Unified Framework for Faster Clustering via Joint Schatten p-Norm Factorization With Optimal Mean

Hengmin Zhang, *Member, IEEE*, Jiaoyan Zhao, Bob Zhang, *Senior Member, IEEE*, Chen Gong, *Senior Member, IEEE*, Jianjun Qian, *Member, IEEE*, and Jian Yang, *Member, IEEE*

***Abstract*— To enhance the effectiveness and efficiency of subspace clustering in visual tasks, this work introduces a novel approach that automatically eliminates the optimal mean, which is embedded in the subspace clustering framework of low-rank representation (LRR) methods, along with the computationally factored formulation of Schatten p-norm. By addressing the issues related to meaningful computations involved in some LRR methods and overcoming biased estimation of the low-rank solver, we propose faster nonconvex subspace clustering methods through joint Schatten p-norm factorization with optimal mean (JSpNFOM), forming a unified framework for enhancing performance while reducing time consumption. The proposed approach employs tractable and scalable factor techniques, which effectively address the disadvantages of higher computational complexity, particularly when dealing with large-scale coefficient matrices. The resulting nonconvex minimization problems are reformulated and further iteratively optimized by multivariate weighting algorithms, eliminating the need for singular value decomposition (SVD) computations in the developed iteration procedures. Moreover, each subproblem can be guaranteed to obtain the closed-form solver, respectively. The theoretical analyses of convergence properties and computational complexity further support the applicability of the proposed methods in real-world scenarios. Finally, comprehensive experimental results demonstrate the effectiveness and efficiency of the proposed nonconvex clustering approaches compared to existing state-of-the-art methods on several publicly available databases. The demonstrated improvements highlight the practical significance of our work in subspace clustering tasks for visual data analysis. The source code for the proposed algorithms is publicly accessible at https://github.com/ZhangHengMin/TRANSUFFC.**

***Index Terms*— Low-rank representation (LRR), matrix factorization, optimal mean, Schatten p-norm, subspace clustering.**

I. Introduction

In RECENT years, low-rank matrix learning has garnered significant interest in various domains, including pattern recognition, machine learning, and data mining [1], [2]. These approaches have found applications in diverse areas, such as robust process monitoring [3], thermography diagnosis systems [4], and manufacturing process monitoring [5]. To address these practical applications, researchers have made progress in both convex and nonconvex relaxation of the rank function [6], [7], [8], [9], [10], [11], [12] for collaborative filtering, image processing, face recognition, and motion segmentation. However, solving low-rank matrix learning problems using the discrete and nonconvex nature of the rank function is generally NP-hard. Consequently, methods have been developed to address the related convex envelope of the rank function, known as the nuclear norm of the matrix [13], [14], [15]. Nonetheless, optimization problems involving the nuclear norm can lead to biased solutions due to the uniform treatment of singular values.

To overcome this limitation, the nuclear norm has been generalized to the Schatten p-norm [11], [16], [17], [18], where $0 < p < 1$ yields a nonconvex problem that better approximates the rank function than the convex nuclear norm in most cases. Several nonconvex rank surrogates have been developed to address this point, including truncated/weighted nuclear norm [19], [20], [21], aiming to achieve better low-rank solutions. Some standard examples of low-rank matrix learning methods using these surrogates include efficient Schatten p-norm minimization (ESpNM) [22], bilinear robust principle analysis (BiRPCA) [16], logarithmic low-rank representation (LogLRR) [23], fast universal low rank representation (FULRR) [24], latent low-rank representation (LatLRR) [25], low-rank sparse subspace clustering (LRRSC) [26], nonnegative subspace clustering (NSC) [27], smoothed low-rank representation (SmLRR) [28], Schatten p-norm factorized low-rank representation (SpNFLRR) [29], as well as other spectral clustering methods without involving rank

Manuscript received 16 November 2022; revised 10 April 2023 and 9 August 2023; accepted 18 October 2023. This work was supported in part by the National Natural Science Fund (NSF) of China under Grant 61906067, Grant 62176124, Grant 61973162, and Grant 12371510; in part by the China Postdoctoral Science Foundation under Grant 2019M651415 and Grant 2020T130191; in part by the Fundamental Research Funds for the Central Universities under Grant 30918014108, Grant 30920032202, and Grant 30921013114; in part by the NSF of Jiangsu Province under Grant BZ2021013; in part by the NSF for Distinguished Young Scholar of Jiangsu Province under Grant BK20220080; and in part by the "111" Program under Grant B13022. *(Corresponding author: Bob Zhang.)*

Hengmin Zhang and Bob Zhang are with the PAMI Research Group, Department of Computer and Information Science, and the Centre for Artificial Intelligence and Robotics, Institute of Collaborative Innovation, University of Macau, Macau, China (e-mail: zhanghengmin@126.com; bobzhang@um.edu.mo).

Jiaoyan Zhao is with the School of Artificial Intelligence, Shenzhen Polytechnic University, Shenzhen 518055, China (e-mail: zhaojy@szu.edu.cn).

Chen Gong, Jianjun Qian, and Jian Yang are with the PCA Laboratory, Key Laboratory of Intelligent Perception and Systems for High-Dimensional Information, Ministry of Education, School of Computer Science and Engineering, Nanjing University of Science and Technology, Nanjing 210094, China (e-mail: chen.gong@njust.edu.cn; csjqian@njust.edu.cn; csjyang@njust.edu.cn).

This article has supplementary material provided by the authors and color versions of one or more figures available at https://doi.org/10.1109/TNNLS.2023.3327716.

Digital Object Identifier 10.1109/TNNLS.2023.3327716

TABLE I
SCHATTEN p-NORM AND THEIR FACTORED FORMULATIONS $\Phi_p(\mathbf{U},\mathbf{V})$ WITH THREE SELECTED p-VALUES, RESPECTIVELY

p-value	$\|\mathbf{Z}\|_{S_p}^p$	$\min_{\mathbf{Z}=\mathbf{U}\mathbf{V}^\top}\Phi_p(\mathbf{U},\mathbf{V})$
$p=1$	$\|\mathbf{Z}\|_*$ [12]	$\min_{\mathbf{Z}=\mathbf{U}\mathbf{V}^\top}\frac{1}{2}(\|\mathbf{U}\|_F^2+\|\mathbf{V}\|_F^2)$
$p=2/3$	$\|\mathbf{Z}\|_{\text{F/N}}^{2/3}$ [16]	$\min_{\mathbf{Z}=\mathbf{U}\mathbf{V}^\top}\frac{1}{3}(\|\mathbf{U}\|_F^2+2\|\mathbf{V}\|_*)$
$p=1/2$	$\|\mathbf{Z}\|_{\text{BiN}}^{1/2}$ [16]	$\min_{\mathbf{Z}=\mathbf{U}\mathbf{V}^\top}\frac{1}{2}(\|\mathbf{U}\|_*+\|\mathbf{V}\|_*)$

relaxation functions, such as sparse subspace clustering (SSC) [30] and structured graph learning (SGL) [31].

To the best of our knowledge, most low-rank matrix learning methods often involve solving rank-relaxed minimization problems, which can lead to higher computational complexity, particularly for large-scale matrix sizes. To address this issue, various factorization techniques have been proposed to replace the rank relaxation norms. For instance, the factored nuclear norm [12], [32] can be expressed as the sum of two Frobenius norms, and the factored Schatten p-norm [16] can be expressed as the sum of two nuclear norms when $p=1/2$, and one nuclear norm and one Frobenius norm when $p=2/3$. The corresponding representation formulas are provided in Table I, where $\|\mathbf{Z}\|_{S_p}^p$ with the three p-values are replaced by $\Phi_p(\mathbf{U},\mathbf{V})$. Without loss of generality, we assume that $\mathbf{Z}=\mathbf{U}\mathbf{V}^\top\in\mathbb{R}^{m\times n}$, where $\mathbf{U}\in\mathbb{R}^{m\times d}$, $\mathbf{V}\in\mathbb{R}^{n\times d}$, and $r<d\ll\min(m,n)$ is an upper bound of the rank number. Furthermore, we investigated the use of $\ell_{2,p}$-norm-based objective functions for the residual term, as discussed in [33] and [34]. In this work, the choice of p-values depends on the specific cases of interest, with a focus on commonly studied cases of $p\in\{1,2/3,1/2\}$. However, it is worth noting that other values of p can also be considered, as long as they satisfy $p\in(0,1)$, especially when dealing with nonconvex cases. To make further comparisons, Table II lists several existing works, such as optimal mean RPCA (OMRPCA) [35] and OM robust active representation (OMRAR) [34], which are designed to correct the incorrect mean calculation by automatically removing the optimal mean. These methods are useful in mitigating the effects of outliers, resulting in higher efficacy and efficiency compared to methodologies that do not incorporate them. Naturally, incorporating the mean calculation process in certain low-rank matrix learning problems, along with factorization strategies as presented in [24], [29], and [36], can boost performance and enhance efficiency.

A. Related Works

To provide a comprehensive understanding of our proposed method and its distinctions and connections to several existing works mentioned above, we will conduct an analysis and comparison with the most relevant methods involving Schatten p-norm, $\ell_{2,p}$-norm, ℓ_p-norm, and/or optimal mean, as listed in Table II. We will focus on four specific aspects, where the variables involved have compatible meanings, and their definitions and explanations can be found in the respective references. Below is a summary of existing methods, presented from three different perspectives.

1) *Variants of Low-Rank Representation (LRR) Methods:* These methods are specifically designed for unsupervised clustering tasks [7] and typically involve rank relaxed optimization problems. Various rank substitutes can lead to different spectral clustering models. Some factored Schatten p-norm matrix relaxations for $0<p\leq 1$ have been demonstrated to offer higher computational efficiency [5], [12], [16], [29], [36], [37] compared to the original low-rank matrix learning models. Additionally, to enhance performance, different types of information have been integrated into mathematical models [3], [4], [24], [25], [38]. Furthermore, non-factored solutions, such as those proposed in [22] and [28], do not require singular value decomposition (SVD) computations for singular value threshold operators, leading to improved computational efficiency. However, it is essential to note that prior research on low-rank matrix recovery problems has indicated that non-factored solutions generally exhibit lower efficiency compared to factorized solutions.
2) *Applications of Optimal Mean:* Several existing methods, such as OMRPCA, FULRR, and OMRAR [24], [34], [35], have incorporated the similar processing techniques for automatically removing optimal means, particularly for group sparse coding and low-rank matrix recovery problems. For example, OMRAR [34] establishes a relationship between optimal mean robust PCA and sample selection by mimicking the self-contained regressions of PCA studied for the optimal mean-centering problem. Moreover, it can use $\ell_{2,p}$-norm to regulate the group sparsity of the residual function in [17], [39], and [40]. Empirical results have validated the improved performance on single and multiple subspaces, further motivating us to introduce the concept of optimal mean to clustering methods.
3) *Optimizations of LRR Models:* Several first-order optimization algorithms are commonly used to optimize learning problems related to low-rank matrices. Among these algorithms, the alternating direction multiplier method (ADMM) and its linear versions [41], [42], [43], [44], accelerated proximal gradient (APG) [45], [46], [47], iteratively reweighted methods (IRwMs) [11], [48], [49], and proximal alternating linearized minimization (PALM) [50], [51] are widely applied and extensively studied. To provide more details, ADMM requires the introduction of Lagrange multiplier variables and a penalty parameter to formulate multivariate unconstrained problems, while both APG and IRwM typically optimize unconstrained single-variable minimization problems. The penalty function strategies [50], [52] can be incorporated to transform constrained problems into unconstrained ones. However, the SVD computations involved in the iterative process lead to higher computational complexity. Thus, developing efficient solvers can address this issue more effectively.

TABLE II
COMPARISON OF SEVERAL METHODS RELATED TO SCHATTEN p-NORM, $\ell_{2,p}$-NORM, ℓ_p-NORM, AND/OR OPTIMAL MEAN FROM VARIOUS ASPECTS

Methods	Objective Function	Factorization ?	Optimal Mean ?	Non-convexity ?	Non-SVD ?
LRR [7]	$\lambda\|\mathbf{Z}\|_* + \|\mathbf{D}-\mathbf{DZ}\|_{2,1}$	×	×	×	×
ESpNM [22]	$\lambda\|\mathbf{Z}\|_{S_p}^p + \|\mathbf{D}-\mathbf{Z}\circ\Omega\|_F^2$	×	×	✓	✓
LatLRR [25]	$\lambda_1\|\mathbf{Z}\|_* + \lambda_2\|\mathbf{Z}\|_* + \|\mathbf{D}-\mathbf{DZ}-\mathbf{LD}\|_1$	×	×	×	×
LRRSC [26]	$\lambda\|\mathbf{Z}\|_{S_0} + \|\mathbf{Z}\|_0 + \|\mathbf{D}-\mathbf{DZ}\|^2\|_F^2$	×	×	✓	✓
FULRR [24]	$\lambda\|\mathbf{Z}\|_{S_p}^p + \|\mathbf{D}-\mathbf{ZU}^\top\mathbf{D}-\mathbf{b1}^\top\|_{2,1}$	✓	✓	✓	×
SmLRR [28]	$\lambda\|\mathbf{Z}\|_{S_p}^p + \|\mathbf{D}-\mathbf{DZ}\|_{2,p}^p$	×	×	✓	✓
NSC [27]	$\lambda_1\|\mathbf{Z}\|_* + \lambda_2 tr(\mathbf{S}^\top\mathbf{E}) + \|\mathbf{D}-\mathbf{DS}^\top\|_F^2$	✓	✓	✓	×
LogLRR [23]	$\lambda\sum_{i=1}^{r} log(w\sigma_i^c+1) + \|\mathbf{D}-\mathbf{DQZ}\|_{2,1}$	✓	×	✓	×
OMRAR [34]	$\lambda\|\mathbf{Z}\|_{2,p} + \|(\mathbf{D}-\mathbf{DX}-\mathbf{b1}^\top)^\top\|_{2,p}$	×	✓	✓	✓
BiRPCA [16]	$\lambda\Phi_p(\mathbf{U},\mathbf{V}) + \|\mathbf{D}-\mathbf{Z}\circ\Omega\|_1$	✓	×	✓	×
SpNFLRR [29]	$\lambda\Phi_p(\mathbf{U},\mathbf{V}) + \|\mathbf{D}-\mathbf{DZ}\|_{2,1}$	✓	×	✓	×
OMRPCA [35]	$\lambda\|\mathbf{Z}\|_* + \|\mathbf{D}-\mathbf{Z}-\mathbf{b1}^\top\|_{2,1}$	×	✓	×	×
JSpNFOM	$\lambda\Phi_p(\mathbf{U},\mathbf{V}) + \|\mathbf{D}-\mathbf{DZ}-\mathbf{b1}^\top\|_{2,p}$	✓	✓	✓	✓

B. Main Contributions

In this work, our primary objective is to develop efficient subspace clustering methods by combining Schatten p-norm factorization and optimal mean into a unified framework, which provides a more accurate solution with lower computational complexity. Our work is mostly related to various important aspects, such as nonconvex low-rank matrix learning, unsupervised clustering learning, matrix factorization, and alternating iteration optimization. By providing this comprehensive overview, readers will have a better understanding of the diverse aspects our work addresses in the field of applying low-rank matrix recovery methods to subspace clustering applications. After carefully examining the statements presented above and making further comparisons, we differentiate this work more distinctly from others based on several aspects listed in Table II, including model formulation, algorithmic procedure, and critical techniques. The main strengths of our work lie in demonstrating the effectiveness of nonconvex representation models based on optimal mean and the efficiency of our optimization algorithms, which are distinct from existing LRR models and their solutions. In summary, we highlight the contributions of our work as follows.

1) To improve the accuracy of clustering, we propose a modified nonconvex LRR framework that combines Schatten p-norm with optimal mean. This framework becomes more general and unbiased in characterizing residual and coefficient terms due to its flexible p-choices and reliable mean calculation capability. Moreover, we directly factorize the Schatten p-norm, resulting in reduced computational complexity at each iteration by reformulating the multiplication of two smaller factor matrices.
2) To avoid the use of Lagrange multiplier variables such as ADMM in solving multivariate constraint problems, we adopt a penalized function strategy and develop an iteratively alternating reweighted splitting algorithm. Notably, our approach features closed-form solvers for all subproblems that do not require SVD computations during low-rank matrix recovery, setting it apart from other methods and allowing for faster convergence rates. Importantly, we also provide a comprehensive explanation of the convergence guarantees of our proposed optimization algorithm, which involves optimizing multiple variables. The derivation and analysis are now elaborated, establishing a solid theoretical foundation for our proposed methods in supporting practical applications. By employing these technologies into the unified framework, we facilitate efficient computations, reduce the computational complexity, and achieve faster convergence, enhancing the practical applicability of our methods.
3) Based on the experiments conducted on several real-world databases, along with theoretical verifications and comparisons of computational efficiency and clustering accuracy, we demonstrate the theoretical consistency and superior performance of our proposed method. We achieve this by comparing it with benchmark methods and the latest clustering works based on nonconvex and factorized low-rank matrix learning methods. The performance of our method is supported by both numerical and visual results, particularly in the context of unsupervised clustering tasks. Moreover, our unified framework allows for efficient and effective applications with further extensions to tensors, opening up possibilities for exploring additional information and inspiring other image low-level vision applications in this area.

C. Outline of This Article

The structure of this article is organized as follows. In Section II, we present our proposed efficient subspace clustering models, which are induced by combining Schatten

p-norm factorization with optimal mean, and we also provide a proposition for the computations of involved variables. Section III focuses on devising faster solutions to unconstrained nonconvex minimization problems and provides their computational complexity and convergence properties under milder assumptions. In Section IV, we analyze the performance and efficiency of the proposed algorithms through numerical and extensive experiments, offering both quantitative and qualitative comparisons. Finally, Section VI concludes this work and outlines future research directions. Additionally, for completeness, detailed proofs of the convergence properties and rates are provided in the Supplementary Materials.

D. Notations

The definitions of some symbols in this article are provided as follows: uppercase bold letters represent matrices, e.g., $\mathbf{X}$, and lowercase bold letters represent vectors, e.g., $\mathbf{x}$. SVD denotes the singular value decomposition of a matrix, and σ_i is the ith singular value. $\mathbf{b}$ represents an optimal mean vector, and $\mathbf{1}$ represents a column vector of ones. $\mathbf{I}$ is a diagonal matrix with all diagonal elements set to 1. $\mathbb{R}_r^{m\times n}$ denotes the set of $m \times n$ matrices with rank at most r, where d is the upper bound of the rank r, and $\text{rank}(\mathbf{X})$ is defined as the number of nonzero singular values of $\mathbf{X}$. The Frobenius norm of a matrix $\mathbf{X}$ is denoted as $\|\mathbf{X}\|_F^2 = \sum_i \|\mathbf{X}^i\|_2^2 = \sum_i \sigma_i^2$, the $\ell_{2,p}$-norm as $\|\mathbf{X}\|_{2,p}^p = \sum_i \|\mathbf{X}^i\|_2^p$, and the Schatten p-norm as $\|\mathbf{X}\|_{S_p}^p = \text{tr}(\mathbf{X}^\top\mathbf{X})^{(p/2)} = \sum_i \sigma_i^p$, where $\mathbf{X}^i$ is the ith column of $\mathbf{X}$. The $L_{2,1}$-norm and nuclear norm, denoted as $\|\mathbf{X}\|_{2,1}$ and $\|\mathbf{X}\|_*$, respectively, are special cases where $p = 1$. Furthermore, we define $\mathbf{W}_u$, $\mathbf{W}_v$, and $\mathbf{W}_e$ as weighted matrices related to $\mathbf{U}$, $\mathbf{V}$, and $\mathbf{E}$, respectively. These matrices will be introduced later in the iteration procedures.

II. Problem Formulations

Suppose that $\mathbf{D} \in \mathbb{R}^{m\times n}$ and $\mathbf{Z} \in \mathbb{R}_r^{n\times n}$ are the data matrix and coefficient matrix, respectively. Additionally, $\mathbf{E} \in \mathbb{R}^{m\times n}$ denotes the residual matrix. Based upon [7] and [50], the original LRR methods are primarily designed to address the regularized rank minimization problems, which can be expressed as

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

where the regularized parameter $\lambda > 0$ is used along with the residual term $\|\mathbf{E}\|_\ell$ to characterize various noise styles in the objective function. Although $\mathbf{A}$ can be replaced by $\mathbf{D}$, doing so may result in lower computational efficiency for large n. To address this issue, faster versions of LRR have been proposed. To this end, several related works replace the rank function with Schatten p-norm using $0 < p \leq 1$. For more detailed descriptions, please refer to the following explanations.

1) The discrete and nonconvex nature of the rank function makes problem (1) NP-hard [13], [53], [54]. To overcome this issue, nuclear norm and other nonconvex rank relaxations [7], [18], [23], [29], [33] have been widely used as substitutes for the rank function in subspace clustering tasks. These proper relaxations can lead to LRR variants' solutions that are closer to the true low-rank solution compared to the nuclear norm.
2) The traditional LRR methods center the data with the mean calculated using the ℓ_2-norm, which may not be accurate since ℓ_{21}-norm is used to measure the residual function [7], [42]. To address this, incorporating a term that automatically removes the optimal mean [17], [24], [35] can be beneficial in correcting the mean calculation.
3) The SVD computations are typically involved in optimizing problem (1) for achieving the coefficient matrix $\mathbf{Z}$ [24], [29], [36]. This can result in high computational complexity, especially when dealing with large-scale coefficient matrices.

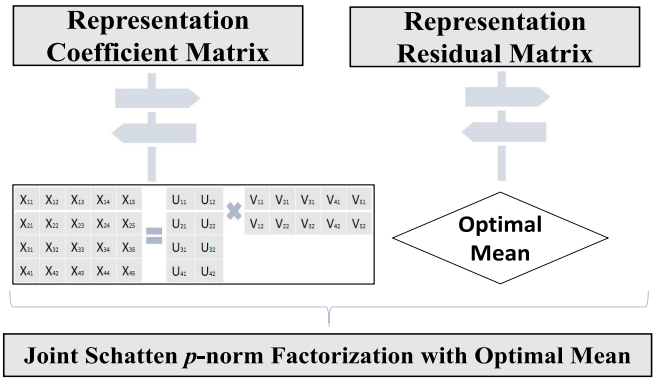


Fig. 1. Proposed model scheme is designed for subspace clustering and employs a computationally efficient factored formulation of Schatten p-norm, as well as automatic removal of optimal mean.

The statements above motivate us to improve the formulation of problem (1) for better ease of calculation, accuracy, and efficiency. To address these issues, we propose a more generalized LRR model that incorporates Schatten p-norm regularization of the $\ell_{2,p}$-norm and optimal mean, as previously proposed in [24], [29], and [34]. This model provides flexibility in the selection of p-values and is formulated as follows:

$$\min_{\mathbf{Z},\mathbf{E}} \lambda\|\mathbf{Z}\|_{S_p}^p + \|\mathbf{E}\|_{2,p}, \quad \text{s.t. } \mathbf{D} - \mathbf{AZ} - \mathbf{b}\mathbf{1}^\top = \mathbf{E} \tag{2}$$

where we subtract the optimal matrix $\mathbf{b}\mathbf{1}^\top$ from the constrained equation, which improves its accuracy. When $p = 1$ and $\mathbf{b} = 0$, Problem (2) reduces to standard single and multiple subspace methods, such as RPCA [55] when $\mathbf{A}$ is an identity matrix, or LRR [7] when $\mathbf{A}$ is a dictionary matrix or substituted by a data matrix. However, optimizing Problem (2) can be computationally intensive, particularly when it involves SVD computation for large-scale coefficient matrices.

To address the third problem mentioned above, we approximate $\|\mathbf{Z}\|_{S_p}^p$ by $\Phi_p(\mathbf{U},\mathbf{V})$ using the bilinear factor matrices presented in Table I. By combining this technique with the optimal mean for the coefficients and residual matrices, as outlined in Fig. 1, we can achieve a more effective and efficient clustering method. In the following, we present a factorized and improved formulation of the LRR problem:

$$\begin{aligned}\min_{\mathbf{U},\mathbf{V},\mathbf{Z},\mathbf{b},\mathbf{E}} & f_{\lambda,p}(\mathbf{U},\mathbf{V},\mathbf{E}) = \lambda\Phi_p(\mathbf{U},\mathbf{V}) + \|\mathbf{E}\|_{2,p} \\ \text{s.t. } & \mathbf{Z} = \mathbf{U}\mathbf{V}^\top, \quad \mathbf{D} - \mathbf{AZ} - \mathbf{b}\mathbf{1}^\top = \mathbf{E}\end{aligned} \tag{3}$$

where we aim to solve the optimization problem (3) efficiently without introducing Lagrange multiplier variables such as ADMM in the augmented Lagrange function to solve (3) indirectly. The optimization problems presented in [29] are specific instances of (3) with different p-values and $\mathbf{b} = 0$. Here, we have observed the need for a more comprehensive discussion of the costs and constraints related to implementing complexity reduction. The related discussion and analysis can be detailed from the following two aspects.

1) On the one hand, the adoption of complexity reduction strategies necessitates matrix factorization, such as $\mathbf{Z} = \mathbf{U}\mathbf{V}^\top$. This motivates us to utilize the widely studied Schatten p-norm factorization, as other nonconvex rank relaxation functions may lack readily available factorization formulas. The existence of closed-form or analytic solutions for the involved subproblems facilitates efficient computation, as demonstrated in Proposition 1 and [10], [18], and [29]. Therefore, we stress that both factorization formulas and closed-form/analytic solutions are critical prerequisites for the successful implementation of complexity reduction strategies.
2) On the other hand, adopting the factorization strategy for low-rank matrices introduces multiple variables, making the theoretical analysis of convergence guarantees challenging, as previously noted in the work [44]. In contrast to commonly used methods like ADMM, we address this challenge by employing a primal iteration scheme and providing theoretical support. Importantly, dealing with multiple variables naturally increases the computational burden. However, it is worth noting that the size of the factor matrices is much smaller compared to the learned low-rank matrix, which is very essential and necessary for reducing computational complexity.

To address the above twofold concerns, we need to transform (3) into an unconstrained problem, represented as

$$\begin{aligned}&\min_{\mathbf{U},\mathbf{V},\mathbf{Z},\mathbf{b},\mathbf{E}} \mathcal{H}^p_{\lambda,\mu}(\mathbf{U},\mathbf{V},\mathbf{Z},\mathbf{b},\mathbf{E})\\ &= f_{\lambda,p}(\mathbf{U},\mathbf{V},\mathbf{E})\\ &\quad + \frac{\mu}{2}\left(\left\|\mathbf{Z}-\mathbf{U}\mathbf{V}^\top\right\|_F^2 + \left\|\mathbf{D}-\mathbf{A}\mathbf{Z}-\mathbf{b}\mathbf{1}^\top-\mathbf{E}\right\|_F^2\right) \quad (4)\end{aligned}$$

where we introduce the parameter $\mu > 0$ using a continuation strategy [56], [57]. We initialize μ with a small value μ^0 and update it at each iteration by multiplying it with ρ (i.e., $\mu^k = \mu^0\rho^k$, where k is the number of iterations). As μ increases, the solution of (4) converges to (3), and the penalty function's value approaches zero. The step-size parameter ρ plays a critical role in balancing the convergence rate and clustering performance. Larger ρ values lead to faster convergence but may compromise accuracy, while smaller values will slow down convergence. Based upon suggestions in the literature, we empirically choose values of $\rho = \{1.1, 1.2, \ldots, 1.8, 1.9\}$ and initial values of μ^0 from the range $\{10^{-6}, 10^{-5}, \ldots, 10^{1}, 10^{2}\}$. Through experimental comparisons, we identify optimal values for both ρ and μ^0 that yield the best performance. Regarding the choice of the penalty function, one can assume that the residual functions $\mathbf{Z}-\mathbf{U}\mathbf{V}^\top$ and $\mathbf{D}-\mathbf{A}\mathbf{Z}-\mathbf{b}\mathbf{1}^\top-\mathbf{E}$ follow either Gaussian or Laplace distributions. If the residuals follow a Gaussian distribution, then measuring them by the squared Frobenius norm is appropriate and easily optimized. However, if they follow a Laplace distribution, a better option would be to measure the noise using the ℓ_1-norm. Then, we specifically focus on the scenario where the residuals follow a Gaussian distribution, as it has been studied in [33], [58], and [59].

Notably, when solving subproblems based on nuclear norm or nonconvex rank relaxation functions, it is common practice to use singular value thresholding (SVT) operators. The same approach can be applied to the proposed problem (4) when $p = 1$, $2/3$, and $1/2$. However, involving SVD computations in the iteration procedures can lead to higher computational complexity. To avoid this limitation, one can employ equivalent model formulations and establish rigorous connections, as demonstrated in the following proposition, as outlined in [34], [37], and [60]. This allows us to optimize the problem without the need for costly SVD computations.

Proposition 1: Observed from Table I that $\|\mathbf{U}\|_*$ only exists for $1/2$, $\|\mathbf{V}\|_*$ exists for $p = 2/3$ and $1/2$, and $\|\mathbf{E}\|_{2,p}$ exists for the three p-values, i.e., 1, $2/3$ and $1/2$, then $f_{\lambda,p}(\mathbf{U},\mathbf{V},\mathbf{E})$ of both (3) and (4) can be replaced by:

1) $(\lambda/2)(\|\mathbf{U}\|_F^2 + \|\mathbf{V}\|_F^2) + \|\mathbf{E}\sqrt{\mathbf{W}_e}\|_F^2$;
2) $(\lambda/3)(\|\mathbf{U}\|_F^2 + 2\mathrm{tr}(\mathbf{V}\mathbf{W}_v\mathbf{V}^\top)) + \|\mathbf{E}\sqrt{\mathbf{W}_e}\|_F^2$;
3) $(\lambda/2)(\mathrm{tr}(\mathbf{U}\mathbf{W}_u\mathbf{U}^\top) + \mathrm{tr}(\mathbf{V}\mathbf{W}_v\mathbf{V}^\top)) + \|\mathbf{E}\sqrt{\mathbf{W}_e}\|_F^2$,

where both $\mathbf{W}_u = (1/2)\mathrm{tr}(\mathbf{U}^\top\mathbf{U})^{-(1/2)}$ and $\mathbf{W}_v = (1/2)\mathrm{tr}(\mathbf{V}^\top\mathbf{V})^{-(1/2)}$ have similar representation formulas. Meanwhile, we compute $\mathbf{W}_e$ by the diagonal matrix according to

$$\begin{bmatrix} \frac{p}{2}\|\mathbf{e}_1\|_2^{p-2} & & & \\ & \frac{p}{2}\|\mathbf{e}_2\|_2^{p-2} & & \\ & & \ddots & \\ & & & \frac{p}{2}\|\mathbf{e}_n\|_2^{p-2} \end{bmatrix} \quad (5)$$

where $\mathbf{e}_i$ $(1 \le i \le n)$ is the ith column vector of $\mathbf{E}$, and the elements of non-diagonal locations are all zeros.

It is worth mentioning that Proposition 1 plays a crucial role in updating the factor and error variables. Below are some illustrations from two different perspectives.

1) Proposition 1 is derived from equivalent formulations of the nuclear norm and $\ell_{2,p}$-norm. It focuses on matrix-level computations during the iterative process by introducing weighting matrices $\mathbf{W}_u$, $\mathbf{W}_v$, and $\mathbf{W}_e$, instead of using threshold operators to achieve the desired variables.
2) The proposed optimization algorithm involves repeated matrix inversion and multiplication during iterative procedures. This work employs a weighted strategy to transform the objective functions under three various p-values and obtain the factor matrices.

III. Optimization Scheme

In this section, we will provide the entire iterative procedure of (4) for three p-values, respectively. Following the updated rules in [50], [52], and [59], we can alternately optimize the

subproblems related to each of the variables in $(\mathbf{U}, \mathbf{V}, \mathbf{Z}, \mathbf{b}, \mathbf{E})$, where we aim to obtain $\mathbf{Z}^*$ for constructing the affinity matrix. Given the kth iteration variables, i.e., $(\mathbf{U}^k, \mathbf{V}^k, \mathbf{Z}^k, \mathbf{b}^k, \mathbf{E}^k)$, the $(k+1)$th iteration variables can be updated by

$$\begin{cases} \mathbf{U}^{k+1} = \operatorname{argmin}_{\mathbf{U}} \lambda \Phi_p(\mathbf{U}, \mathbf{V}^k) + \frac{\mu}{2} \left\| \mathbf{Z}^k - \mathbf{U}(\mathbf{V}^k)^\top \right\|_F^2 & (6) \\ \mathbf{V}^{k+1} = \operatorname{argmin}_{\mathbf{V}} \lambda \Phi_p(\mathbf{U}^{k+1}, \mathbf{V}) \\ \qquad + \frac{\mu}{2} \left\| \mathbf{Z}^k - \mathbf{U}^{k+1}\mathbf{V}^\top \right\|^2 & (7) \\ \mathbf{Z}^{k+1} = \operatorname{argmin}_{\mathbf{Z}} \left\| \mathbf{Z} - \mathbf{U}^{k+1}(\mathbf{V}^{k+1})^\top \right\|^2 + g(\mathbf{Z}, \mathbf{b}^k, \mathbf{E}^k) & (8) \\ \mathbf{b}^{k+1} = \operatorname{argmin}_{\mathbf{b}} g(\mathbf{Z}^{k+1}, \mathbf{b}, \mathbf{E}^k) & (9) \\ \mathbf{E}^{k+1} = \operatorname{argmin}_{\mathbf{E}} \|\mathbf{E}\|_{2,p} + \frac{\mu}{2} g(\mathbf{Z}^{k+1}, \mathbf{b}^{k+1}, \mathbf{E}) & (10) \end{cases}$$

where we define the function $g(\mathbf{Z}, \mathbf{b}, \mathbf{E}) = \|\mathbf{D} - \mathbf{A}\mathbf{Z} - \mathbf{b}\mathbf{1}^\top - \mathbf{E}\|_F^2$ to relate the computations of the involved variables. The closed-form solvers of sub-problems (6)–(10) can be computed using the formulas of both Table I and Proposition 1. Naturally, different p-values lead to the various solver formulations for $\mathbf{U}^{k+1}$ and $\mathbf{V}^{k+1}$ along with some modifications for updating the variables to the nuclear norm and $\ell_{2,p}$-norm. Among them, we have some formulas presented below.

1) *Updating* $\mathbf{U}^{k+1}$ *in (6):* Fixing the kth iteration variables and omitting unrelated variables with $\mathbf{U}$ in $\Phi_p(\mathbf{U}, \mathbf{V}^k)$ for various p-values provided in Table I, we can choose the function formulas as

$$\Phi_p'(\mathbf{U}) \in \left\{ \frac{1}{2}\|\mathbf{U}\|_F^2, \frac{1}{3}\|\mathbf{U}\|_F^2, \frac{1}{2}\|\mathbf{U}\|_* \right\} \tag{11}$$

where we will give the closed-form solvers with respect to the chosen p-values. The detailed steps are provided as below.

a) When $p = 1$ and $2/3$, $\mathbf{U}^{k+1}$ can be denoted as

$$\mathbf{U}^{k+1} = \mathbf{Z}^k\mathbf{V}^k\left(\frac{\lambda}{\mu}\mathbf{I} + (\mathbf{V}^k)^\top\mathbf{V}^k\right)^{-1} \tag{12}$$

$$\mathbf{U}^{k+1} = \mathbf{Z}^k\mathbf{V}^k\left(\frac{2\lambda\mathbf{I}}{3\mu} + (\mathbf{V}^k)^\top\mathbf{V}^k\right)^{-1}. \tag{13}$$

b) When $p = 1/2$, it follows from Proposition 1 that $\mathbf{U}^{k+1}$ can be computed by solving the minimization problem:

$$\min_{\mathbf{U}} \lambda \operatorname{tr}(\mathbf{U}\mathbf{W}_u^k\mathbf{U}^\top) + \mu \operatorname{tr}\left(\left(\mathbf{Z}^k - \mathbf{U}(\mathbf{V}^k)^\top\right)^\top\left(\mathbf{Z}^k - \mathbf{U}(\mathbf{V}^k)^\top\right)\right) \tag{14}$$

where $\mathbf{W}_u^k = (1/2)((\mathbf{U}^k)^\top\mathbf{U}^k)^{-(1/2)}$ is defined but relied on $\mathbf{U}^k$. Computing the derivative of (14) with respect to $\mathbf{U}$ and let it be $\mathbf{0}$, it is easy to obtain

$$\mathbf{U}^{k+1} = \mathbf{Z}^k\mathbf{V}^k\left(\frac{\lambda}{\mu}\mathbf{W}_u^k + (\mathbf{V}^k)^\top\mathbf{V}^k\right)^{-1}. \tag{15}$$

2) *Updating* $\mathbf{V}^{k+1}$ *in (7):* Similar to the processing way used in a) at first, we here redefine the proper formulas of $\Phi_p(\mathbf{U}^{k+1}, \mathbf{V})$ followed by:

$$\Phi_p'(\mathbf{V}) \in \left\{ \frac{1}{2}\|\mathbf{V}\|_F^2, \frac{2}{3}\|\mathbf{V}\|_*, \frac{1}{2}\|\mathbf{V}\|_* \right\}. \tag{16}$$

a) When $p = 1$, $\mathbf{V}^{k+1}$ can be denoted as

$$\mathbf{V}^{k+1} = (\mathbf{Z}^k)^\top\mathbf{U}^{k+1}\left(\frac{\lambda}{\mu}\mathbf{I} + (\mathbf{U}^{k+1})^\top\mathbf{U}^{k+1}\right)^{-1}. \tag{17}$$

b) When $p = 2/3$ and $p = 1/2$, combining (14) with (7), $\mathbf{V}^{k+1}$ can be, respectively, computed as

$$\mathbf{V}^{k+1} = (\mathbf{Z}^k)^\top\mathbf{U}^{k+1}\left(\frac{4\lambda\mathbf{W}_v^k}{3\mu} + (\mathbf{U}^{k+1})^\top\mathbf{U}^{k+1}\right)^{-1} \tag{18}$$

$$\mathbf{V}^{k+1} = (\mathbf{Z}^k)^\top\mathbf{U}^{k+1}\left(\frac{\lambda}{\mu}\mathbf{W}_v^k + (\mathbf{U}^{k+1})^\top\mathbf{U}^{k+1}\right)^{-1} \tag{19}$$

where $\mathbf{W}_v^k = (1/2)((\mathbf{V}^k)^\top\mathbf{V}^k)^{-(1/2)}$. Following from both (14)–(16), the differentiable property is employed again to guarantee the analytic solvers directly.

Algorithm 1 Solutions of Problem (4)

Input: $\mathbf{D}$, d, $p \in \{1, 2/3, 1/2\}$, λ, and ρ.
Initialize: $\mathbf{U} = \text{rand}(m, d)$, $\mathbf{V} = \text{rand}(n, d)$, $\mathbf{Z}^0 = \mathbf{U}\mathbf{V}^\top$, $\mathbf{b}^0 = \mathbf{0}$, $\mathbf{E}^0 = \mathbf{0}$, and μ^0.
Output: $\mathbf{Z}^* \leftarrow \mathbf{Z}^{k+1}$.

```
Let k = 0.
while not converged do
    ♯1: update U^{k+1} and V^{k+1}
    if p = 1
        Update U^{k+1} by (12);
        Update V^{k+1} by (17);
    elseif p = 2/3
        Update U^{k+1} by (13);
        Update W_v^k = (1/2)((V^k)^⊤V^k)^{-1/2};
        Update V^{k+1} by (18);
    else
        Update W_u^k = (1/2)((U^k)^⊤U^k)^{-1/2};
        Update W_v^k = (1/2)((V^k)^⊤V^k)^{-1/2};
        Update U^{k+1} by (15);
        Update V^{k+1} by (19);
    end
    ♯2: update Z^{k+1} by (20);
    ♯3: update b^{k+1} by (21);
    ♯4: update W_e^k by (5);
    ♯5: update E^{k+1} by (23);
    ♯6: update μ^{k+1} by min(ρμ^k, 10^8);
    k ← k + 1
end while
```

3) *Updating* $\mathbf{Z}^{k+1}$ *in (8):* By computing the objective function of (8) with $\mathbf{Z}$ and set it to be $\mathbf{0}$, we have the simplified representation formula

$$\mathbf{Z}^{k+1} = \left(\mathbf{I} + \mathbf{A}^\top\mathbf{A}\right)^{-1}\left(\varphi_{k+1} + \mathbf{A}^\top\psi_k\right) \tag{20}$$

where $\varphi_{k+1} = \mathbf{U}^{k+1}(\mathbf{V}^{k+1})^\top$ and $\psi_k = \mathbf{D} - \mathbf{b}^k\mathbf{1}^\top - \mathbf{E}^k$.

TABLE III
FORMULAS OF DERIVATIVES RELATED WITH CONVEX AND NONCONVEX NORMS. NOTE THAT $f(\mathbf{Z},\mathbf{b},\mathbf{E})=\mathbf{AZ}+\mathbf{b1}^\top+\mathbf{E}$

p-values	$\nabla\Phi_p'(\mathbf{U})$	$\nabla\Phi_p'(\mathbf{V})$	$\nabla\|\mathbf{E}\|_{2,p}$	Derivative	Formulas
1	$\mathbf{U}$	$\mathbf{V}$	$2\mathbf{EW}_e$	$\nabla g(\cdot)\|_{\mathbf{Z}}$	$2\mathbf{A}^\top(f(\mathbf{Z},\mathbf{b},\mathbf{E})-\mathbf{D})$
2/3	$\frac{2}{3}\mathbf{U}$	$\frac{4}{3}\mathbf{VW}_v$	$2\mathbf{EW}_e$	$\nabla g(\cdot)\|_{\mathbf{b}}$	$2(f(\mathbf{Z},\mathbf{b},\mathbf{E})-\mathbf{D})\mathbf{1}$
1/2	$\mathbf{UW}_u$	$\mathbf{VW}_v$	$2\mathbf{EW}_e$	$\nabla g(\cdot)\|_{\mathbf{E}}$	$2(f(\mathbf{Z},\mathbf{b},\mathbf{E})-\mathbf{D})$

4) *Updating* $\mathbf{b}^{k+1}$ *in (9):* We here let the derivative of $g(\mathbf{Z}^{k+1},\mathbf{b},\mathbf{E}^k)$ with respect to $\mathbf{b}$ be 0, we can get

$$\mathbf{b}^{k+1}=\left(\mathbf{D}-\mathbf{AZ}^{k+1}-\mathbf{E}^k\right)\mathbf{1}\left(\mathbf{1}^\top\mathbf{1}\right)^{-1}. \tag{21}$$

5) *Updating* $\mathbf{E}^{k+1}$ *in (10):* It can be concluded from (10) together with Proposition 1 that we need to optimize the minimization problem followed by

$$\min_{\mathbf{E}}\left\|\mathbf{E}\sqrt{\mathbf{W}_e^k}\right\|^2+\frac{\mu}{2}\left\|\mathbf{E}-\left(\mathbf{D}-\mathbf{AZ}^{k+1}-\mathbf{b}^{k+1}\mathbf{1}^\top\right)\right\|_F^2 \tag{22}$$

where $\mathbf{W}_e^k$ is a diagonal matrix derived from $\mathbf{E}^k$, and we use $\mathbf{e}_i^k$ to substitute $\mathbf{e}_i$ of (5). Then, we have the following solver formula, represented as

$$\mathbf{E}^{k+1}=\left(\mathbf{D}-\mathbf{AZ}^{k+1}-\mathbf{b}^{k+1}\mathbf{1}^\top\right)\left(\frac{2\mathbf{W}_e^k}{\mu}+\mathbf{I}\right)^{-1}. \tag{23}$$

Based upon 1)–5) given above, we can summarize the iterative rules of the optimization scheme in Algorithm 1. Only the $\mathbf{U}$ and $\mathbf{V}$ variables need to be updated according to various p-values. The other variables remain fixed, as shown above. The proposed algorithm can be considered as an unconstrained extension of [29], with the optimal mean incorporated. After updating all the variables, the iterations will be repeated until the stopping criteria are met, i.e.,

$$\left\|g\left(\mathbf{Z}^{k+1},\mathbf{b}^{k+1},\mathbf{E}^{k+1}\right)\right\|_\infty<\epsilon \tag{24}$$

where we empirically set $\epsilon=1e^{-8}$ in (24) or the number of maximum iterations is reached before stopping.

IV. THEORETICAL ANALYSIS

In this section, we next present the computational complexity and convergence properties of Algorithm 1, which rely on the updated rules and properties of the objective function. The analysis is presented in the following manner.

1) For the computational complexity of updating $\mathbf{U}$, $\mathbf{V}$, $\mathbf{Z}$, $\mathbf{b}$, and $\mathbf{E}$, they mainly depend on the values of p and are, respectively, given as follows.
   a) The computational complexity of updating $\mathbf{U}$ is $o(2d^3+n^2d+3nd^2)$ for $p=1/2$, and $o(d^3+n^2d+2nd^2)$ for both $p=1$ and $2/3$. The computational complexity of updating $\mathbf{V}$ is $o(d^3+n^2d+2nd^2)$ for both $p=1$, and $o(2d^3+n^2d+3nd^2)$ for both $p=1/2$ and $2/3$. Here, given that $d\ll\min\{m,n\}$, the above-computed complexity can omit $o(d^3)$.
   b) The computational complexity of updating $\mathbf{Z}$, $\mathbf{b}$, and $\mathbf{E}$ is $o(2mn^2+2n^3+n^2d+mn)$, $o(mn^2+mn+n)$, and $o(2mn^2+mn+n^3)$, respectively.
2) For the convergence analysis, we next present a sketch of the evidence from three aspects.
   a) First, we prove the non-increasing property of the objective function provided by the problem (4).
   b) Second, we prove that the change of variables in adjacent iterations tend to be zero.
   c) Third, we prove that any accumulation point $(\mathbf{U}^*,\mathbf{V}^*,\mathbf{Z}^*,\mathbf{b}^*,\mathbf{E}^*)$ of $\{\mathbf{U}^k,\mathbf{V}^k,\mathbf{Z}^k,\mathbf{b}^k,\mathbf{E}^k\}$ is a stationary point of problem (4).

Subsequently, we present the first-order optimality conditions, along with the related subproblems in Algorithm 1, which will be used repeatedly. We then use (4) and (6)–(10) to compute the (sub)derivative of $\mathcal{H}^p_{\lambda,\mu}(\mathbf{U},\mathbf{V},\mathbf{Z},\mathbf{b},\mathbf{E})$ with respect to each variable, leading to representation formulas. Using the fact that $\mathbf{U}^{k+1}$, $\mathbf{V}^{k+1}$, $\mathbf{Z}^{k+1}$, $\mathbf{b}^{k+1}$, and $\mathbf{E}^{k+1}$ are the

Fig. 2. Partially selected images from various experimental databases, including AR, ExtYaleB, COIL10, USPS, MNIST, FLAVIA, and Hopkins 155, are shown from top to bottom rows.

TABLE IV
STATISTICS OF SEVERAL IMAGE DATABASES AND HOPKINS 155 DATABASE USED IN THE EXPERIMENTS

Databases	Data Size	Sample Size	Class Number	Type
AR	2580×700	60×43	100	face
ExtYaleB	2016×640	48×42	10	face
COIL10	1024×720	32×32	10	object
USPS	256×1000	16×16	10	digital
MNIST	784×2000	28×28	10	digital
FLAVIA	1200×1907	30×40	32	object
//	Motions	Points	Frames	Sequences
Hopkins155	2	266	30	120
	3	398	29	35

TABLE V
COMPARISON OF CLUSTERING EFFICACY AND EFFICIENCY AMONG VARIOUS METHODS FOR THE AR, EXTYALEB, AND COIL10 DATABASES

Method	AR				ExtYaleB				COIL10			
	ACC	NMI	TIME	ITER	ACC	NMI	TIME	ITER	ACC	NMI	TIME	ITER
SSC	39.57	65.39	16s	162	47.66	39.60	31s	424	60.28	74.99	10s	158
LRR	61.71	81.99	57s	232	80.63	82.39	37s	197	65.97	76.39	35s	175
LRRSC	63.43	79.26	16s	73	70.78	63.62	17s	101	67.92	72.20	16s	70
FULRR	64.57	82.49	101s	233	74.38	73.77	73s	220	72.08	77.50	71s	234
NSC	52.86	76.03	22s	471	72.35	74.86	20s	587	64.72	73.73	11s	248
SGL	61.57	79.54	138s	148	81.56	84.49	61s	218	53.06	61.44	28s	82
LogLRR	65.29	80.43	100s	254	76.09	77.92	75s	245	65.56	76.37	82s	245
LatLRR	67.71	86.18	62s	75	84.22	87.62	45s	75	78.61	82.39	39s	75
IRLS	60.29	80.95	90s	74	81.09	82.76	66s	64	65.97	75.59	91s	68
ARM	43.14	69.42	11s	26	55.31	46.25	8s	26	65.83	77.74	8s	27
S_1NFLRR	61.57	78.82	20s	178	80.94	81.89	5s	69	66.94	68.67	7s	123
$S_{1/2}$NFLRR	68.14	83.97	21s	169	87.19	81.24	8s	97	73.61	75.95	6s	91
$S_{2/3}$NFLRR	68.43	79.86	20s	167	87.82	82.04	8s	89	74.17	79.62	5s	69
JS_1NFOM	62.71	81.15	11s	73	82.50	85.65	9s	74	79.17	86.93	6s	71
$JS_{1/2}$NFOM	68.29	**86.46**	11s	74	90.13	86.98	8s	75	83.06	**90.92**	6s	73
$JS_{2/3}$NFOM	**69.00**	85.13	12s	74	**92.97**	**88.63**	8s	75	**83.33**	90.75	6s	71

minimized solutions, then we have

$$\begin{cases} \mathbf{0} = \nabla\lambda\Phi_p'(\mathbf{U})\big|_{\mathbf{U}=\mathbf{U}^{k+1}} - \mu\Big(\mathbf{Z}^k - \mathbf{U}^{k+1}\big(\mathbf{V}^k\big)^\top\Big)\mathbf{V}^k & (25) \\ \mathbf{0} = \nabla\lambda\Phi_p'(\mathbf{V})\big|_{\mathbf{V}=\mathbf{V}^{k+1}} - \mu\Big(\mathbf{Z}^k - \mathbf{U}^{k+1}\big(\mathbf{V}^{k+1}\big)^\top\Big)^\top\mathbf{U}^{k+1} & (26) \\ \mathbf{0} = \mathbf{Z}^{k+1} - \mathbf{U}^{k+1}\big(\mathbf{V}^{k+1}\big)^\top + \nabla g\big(\mathbf{Z}, \mathbf{b}^k, \mathbf{E}^k\big)\big|_{\mathbf{Z}=\mathbf{Z}^{k+1}} & (27) \\ \mathbf{0} = \nabla g\big(\mathbf{Z}^{k+1}, \mathbf{b}, \mathbf{E}^k\big)\big|_{\mathbf{b}=\mathbf{b}^{k+1}} & (28) \\ \mathbf{0} = \nabla\|\mathbf{E}\|_{2,p}\big|_{\mathbf{E}=\mathbf{E}^{k+1}} + \mu\nabla g\big(\mathbf{Z}^{k+1}, \mathbf{b}^{k+1}, \mathbf{E}\big)\big|_{\mathbf{E}=\mathbf{E}^{k+1}} & (29) \end{cases}$$

where in (11) and (16), $\Phi_p'(\mathbf{U})$ and $\Phi_p'(\mathbf{V})$ are both defined, respectively, and ∇ represents the derivative of differentiable functions. Proposition 1 shows that each of the subproblems is differentiable, and we omit the existence of the penalty parameter μ in (27) and (28), as it has no effect on the satisfaction of "=." The derivatives of the differentiable function $g(\mathbf{Z}, \mathbf{b}, \mathbf{E})$ with respect to each variable can be computed easily. Some concrete derivatives are provided in Table III, which are used in algorithmic designs and convergence analysis. The theoretical proofs are provided in the Supplementary Materials.

V. EXPERIMENTAL RESULTS

In this section, we mainly present experimental validations to demonstrate the computational efficiency and clustering accuracy of our proposed joint Schatten p-norm factorization with optimal mean (JSpNFOM) methods with three selected p-values: $p \in \{1, 1/2, 2/3\}$. We conduct these experiments on image databases and the Hopkins155 database for comprehensive evaluation. Fig. 2 displays partial images from seven databases, and their statistics are listed in Table IV. These datasets are subsets of the original databases, and we provide brief descriptions of each below for reference.

1) Two face databases, AR and ExtYaleB, are used for the clustering tasks. These databases contain images with variations in illuminations, occlusions, expressions, etc.
2) Two object databases, COIL10 and FLAVIA, are used to evaluate the performance of the involved methods, similar to other image databases.
3) Two digital databases, USPS and MNIST, are employed to validate comparisons, with the numbers ranging from "0" to "9," respectively.
4) The Hopkins155 database serves as a widely used benchmark for motion segmentation. It consists of 155 video sequences, each containing two or three motions corresponding to two or three low-dimensional subspaces.

To conduct the evaluations, we selected several databases, with all of them being subsets of the original databases, except for the Hopkins155 database. To categorize the selected subsets based on their sample dimensions, we divided them into two groups: one with larger sample dimensions than sample numbers and the other with smaller sample dimensions than sample numbers. These subsets are commonly used for testing subspace clustering methods. We then compared the involved methods across three classes, as described below.

1) *Traditional Clustering Methods:* This class includes traditional clustering methods such as SSC [30] and LRR [7], as well as their improved variants like LatLRR [25] and LRRSC [26], all of which were incorporated to enhance clustering performance.
2) *Nonconvex Clustering Methods:* We considered methods optimized using IRLS [28], which extends LRR based on Schatten p-norm and $\ell_{2,p}$-norm, as well as ARM [61], a nonconvex clustering method based on the arctangent rank function. We also compared the recently proposed clustering methods, such as NSC [27] and SGL [31].
3) *Factorization Methods:* In this class, we compared the recently proposed factorization methods, including FULRR [24] and LogLRR [23]. Additionally, we considered previously proposed methods like SpNFLRR [29]

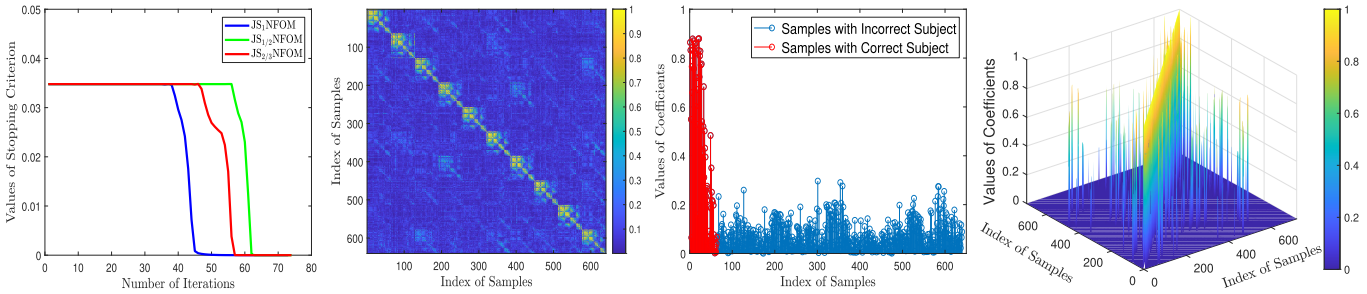


Fig. 3. Visual results of the proposed methods on the image databases are shown in (a) convergence curves of stopping criteria for AR, (b) block-diagonal structures, and (c) selected column vector of the coefficient matrix for ExtYaleB. (d) 3-D surface plot of the coefficient matrix for COIL10.

TABLE VI
COMPARISON OF CLUSTERING EFFICACY AND EFFICIENCY AMONG VARIOUS METHODS FOR THE USPS, MNIST, AND FLAVIA DATABASES

Method	USPS				MNIST				FLAVIA			
	ACC	NMI	TIME	ITER	ACC	NMI	TIME	ITER	ACC	NMI	TIME	ITER
SSC	52.40	64.21	29s	447	57.60	59.57	161s	411	51.23	65.76	150s	346
LRR	74.90	63.62	75s	225	56.35	58.98	811s	224	51.81	65.02	538s	165
LRRSC	38.30	27.65	35s	68	27.50	18.56	459s	101	44.20	59.17	282s	72
FULRR	64.40	62.63	13s	205	54.30	58.24	100s	213	52.54	65.28	328s	223
NSC	79.30	75.94	36s	335	65.75	69.98	85s	138	51.92	57.60	297s	475
SGL	76.10	66.02	134s	310	55.20	55.35	86s	101	52.18	63.85	129s	137
LogLRR	63.20	54.36	18s	218	55.00	53.14	127s	221	52.70	64.97	533s	248
LatLRR	72.60	59.88	6s	74	57.45	57.95	48s	69	49.61	64.15	172s	68
IRLS	74.90	63.57	240s	64	56.35	59.37	4947s	73	50.60	64.18	169s	60
ARM	74.70	63.68	16s	35	62.15	64.02	138s	31	55.06	65.33	128s	31
S_1NFLRR	74.70	63.65	3s	51	56.95	54.09	47s	105	52.75	63.84	58s	135
$S_{1/2}$NFLRR	77.00	63.83	10s	103	57.45	52.58	68s	155	55.58	63.82	57s	123
$S_{2/3}$NFLRR	74.80	63.60	5s	75	58.60	54.76	55s	117	55.01	63.91	59s	124
JS$_1$NFOM	80.30	81.71	6s	73	66.15	68.89	37s	73	53.44	65.99	40s	73
JS$_{1/2}$NFOM	**80.70**	**82.14**	7s	74	67.05	66.09	37s	73	**57.47**	67.70	42s	75
JS$_{2/3}$NFOM	79.50	80.18	6s	75	**67.20**	**69.86**	37s	73	56.16	**68.10**	41s	74

and our proposed JSpNFOM methods. Their performance was compared across the selected three p-values.

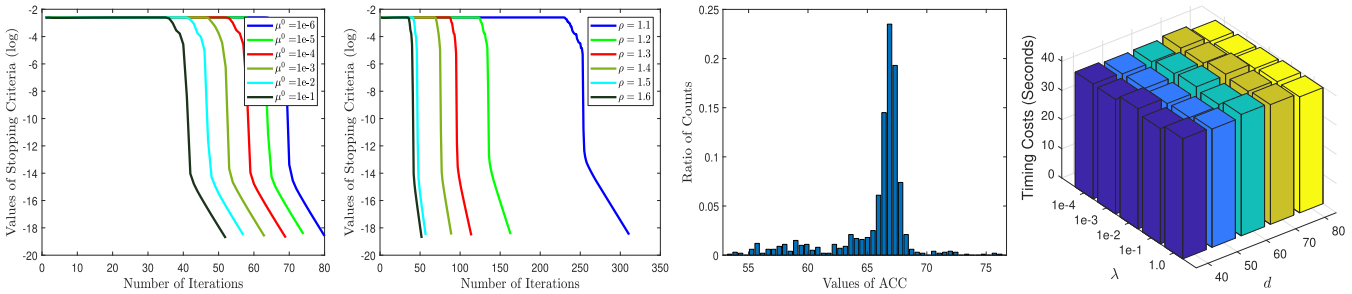
All experiments were conducted to compare our proposed method with 11 clustering methods. The experiments were performed on a personal computer equipped with an Intel[1] Core[2] i7-7700 CPU at 3.6 GHz and 8.0 GB RAM, using the MATLAB R2021b environment. For the comparison methods, we either tuned the model and algorithmic parameters based on λ, μ, ρ, and p, or used the default values of the released codes for relevant optimization solutions. The factorization methods required an estimate of the upper bound of the rank number, denoted as d, which was set to be larger than the number of classes. Moreover, most of the involved methods applied different post-processing techniques on the learned affinity matrices when using the spectral clustering technique on the affinity matrix $\mathbf{Z}^*$ to achieve the final segmentation of the data points. The normalized cut strategy was used to segment the affinity matrix, and as verified in the previous studies [7], [29], [61], different clustering methods could lead to different clustering results. Subsequently, we present both qualitative and quantitative observations and analyses for two tasks: image clustering and motion segmentation. The best results are reported in bold, and the second-best results are underlined in all subsequent tables. We highlight their superior clustering performance and computational efficiency compared to existing approaches. Additionally, we have included detailed validations of the theoretical properties, including convergence analysis, the property of the coefficient matrix, and the impact of each modulus of the objective term, among other relevant viewpoints. These improvements aim to provide a more thorough evaluation of our proposed methods.

A. Image Clustering

In this section, we present the results of numerical experiments conducted on six image databases and then report the quantitative and qualitative outcomes obtained from various settings, including clustering accuracy (ACC) and normalized mutual information (NMI) for performance evaluation, as well as timing costs (TIME) and iteration numbers (ITER) for efficiency comparison. We here stress that higher values of ACC (%) and NMI (%), along with lower values of TIME and ITER, indicate superior computational efficacy and efficiency for the clustering methods.

Table V compares the clustering performance and computational efficiency of various methods on three different image databases: AR, ExtYaleB, and COIL10, which have

[1]Registered trademark.
[2]Trademarked.



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

Fig. 4. Visual comparisons of our methods on the databases, including the influences of (a) initial parameter μ^0 and (b) step-size ρ for USPS. (c) Randomly initialized variables on the values of ACC for MNIST. (d) Parameter pairs (λ, d) on the timing costs for FLAVIA.

TABLE VII

COMPARISON OF SEGMENTATION ERROR AND COMPUTATIONAL TIMING FOR THE INVOLVED METHODS ON THE HOPKINS155 DATABASE

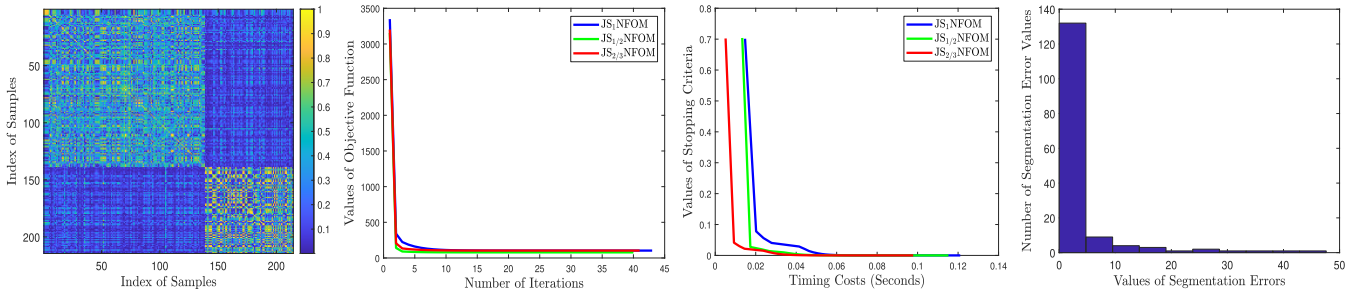
Method	Two Motions				Three Motions				All Motions			
	MEAN	STD	MED	TIME	MEAN	STD	MED	TIME	MEAN	STD	MED	TIME
SSC	2.39	6.08	0.35	1.03s	8.45	12.77	2.57	2.32s	3.77	8.45	0.52	1.32s
LRR	2.60	6.70	0.00	3.85s	6.47	9.61	1.14	8.26s	3.48	7.60	0.00	4.85s
LRRSC	21.69	16.50	22.50	2.49s	34.05	17.06	35.85	5.10s	24.48	17.41	26.51	3.07s
FULRR	15.88	17.62	6.43	0.88s	15.63	17.49	9.13	1.14s	15.82	17.53	7.14	0.94s
NSC	18.93	18.43	17.80	0.21s	26.93	14.76	28.30	0.62s	20.74	17.94	21.95	0.31s
SGL	6.01	11.01	0.10	0.33s	13.68	13.50	9.89	0.47s	7.74	12.01	1.16	0.36s
LogLRR	3.11	9.04	0.00	1.31s	**4.97**	9.15	0.91	1.71s	3.53	9.07	0.20	1.40s
LatLRR	1.83	5.72	0.00	0.27s	5.81	8.89	2.37	0.34s	2.73	6.77	0.00	0.29s
IRLS	2.23	6.40	0.00	0.23s	5.06	8.77	1.42	0.47s	2.87	7.07	0.00	0.28s
ARM	6.89	10.92	0.00	1.63s	23.33	16.23	25.32	3.49s	10.60	14.07	0.77	2.05s
S_1NFLRR	2.89	9.48	0.00	0.59s	5.70	10.39	0.37	1.30s	3.52	9.73	0.00	0.75s
$S_{1/2}$NFLRR	2.35	7.67	0.00	0.72s	6.86	11.21	0.71	1.54s	3.37	8.78	0.00	0.90s
$S_{2/3}$NFLRR	2.18	7.46	0.00	0.80s	6.25	10.63	0.71	1.74s	3.10	8.44	0.00	1.01s
JS_1NFOM	1.67	6.14	0.00	0.17s	6.35	10.58	0.93	0.42s	2.73	7.60	0.00	0.23s
$JS_{1/2}$NFOM	**1.63**	6.13	0.00	0.20s	6.25	10.58	0.71	0.43s	**2.67**	7.58	0.00	0.25s
$JS_{2/3}$NFOM	1.67	6.13	0.00	0.19s	6.16	10.54	0.48	0.39s	2.68	7.55	0.00	0.23s

larger sample dimensions than sample numbers. It can be observed that the LatLRR method achieved the higher NMI values across three databases, with the proposed $JS_{1/2}$NFOM and $JS_{2/3}$NFOM methods following closely behind. In terms of computational efficiency, the SSC, LRRSC, and ARM methods exhibited lower timing costs and fewer iterations in comparison to other techniques. Conversely, the NSC, LogLRR, and FULRR methods required a greater number of iterations. In general, factorization approaches such as SpNFLRR and JSpNFOM exhibit efficient performance. The integration of prior information can further enhance clustering accuracy, ultimately leading to better results.

To provide some visual comparisons, Fig. 3 presents the plotted curves of the stopping criteria for our proposed methods with $p \in \{1, 1/2, 2/3\}$ in Fig. 3(a) for the AR database, the block-diagonal structures of the coefficient matrix related to the number of subjects in Fig. 3(b) for the ExtYaleB database, and the statistical distribution of representation coefficients related to the index of all clustering samples in Fig. 3(c) for the ExtYaleB database. Besides this, we display the coefficient matrix as a 3-D surface plot with color in Fig. 3(d) to visualize the differences among the coefficient values. These visual results offer valuable insights for our proposed methods.

Table VI compares the clustering performance and computational efficiency of various methods on three different image databases: USPS, MNIST, and FLAVIA, which have smaller sample dimensions than sample numbers. The $JS_{1/2}$NFOM method achieves the best performance in terms of ACC and NMI on the USPS and FLAVIA databases, while the $JS_{2/3}$NFOM method achieves the best performance in terms of NMI on the MNIST database. Additionally, the $JS_{2/3}$NFOM method achieves the best performance in terms of computational efficiency, with the nearly lowest computational time and number of iterations. Among the compared methods, NSC, SGL, and ARM exhibit superior clustering accuracy compared to other methods, while FULRR, LogLRR, LatLRR, and SpNFLRR have higher computational efficiency compared to other methods. This finding suggests that incorporating additional information and utilizing factorization techniques can make the experiments more effective and efficient. Thus, it is reasonable to combine the Schatten p-norm and optimal mean to achieve superior quantitative results.

To provide the visual comparisons, Fig. 4 first presents the non-increasing property of the plotted curves for the stopping criteria of our proposed methods under $p \in \{1, 1/2, 2/3\}$, focusing on Fig. 4(a) with initial values of μ^0, and Fig. 4(b)



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

Fig. 5. Visual comparison results of the proposed methods on the Hopkins 155 database include (a) block-diagonal coefficient matrix, (b) objective function over the number of iterations, (c) stopping criteria over timing costs, and (d) histogram of segmentation error values.

TABLE VIII

COMPARISON OF RESULTS FOR ABLATION EXPERIMENTS ON CLUSTERING ACCURACY (%) OVER SIX IMAGE DATASETS AND SEGMENTATION ERROR (%) OVER HOPKINS 155 DATASET. NOTE THAT HIGHER ACCURACY AND LOWER ERROR IMPLY BETTER PERFORMANCE

//	p	AR	ExtYaleB	COIL10	USPS	MNIST	FLAVIA	Hopkins 155		
								Two Motions	Three Motions	All Motions
Proposed	1	62.71	82.50	79.17	80.30	66.15	53.44	1.67	6.35	2.73
	1/2	68.29	90.13	83.06	**80.70**	67.05	**57.47**	**1.63**	6.25	**2.67**
	2/3	**69.00**	**92.97**	**83.33**	79.50	**67.20**	56.16	1.67	**6.16**	2.68
Case 1	1	61.14	82.50	78.72	69.70	60.15	50.66	2.14	11.87	4.34
	1/2	65.57	83.91	81.53	74.40	63.65	54.80	2.07	12.01	4.31
	2/3	66.71	84.38	81.08	76.10	63.05	52.54	1.93	11.02	3.98
Case 2	1/2	67.29	86.59	82.44	78.50	65.81	55.79	1.93	9.12	3.55
	2/3	67.71	87.53	81.97	78.60	65.45	54.10	1.74	8.39	3.24
Case 3	1/2	67.29	90.00	81.28	79.20	66.20	55.06	1.85	9.43	3.56
	2/3	68.86	89.38	81.66	78.70	64.90	55.21	1.81	8.67	3.36

with various values of ρ on the USPS database based on the acceleration continuity strategy. In addition, we illustrate the influences of the randomly initial variables by running 1000 times in Fig. 4(c) for the MNIST database. Moreover, we showcase the influences of parameters (λ, d) on the timing costs for the FLAVIA database in Fig. 4(d). These results validate the convergence property and the effects of clustering accuracy and computation time derived from initial variables and parameter pairs, respectively.

B. Motion Segmentation

In this section, we will evaluate the performance of the JSpNFOM methods in motion segmentation by analyzing the clustering error values (%) measured by mean (MEAN), standard deviation (STD), and median (MED), as well as the average timing costs (TIME) measured in seconds, following the approach used in [29] and [61]. Motion segmentation refers to the unsupervised analysis of video sequences, where points in multiple image frames of a dynamic scene are clustered into different motions of moving objects. The observations from the evaluation are presented below.

Table VII presents a comprehensive comparison of our proposed JSpNFOM methods with several subspace clustering methods. Our methods demonstrate effective and robust performance, achieving the best segmentation accuracy with the lowest mean error and execution time in most cases. Additionally, the IRLS method also performs well, exhibiting low execution time and competitive segmentation accuracy. However, the LatLRR and LogLRR methods display good segmentation accuracy but are relatively slower compared to other methods. On the other hand, the NSC and SGL methods exhibit the lowest execution time but with slightly lower segmentation accuracy than some other comparison methods. Further analysis reveals that incorporating additional information effectively reduces segmentation errors in the LRRSC method, while utilizing factorization techniques significantly improves the computational efficiency of the FULRR and SpNFLRR methods. These findings provide explanations for the superior performance of our proposed methods.

In addition to the quantitative comparison results, Fig. 5 presents the visual results of the proposed methods. Fig. 5(a) shows the block-diagonal coefficient matrix for two motions, while Fig. 5(b) and 5(c) provide plotted comparisons of the three proposed methods, demonstrating their non-increasing property. Fig. 5(d) illustrates that the largest ratios of segmentation error values are smaller than 10. These qualitative comparisons provide additional evidence for the interpretability and rationality of the proposed methods and offer valuable insights into this experimental task.

C. Further Analysis and Discussion

In this section, we thoroughly assess our proposed methods from several viewpoints. First, we perform ablation validations for each component, as listed in Table VIII. Moreover, we demonstrate the reconstruction ability of our proposed methods on the ExtYaleB database, presented in Fig. 6. Additionally, we validate our methods on synthetic data based

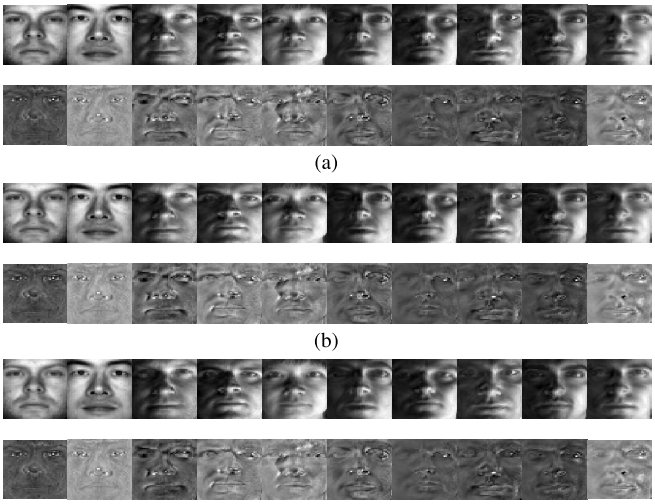


Fig. 6. Visual comparisons of the reconstruction and residual images for three p-values of proposed methods in (a)–(c) on the ExtYaleB database. (a) Achieved by JS_1NFOM. (b) Achieved by $JS_{1/2}$NFOM. (c) Achieved by $JS_{2/3}$NFOM.

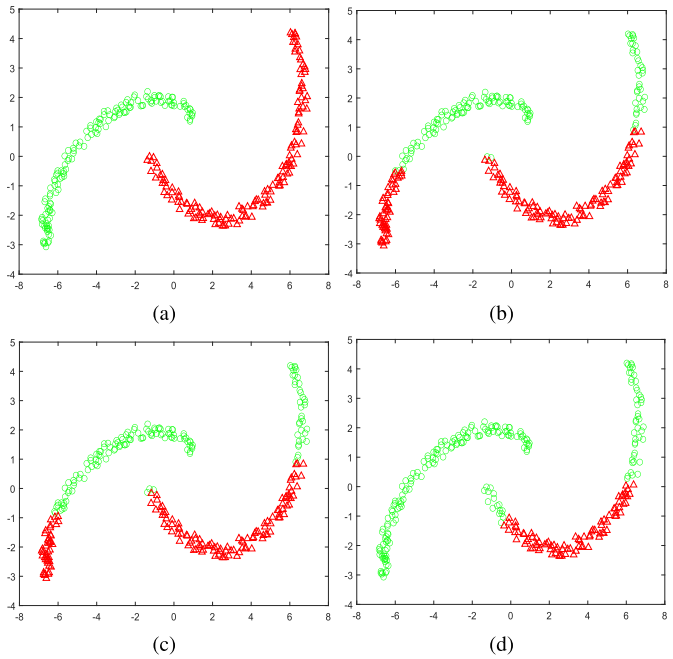


Fig. 7. Visual comparison results for a generated two-moon dataset are presented in (a) original state and (b)–(d) applying three proposed methods by selecting a parameter $p \in \{1, 1/2, 2/3\}$ in sequence.

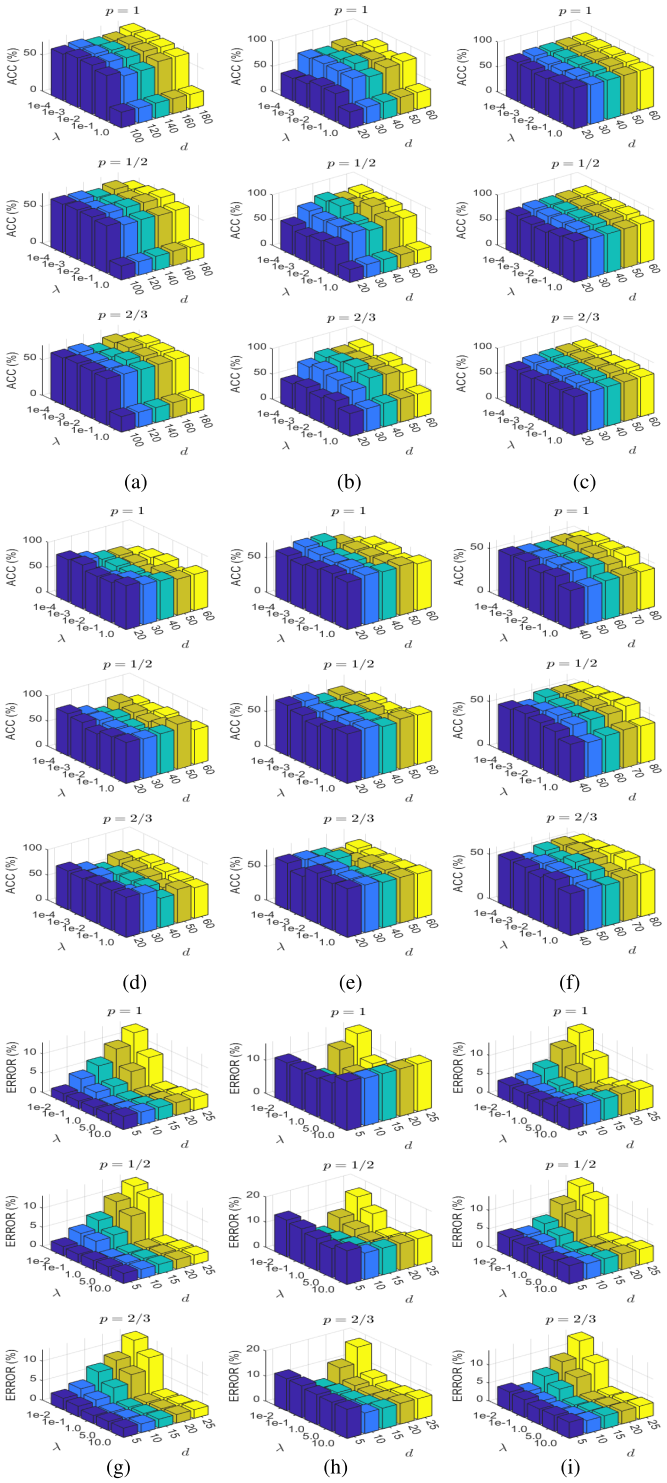


Fig. 8. Analysis of parameter pairs (λ, d) performed for the proposed methods with three p-values on all experimental databases. (a) AR. (b) ExtYaleB. (c) COIL10. (d) USPS. (e) MNIST. (f) FLAVIA. (g) Two motions. (h) Three motions. (i) All motions.

on the two-moons dataset, as depicted in Fig. 7. Furthermore, we explore the effect of selecting $p \in \{1, 1/2, 2/3\}$ and conduct a sensitivity analysis on all databases, as shown in Fig. 8. Then, we make the following observations.

1) *Ablation Validations:* We performed experiments to assess the effectiveness of each component in our proposed method, using JS_1NFOM as the baseline. To illustrate it, we further evaluated the contribution of our method through three different cases as follows.
   a) *Case 1:* The proposed methods without the use of optimal mean $\mathbf{b}$ under $p = 1$, $1/2$, and $2/3$.
   b) *Case 2:* Fixing the $\ell_{2,1}$ norm and changing the Schatten p-norm factorization for $p = 1/2$ and $2/3$.
   c) *Case 3:* Fixing the nuclear norm factorization and changing the $\ell_{2,p}$ norm for $p = 1/2$ and $2/3$.

The results presented in Table VIII validate that the incorporation of the optimal mean, $\ell_{2,p}$-norm, and

Schatten p-norm has a positive impact on clustering performance compared to our proposed method. Notably, removing the optimal mean in Case 1 resulted in a reduction in performance. Moreover, using $p = 1/2$ and $p = 2/3$ in Cases 2 and 3, respectively, led to a significant increase in clustering accuracy for both image clustering and motion segmentation, when compared to using $p = 1$.

2) *Validation of Reconstruction Ability:* To illustrate the effectiveness of our JS$_p$NFOM methods in image reconstruction for three p-values, we present visual examples in Fig. 6(a)–(c). These examples demonstrate a series of reconstructed images and their corresponding residual images from the ExtYaleB database with illumination corruptions. Upon visual inspection of these samples, it is evident that our methods successfully eliminate illuminations from the selected clustering images compared to the raw data shown in Fig. 2 (the second row), showcasing their ability to preserve image fidelity. This is due to the fact that the weakness of illuminations is captured by the residual images, which are subtracted from the raw face images during the reconstruction process.
3) *Validation of Two Moons:* The visualization of the Ncut data clustering results is presented in Fig. 7, where two original clusters of data are distributed in a two-moons shape, as seen in Fig. 7(a). This synthetic dataset is used to evaluate our algorithms with three different p-values. The data was randomly generated from two sine-shaped curves, with the noise percentage set to 0.10, and each cluster containing 300 samples. Visual differences between the results are apparent, with clustering accuracies of 69.67%, 72.33%, and 79.67% corresponding to Fig. 7(b)–(d). These results confirm the influence of the p-value choice on the clustering performance.
4) *Sensitivity of Parameters:* The minimization problem (4) involves the parameter pairs (p, d, λ), which we analyze under three selected p-values shown in Fig. 8. Among these parameters, we recognize that d plays a crucial role as it represents an upper bound on the true rank r of the coefficient matrix, where r is typically associated with the number of subjects in each clustering database. The specific value of d varies depending on the number of subjects in different databases, necessitating different choices for d. As for the selection of the parameter λ, we determine its value through empirical analysis. Specifically, we selected values from $\{1e^{-4}, 1e^{-3}, 1e^{-2}, 1e^{-1}, 1.0\}$ for the six image datasets. Additionally, for the AR dataset, we chose d from the set $\{100, 120, 140, 160, 180\}$, for the FLAVIA dataset, from $\{40, 50, 60, 70, 80\}$, and for the remaining image datasets, from $\{20, 30, 40, 50, 60\}$. As for the Hopkins 155 dataset, we set $\lambda \in \{1e^{-2}, 1e^{-1}, 1.0, 5.0, 10.0\}$ and $d \in \{5, 10, 15, 20, 30\}$. Notably, the proper choices of λ and d play a critical role in ensuring the robust performance of our proposed methods.

VI. Conclusion and Future Work

This work introduces a novel and efficient iteratively reweighted optimization algorithm for addressing unconstrained nonconvex problems. The proposed approach, JSpNFOM, leverages Schatten p-norm factorization and $\ell_{2,p}$-norm to enhance the robustness of clustering performance. We also integrate the optimal mean function into the residual measurements, further improving the clustering accuracy in the multivariate nonconvex LRR model. To reduce computational complexity, we factorize the large-scale coefficient matrix into two smaller matrices. In addition to presenting theoretical analyses for computational complexity and convergence properties, we conducted experiments on real-world databases. The results demonstrate the consistency of our proposed method's quantitative and qualitative performance for subspace clustering, validating the theoretical properties. Moreover, our method outperforms most related spectral clustering approaches in terms of both computational efficacy and efficiency.

As part of our future work, we acknowledge the challenges posed by the nonconvexity and multivariate nature of our proposed methods, making it difficult to provide provable convergence guarantees. Therefore, we find it highly meaningful and valuable to focus on developing efficient optimization algorithms with both local and global theoretical analyses, as demonstrated in previous works [50], [62]. Additionally, we have observed that our proposed methods involve multiple parameters, and tuning them to select optimal values can be a slightly complex process. To address this issue, we suggest leveraging neural network technologies, as showcased in [63] and [64], to learn the optimal parameters corresponding to each iteration. By adopting these improvements, we aim to significantly enhance the optimization process, making our methods more versatile and applicable beyond the research domain of unsupervised learning.

Acknowledgment

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

References

[1] T. Bouwmans, A. Sobral, S. Javed, S. K. Jung, and E.-H. Zahzah, "Decomposition into low-rank plus additive matrices for background/foreground separation: A review for a comparative evaluation with a large-scale dataset," *Comput. Sci. Rev.*, vol. 23, pp. 1–71, Feb. 2017.
[2] Z. Hu, F. Nie, R. Wang, and X. Li, "Low rank regularization: A review," *Neural Netw.*, vol. 136, pp. 218–232, Apr. 2021.
[3] Y. Fu, C. Luo, and Z. Bi, "Low-rank joint embedding and its application for robust process monitoring," *IEEE Trans. Instrum. Meas.*, vol. 70, pp. 1–13, 2021.
[4] J. Ahmed, B. Gao, W. L. Woo, and Y. Zhu, "Ensemble joint sparse low-rank matrix decomposition for thermography diagnosis system," *IEEE Trans. Ind. Electron.*, vol. 68, no. 3, pp. 2648–2658, Mar. 2021.
[5] H. Yan, K. Paynabar, and J. Shi, "Image-based process monitoring using low-rank tensor decomposition," *IEEE Trans. Autom. Sci. Eng.*, vol. 12, no. 1, pp. 216–227, Jan. 2015.
[6] Z. Zha, B. Wen, X. Yuan, J. Zhou, C. Zhu, and A. C. Kot, "Low-rankness guided group sparse representation for image restoration," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 34, no. 10, pp. 7593–7607, Oct. 2023, doi: 10.1109/TNNLS.2022.3144630.

[7] G. Liu, Z. Lin, S. Yan, J. Sun, Y. Yu, and Y. Ma, "Robust recovery of subspace structures by low-rank representation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 1, pp. 171–184, Jan. 2013.

[8] H. Zhang, J. Yang, J. Qian, and W. Luo, "Nonconvex relaxation based matrix regression for face recognition with structural noise and mixed noise," *Neurocomputing*, vol. 269, pp. 188–198, Dec. 2017.

[9] 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.

[10] 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, Jan. 2010.

[11] H. Zhang, C. Gong, J. Qian, B. Zhang, C. Xu, and J. Yang, "Efficient recovery of low-rank matrix via double nonconvex nonsmooth rank minimization," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 30, no. 10, pp. 2916–2925, Oct. 2019.

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

[13] 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, Jan. 2010.

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

[15] J. Qian, W. K. Wong, H. Zhang, J. Xie, and J. Yang, "Joint optimal transport with convex regularization for robust image classification," *IEEE Trans. Cybern.*, vol. 52, no. 3, pp. 1553–1564, Mar. 2022.

[16] 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.*, vol. 40, no. 9, pp. 2066–2080, Sep. 2018.

[17] H. Zhang et al., "Efficient and effective nonconvex low-rank subspace clustering via SVT-free operators," *IEEE Trans. Circuits Syst. Video Technol.*, early access, May 11, 2023, doi: 10.1109/TCSVT.2023.3275299.

[18] 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.

[19] 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.

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

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

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

[23] Q. Shen, Y. Chen, Y. Liang, S. Yi, and W. Liu, "Weighted Schatten p-norm minimization with logarithmic constraint for subspace clustering," *Signal Process.*, vol. 198, Sep. 2022, Art. no. 108568.

[24] Q. Shen, Y. Liang, S. Yi, and J. Zhao, "Fast universal low rank representation," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 32, no. 3, pp. 1262–1272, Mar. 2022.

[25] G. Liu and S. Yan, "Latent low-rank representation for subspace segmentation and feature extraction," in *Proc. Int. Conf. Comput. Vis.*, Nov. 2011, pp. 1615–1622.

[26] M. Brbic and I. Kopriva, "ℓ_0-motivated low-rank sparse subspace clustering," *IEEE Trans. Cybern.*, vol. 50, no. 4, pp. 1711–1725, Apr. 2020.

[27] H. Jia, D. Zhu, L. Huang, Q. Mao, L. Wang, and H. Song, "Global and local structure preserving nonnegative subspace clustering," *Pattern Recognit.*, vol. 138, Jun. 2023, Art. no. 109388.

[28] 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.

[29] H. Zhang, J. Yang, F. Shang, C. Gong, and Z. Zhang, "Lrr for subspace segmentation via tractable Schatten-p norm minimization and factorization," *IEEE Trans. Cybern.*, vol. 49, no. 5, pp. 1722–1734, May 2019.

[30] 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.

[31] Z. Kang, Z. Lin, X. Zhu, and W. Xu, "Structured graph learning for scalable subspace clustering: From single view to multiview," *IEEE Trans. Cybern.*, vol. 52, no. 9, pp. 8976–8986, Sep. 2022.

[32] E. Kim, M. Lee, and S. Oh, "Elastic-net regularization of singular values for robust subspace learning," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 915–923.

[33] W. Jiang, J. Liu, H. Qi, and Q. Dai, "Robust subspace segmentation via nonconvex low rank representation," *Inf. Sci.*, vols. 340–341, pp. 144–158, May 2016.

[34] J. Zhao, S. Yi, Y. Liang, W. Liu, and X. Cao, "Robust active representation via $\ell_{2,p}$-norm constraints," *Knowl.-Based Syst.*, vol. 235, Jan. 2022, Art. no. 107639.

[35] F. Nie, J. Yuan, and H. Huang, "Optimal mean robust principal component analysis," in *Proc. Int. Conf. Mach. Learn. (ICML)*, 2014, pp. 1062–1070.

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

[37] S. Yi, F. Nie, Y. Liang, W. Liu, Z. He, and Q. Liao, "Fast extended inductive robust principal component analysis with optimal mean," *IEEE Trans. Knowl. Data Eng.*, vol. 34, no. 10, pp. 4812–4825, Oct. 2022.

[38] 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.

[39] S. Yi, Z. He, X. Y. Jing, Y. Li, Y. M. Cheung, and F. Nie, "Adaptive weighted sparse principal component analysis for robust unsupervised feature selection," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 31, no. 6, pp. 2153–2163, Jun. 2020.

[40] P. Li, W. Zhang, C. Lu, R. Zhang, and X. Li, "Robust kernel principal component analysis with optimal mean," *Neural Netw.*, vol. 152, pp. 347–352, Aug. 2022.

[41] S. Boyd, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Found. Trends Mach. Learn.*, vol. 3, no. 1, pp. 1–122, 2010.

[42] 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.

[43] 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, Mar. 2012.

[44] 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, Jan. 2016.

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

[46] 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.

[47] H. Zhang, F. Qian, F. Shang, W. Du, J. Qian, and J. Yang, "Global convergence guarantees of (A)GIST for a family of nonconvex sparse learning problems," *IEEE Trans. Cybern.*, vol. 52, no. 5, pp. 3276–3288, May 2022.

[48] C. Lu, J. Tang, S. Yan, and Z. Lin, "Nonconvex nonsmooth low rank minimization via iteratively reweighted nuclear norm," *IEEE Trans. Image Process.*, vol. 25, no. 2, pp. 829–839, Feb. 2016.

[49] H. Zhang, J. Qian, B. Zhang, J. Yang, C. Gong, and Y. Wei, "Low-rank matrix recovery via modified Schatten-p norm minimization with convergence guarantees," *IEEE Trans. Image Process.*, vol. 29, pp. 3132–3142, 2020.

[50] H. Zhang, J. Qian, J. Gao, J. Yang, and C. Xu, "Scalable proximal Jacobian iteration method with global convergence analysis for nonconvex unconstrained composite optimizations," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 30, no. 9, pp. 2825–2839, Sep. 2019.

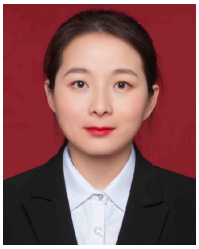
[51] H. Attouch, J. Bolte, P. Redont, and A. Soubeyran, "Proximal alternating minimization and projection methods for nonconvex problems: An approach based on the Kurdyka–Łojasiewicz inequality," *Math. Oper. Res.*, vol. 35, no. 2, pp. 438–457, May 2010.

[52] J. Wang, M. Wang, X. Hu, and S. Yan, "Visual data denoising with a unified Schatten-p norm and ℓ_q norm regularized principal component pursuit," *Pattern Recognit.*, vol. 48, no. 10, pp. 3135–3144, Oct. 2015.

[53] E. J. Candès, X. Li, Y. Ma, and J. Wright, "Robust principal component analysis?" *J. ACM*, vol. 58, no. 1, pp. 1–37, 2009.
[54] M. Fazel, H. Hindi, and S. P. Boyd, "A rank minimization heuristic with application to minimum order system approximation," in *Proc. Amer. Control Conf.*, Jun. 2001, pp. 4734–4739.
[55] H. Xu, C. Caramanis, and S. Sanghavi, "Robust PCA via outlier pursuit," in *Proc. Adv. Neural. Inf. Process. Syst (NeurIPS)*, vol. 23, 2010, pp. 2496–2504.
[56] D. G. Luenberger and Y. Ye, *Linear and Nonlinear Programming*. Cham, Switzerland: Springer, 1984.
[57] S. Zhang, H. Qian, and X. Gong, "An alternating proximal splitting method with global convergence for nonconvex structured sparsity optimization," in *Proc. Assoc. Adv. Artif. Intell. (AAAI)*, 2016, pp. 2330–2336.
[58] 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.
[59] J. Bolte, S. Sabach, and M. Teboulle, "Proximal alternating linearized minimization for nonconvex and nonsmooth problems," *Math. Program.*, vol. 146, nos. 1–2, pp. 459–494, Aug. 2014.
[60] K. Mohan and M. Fazel, "Iterative reweighted algorithms for matrix rank minimization," *J. Mach. Learn. Res.*, vol. 13, no. 1, pp. 3441–3473, 2012.
[61] Z. Kang, C. Peng, and Q. Cheng, "Robust subspace clustering via tighter rank approximation," in *Proc. 24th ACM Int. Conf. Inf. Knowl. Manage.*, Oct. 2015, pp. 393–401.
[62] 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, Jan. 2015.
[63] H. Cai, J. Liu, and W. Yin, "Learned robust PCA: A scalable deep unfolding approach for high-dimensional outlier detection," in *Proc. Adv. Neural. Inf. Process. Syst. (NeurIPS)*, vol. 34, 2021, pp. 16977–16989.
[64] Z. Xu and J. Sun, "Model-driven deep-learning," *Nat. Sci. Rev.*, vol. 5, no. 1, pp. 22–24, Jan. 2018.

Hengmin Zhang (Member, IEEE) received the Ph.D. degree from the School of Computer Science and Engineering, Nanjing University of Science and Technology (NJUST), Nanjing, China, in 2019.

He was a Post-Doctoral Fellow with the School of Information Science and Engineering, East China University of Science and Technology (ECUST), Shanghai, China, and the PAMI Research Group, Department of Computer and Information Science, University of Macau (UM), Macau, China, from 2019 to 2022. He is currently a Research Fellow with the School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore. He has published more than 30 technical papers at prominent journals and conferences. His research interests include sparse coding and low-rank matrix recovery, nonconvex optimizations, and large-scale representation learning methods.



Jiaoyan Zhao received the M.S. degree from the Hunan Institute of Science and Technology, Yueyang, China, in 2017, and the Ph.D. degree in information and communication engineering from Shenzhen University, Shenzhen, China, in 2023.

She is currently a Lecturer with the School of Artificial Intelligence, Shenzhen Polytechnic University, Shenzhen. Her research interests include active learning and machine learning.

Bob Zhang (Senior Member, IEEE) received the Ph.D. degree in electrical and computer engineering from the University of Waterloo, Waterloo, ON, Canada, in 2011.

He was with the Center for Pattern Recognition and Machine Intelligence and later was a Post-Doctoral Researcher with the Department of Electrical and Computer Engineering, Carnegie Mellon University, Pittsburgh, PA, USA. He is currently an Associate Professor with the Department of Computer and Information Science, University of Macau, Macau, China. His research interests include biometrics, pattern recognition, feature extraction/detection, and image processing.

Dr. Zhang is/was a Technical Committee Member of the IEEE Systems, Man, and Cybernetics Society; and an Associate Editor of IEEE TRANSACTIONS ON SYSTEMS, MAN, AND CYBERNETICS: SYSTEMS, IEEE TRANSACTIONS ON NEURAL NETWORKS AND LEARNING SYSTEMS, *Artificial Intelligence Review*, and *IET Computer Vision*.

Chen Gong (Senior Member, IEEE) received the dual Ph.D. degree from Shanghai Jiao Tong University (SJTU), Shanghai, China, and the University of Technology Sydney (UTS), Ultimo, NSW, Australia, in 2016 and 2017, respectively.

He is currently a Full Professor with the School of Computer Science and Engineering, Nanjing University of Science and Technology, Nanjing, China. He has published more than 100 technical papers at prominent journals and conferences, such as JMLR, IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE (TPAMI), IEEE TRANSACTIONS ON NEURAL NETWORKS AND LEARNING SYSTEMS (TNNLS), IEEE TRANSACTIONS ON IMAGE PROCESSING (TIP), ICML, NeurIPS, ICLR, CVPR, AAAI, and IJCAI.

Dr. Gong has received the "Excellent Doctoral Dissertation Award" of Chinese Association for Artificial Intelligence, the "Young Elite Scientists Sponsorship Program" of China Association for Science and Technology, the "Wu Wen-Jun AI Excellent Youth Scholar Award," and the Science Fund for Distinguished Young Scholars of Jiangsu Province. He was selected as the "Global Top Chinese Young Scholars in AI" released by Baidu. He serves as the Area Chair or a Senior PC Member for several top-tier conferences, such as AAAI, IJCAI, ACM MM, ECML, and AISTATS. He also serves as an Associate Editor for IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY (TCSVT) and NePL.

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

He is currently an Associate Processor with the Key Laboratory of Intelligent Perception and Systems for High-Dimensional Information, Ministry of Education, School of Computer Science and Engineering, NJUST. His research interests include pattern recognition, computer vision, and machine learning.

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

Jian Yang (Member, IEEE) received the Ph.D. degree in pattern recognition and intelligence systems from the Nanjing University of Science and Technology (NJUST), Nanjing, China, in 2002.

In 2003, he was a Post-Doctoral Researcher with the University of Zaragoza, Zaragoza, Spain. He was a Post-Doctoral Fellow with the Biometrics Centre, The Hong Kong Polytechnic University, Hong Kong, and the Department of Computer Science, New Jersey Institute of Technology, Newark, NJ, USA, from 2004 to 2006 and 2006 to 2007, respectively. He is currently a Chang-Jiang Professor with the School of Computer Science and Engineering, NJUST. He has authored more than 380 scientific articles in pattern recognition and computer vision. His articles have been cited more than 39 000 times in the Scholar Google. His research interests include pattern recognition, computer vision, biometrics, and machine learning.

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