p} [75]</td>
<td>2</td>
<td>55.70</td>
<td>5</td>
<td>76.70</td>
<td>5</td>
</tr>
</tbody>
</table>

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], $S\textsubscript{p}NF\_LRR$ 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 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 $S\textsubscript{p}NM\_LRR$ 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 $S\textsubscript{p}NM\_LRR$ and $S\textsubscript{p}NF\_LRR$ 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 $S\textsubscript{p}NM\_LRR$ and $S\textsubscript{p}NF\_LRR$, there exist the common parameter $\lambda$, and an additional ingredient for the expected rank number $d$ of the coefficient matrix in $S\textsubscript{p}NF\_LRR$. Thus, the discussions of $\lambda$ for $S\textsubscript{p}NM\_LRR$ and $(\lambda, d)$ for $S\textsubscript{p}NF\_LRR$ are given independently as follows.

For $S\textsubscript{p}NM\_LRR$, the choice of the parameter $\lambda$ is not unique on the above each of experiments. There exists a recent study in [77], which states that estimating $\lambda$ 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 $\lambda \in [\lambda_{\text{min}}, \lambda_{\text{max}}]$, where both $\lambda_{\text{min}}$ and $\lambda_{\text{max}}$ 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 $S\textsubscript{p}NM\_LRR$ with...
Fig. 10. Clustering accuracies are obtained by the $S_p$NF 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 $\lambda$ 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 $\lambda$. The results are insensitive to $\lambda$ 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 $S_p$NF LRR, we analyze the affects of the parameter pairs $(\lambda, d)$ for the clustering accuracy. The results are reported in Fig. 10(a)–(c) for various $(\lambda, d)$. Since different clustering data accords to different rank number of representation matrix, which has various effect on the clustering accuracy even for the same $\lambda$. Specially, $S_p$NF 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 $d$ ($\geq 80$), $S_p$NF 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 $S_p$NM LRR and $S_p$NF 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 $S_p$NM LRR and $S_p$NF 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


S. Burer and R. D. C. Monteiro, “Local minima and convergence in 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 *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.