Research Article

One-Step Robust Low-Rank Subspace Segmentation for Tumor Sample Clustering

Jian Liu [ID], **Yuhu Cheng, Xuesong Wang** [ID], **and Shuguang Ge**

School of Information and Control Engineering, China University of Mining and Technology, Xuzhou 221116, China

Correspondence should be addressed to Xuesong Wang; wangxuesongcumt@163.com

Clustering of tumor samples can help identify cancer types and discover new cancer subtypes, which is essential for effective cancer treatment. Although many traditional clustering methods have been proposed for tumor sample clustering, advanced algorithms with better performance are still needed. Low-rank subspace clustering is a popular algorithm in recent years. In this paper, we propose a novel one-step robust low-rank subspace segmentation method (ORLRS) for clustering the tumor sample. For a gene expression data set, we seek its lowest rank representation matrix and the noise matrix. By imposing the discrete constraint on the low-rank matrix, without performing spectral clustering, ORLRS learns the cluster indicators of subspaces directly, i.e., performing the clustering task in one step. To improve the robustness of the method, capped norm is adopted to remove the extreme data outliers in the noise matrix. Furthermore, we conduct an efficient solution to solve the problem of ORLRS. Experiments on several tumor gene expression data demonstrate the effectiveness of ORLRS.

1. Introduction

Tumor is a group of cells that have undergone unregulated growth and often form a mass or lump. It is critical to reveal the pathogenesis of cancer by analyzing tumor gene expression data. The advances of various sequencing technologies have made it possible to measure the expression levels of thousands of genes simultaneously [1]. Increasingly, one challenge is how to interpret these gene expression data to gain insights into mechanisms of tumors [2]. Many advanced machine learning algorithms [3–9] have thus been proposed to analyze various data. Among them, clustering can be used for discovering tumor samples with similar molecular expression patterns [10, 11].

Many traditional clustering methods, such as hierarchical clustering (HC) [12, 13], self-organizing maps (SOM) [14], nonnegative matrix factorization (NMF) [15, 16], and principal component analysis (PCA) [17–20] have been used for gene expression data clustering. The gene expression data often contains structures that can be represented and processed by some parametric models. The linear subspaces are possible to characterize a given set of data since they are easy

to calculate and often effective in real applications. The subspace methods, such as NMF, are essentially based on the assumption that the data is approximately drawn from a low-dimensional subspace. In recent years, these methods have been gaining much attention. For example, Yu et al. proposed a correntropy-based hypergraph regularized NMF (CHNMF) method for clustering and feature selection [21]. Specifically, the correntropy is used in the loss term of CHNMF instead of the Euclidean norm to improve the robustness of the algorithm. And, CHNMF also uses the hypergraph regularization to explore the high-order geometric information in more sample points. Jiao et al. proposed a hypergraph regularized constrained nonnegative matrix factorization (HCNMF) method for selecting differentially expressed genes and tumor sample classification [22]. HCNMF incorporates a hypergraph regularization constraint to consider the higher order data sample relationships. A nonnegative matrix factorization framework based on multisubspace cell similarity learning for unsupervised scRNA-seq data analysis (MscNMF) was proposed by Wang et al. [23]. MscNMF can learn the gene features and cell features of different subspaces, and the correlation and

heterogeneity between cells will be more prominent in multisubspaces, resulting in the final cell similarity learning will be more satisfactory.

However, real data rarely can be well represented by a single subspace. A more reasonable model is to assume that the data are lying near multiple subspaces (i.e., the data are considered as samples approximately drawn from a mixture of multiple low-dimensional subspaces). Subspace clustering (or segmentation) has been proposed to improve clustering accuracy. It is assumed that the data points are drawn from the combination of multiple low-dimensional subspaces. The goal of subspace clustering is to obtain such multiple low-dimensional subspaces with each subspace corresponding to a cluster. Subspace clustering has obtained promising results in previous studies, and subspace clustering methods have been found widespread applications in many areas, such as pattern recognition [24], image processing [25], and bioinformatics [26].

When the data are clean, i.e., the samples can be strictly drawn from multiple subspaces, several existing methods, such as sparse subspace clustering (SSC) [27], low-rank representation (LRR) [5], and low-rank model with discrete group structure constraint (LRS) [28], are able to solve the subspace clustering problem. SSC clusters the data drawn from multiple low-dimensional subspaces based on sparse representation (SR) [29]. Since low-rank structure can well perform matrix recover, the multiple subspaces can be exactly recovered by LRR. Recently, many excellent works based on low-rank representation are published. For example, Tang et al. proposed a multiview subspace clustering model by learning a joint affinity graph for multiview subspace clustering based on low-rank representation with diversity regularization and rank constraint [30]. This method can effectively suppress redundancy and enhance the diversity of different feature views. In addition, the cluster number is used to promote affinity graph learning by using a rank constraint. In [31], an unsupervised linear feature selective projection (FSP) method was proposed for feature extraction with low-rank embedding and dual Laplacian regularization. FSP can take advantage of the inherent relationship between data and can effectively suppress the influence of noise. LRR have two steps in the clustering task: building the affinity matrix and performing spectral clustering. How to define an excellent affinity matrix is crucial. Furthermore, the clustering problem will be transformed into a segmentation problem of graph by using spectral clustering. The choice of segmentation criteria will directly affect the clustering results. To address the above concerns, LRS directly grasps the indicators of different subspaces via the discrete constraint. As a result, multiple low-rank subspaces can be obtained clearly. Furthermore, Nie et al. introduced a piecewise function to relax the rank constraint which makes LRS better at handling the noisy dataset than the preliminary version [32].

As pointed out in [33], one major challenge of subspace clustering is to deal with the outliers that exist in data. Therefore, robust subspace clustering has become an active research topic. To address the robustness issue, the main idea is to explore the $L_{2,1}$-norm based objective functions since the nonsquared residuals of $L_{2,1}$-norm can reduce the effects of data outliers. In [34, 35], the $L_{2,1}$-norm is adopted in robust PCA (RPCA) for detecting outliers. In [33], Liu et al. proposed a robust LRR model via $L_{2,1}$-norm for subspace clustering. Although the $L_{2,1}$-norm is robust to outliers, it still suffers from the extreme data outliers. The $L_{2,1}$-norm just reduces, not completely removes, the effects of the outliers. Capped norm is a more robust strategy than $L_{2,1}$-norm due to the fact that it can remove the effects of the outliers. It has been recently studied in many applications [36, 37].

In this paper, a one-step robust low-rank subspace segmentation (ORLRS) method via the discrete constraint and capped norm is proposed for clustering tumor sample. For a data set $\mathbf{X} \in \mathbb{R}^{m \times n}$ with m genes and n samples, a low-rank representation matrix $\mathbf{A} \in \mathbb{R}^{m \times n}$ and a noise matrix $\mathbf{E} \in \mathbb{R}^{m \times n}$, i.e., $\mathbf{X} = \mathbf{A} + \mathbf{E}$, are being sought. The low-rank representation of the i-th subspace $\mathbf{A}_i$ can be denoted as $\mathrm{rank}(\mathbf{A}_i)$. Here, we impose the discrete constraint on a diagonal matrix $\mathbf{I}_i \in \mathbb{R}^{n \times n}$ to obtain the low-rank representation $\mathrm{rank}(\mathbf{X}\mathbf{I}_i)$, where $\mathbf{I}_i \subseteq \{0, 1\}$ and $\sum_{i=1}^{c} \mathbf{I}_i = \mathbf{I}$ (c is the number of total subspaces and $\mathbf{I}$ is an identify matrix). The indicators of the i-th cluster are included in $\mathbf{I}_i$. In contrast to traditional low-rank based models, we can directly learn the cluster indicators. To avoid trivial solutions and approximate the low-rank constraint, the rank of all subspace simultaneously can be minimized as $\sum_{i=1}^{c} (\|\mathbf{A}\mathbf{I}_i\|_{S_p}^{k})^2$, where $\|\bullet\|_{S_p}^{k}$ denotes the Schatten p-norm which has a better relaxation than the nuclear norm [38]. For the noise matrix $\mathbf{E}$, capped norm is used to improve the robustness. We define θ as a thresholding parameter for choosing the extreme data outliers, and then the capped norm of $\mathbf{E}$ can be formulated as $\|\mathbf{E}\|_{\mathrm{Capped}} = \min_{\mathbf{E}} \sum_{i=1}^{n} \min\{\|\mathbf{E}_i\|_2, \theta\}$. This function treats $\mathbf{E}_i$ equally if $\|\mathbf{E}_i\|_2$ is smaller than θ. Hence, it is more robust to outliers than $L_{2,1}$-norm. Meanwhile, we derive an efficient optimization algorithm to solve ORLRS with a rigorous theoretical analysis.

The main contributions of our paper are given as follows:

① Compared with traditional low-rank representation-based methods, ORLRS can obtain the clustering result directly by learning a subspace indicator matrix from the low-rank representation matrix without spectral clustering. This avoids the graph construction process in spectral clustering and makes the clustering process simpler.

② We introduced the capped norm into our model and formed a novel objective function for the gene expression data clustering task. Capped norm is used to constrain the noise matrix to improve the robustness of ORLRS.

③ Optimizing the objective function of ORLRS is a nontrivial problem, thus we derive a new optimization algorithm to solve the problem. Furthermore, we have also given a rigorous convergence analysis of ORLRS.

The remainder of the paper is structured as follows. In Section 2, the proposed ORLRS is presented, and the theoretical analysis of the proposed method is provided.

Experimental results are presented in Section 3. In Section 4, the conclusions are given.

2. Methods

We start with a brief introduction of several classical clustering methods. Then, the proposed ORLRS is presented, and the optimal solution and convergence analysis of ORLRS is provided.

2.1. Subspace Clustering via LRR.

Denote $\mathbf{X} \in \mathbb{R}^{m \times n}$ as a data set with m features and n samples. LRR can be defined as

$$\min_{A,E} \|\mathbf{A}\|_* + \lambda \|\mathbf{E}\|_{2,1},$$
$$\text{s.t.} \mathbf{X} = \mathbf{DA} + \mathbf{E}, \tag{1}$$

where $\|\mathbf{A}\|_* = \sum_i \sigma_i(\mathbf{A})$, i.e., the nuclear norm of $\mathbf{A}$ [33], $\|\mathbf{E}\|_{2,1} = \sum_{j=1}^n \sqrt{\sum_{i=1}^m \mathbf{E}_{ij}^2}$ can detect outliers with column-wise sparsity, $\mathbf{D}$ is a dictionary, and $\lambda > 0$ is a balance parameter.

A brief explanation of LRR subspace clustering process is provided as follows. Firstly, the low-rank problem is solved by equation (1). Then, the optimal solution $\mathbf{A}^*$ to equation (1) is used to calculate the affinity matrix by $(|\mathbf{A}^*| + |(\mathbf{A}^*)^T|)/2$, where $|\mathbf{A}^*|$ is the absolute value function. Finally, the data are clustered by using spectral clustering [39].

2.2. One-Step Robust Low-Rank Subspace Clustering.

In this paper, we propose the one-step robust low-rank subspace clustering (ORLRS) method via discrete constraint and capped norm. Different from LRR, ORLRS was proposed for clustering the data by learning the indicators.

Suppose the data matrix $\mathbf{X}$ has c subspaces $\{\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_c\}$, the low-rank representation of each subspace needs to be optimized. In the clustering task, we want each subspace to belong to its own cluster. To obtain a low-rank representation of each subspace, the following formula should be computed: $\text{rank}(\mathbf{X}_i)$, which has trivial solution. Therefore, we need to solve the problem in another way. We define a cluster indicator matrix as $\mathbf{C} \in \mathbb{R}^{c \times n}$: $\mathbf{C}(i, j) = 1$ if the j-th sample belongs to the i-th subspace, and $\mathbf{C}(i, j) = 0$ otherwise. And, the c diagonal matrices are defined as $\mathbf{I}_1, \mathbf{I}_2, \ldots, \mathbf{I}_c \in \mathbf{I}$, where the diagonal elements of $\mathbf{I}_i$ ($1 \leq i \leq c$) are formed by the i-th row of $\mathbf{C}$ and $\mathbf{I}$ is the identity matrix. Then, $\mathbf{XI}_i$ can be represented as the i-th subspace of $\mathbf{X}$. That is, $\text{rank}(\mathbf{X}_i)$ can be rewritten as $\text{rank}(\mathbf{XI}_i)$. We can get the clustering label in one step by directly optimizing $\mathbf{I}_i$ [28].

Finally, the problem of the one-step low-rank subspace clustering method can be defined as

$$\min_{\mathbf{I}_i|_{I=1}^c} \sum_{i=1}^c \left(\|\mathbf{XI}_i\|_{S_p}^p \right)^2,$$
$$\text{s.t.} \mathbf{I}_i|_{I=1}^c \subseteq \{0, 1\}^{n \times n}, \sum_{i=1}^c \mathbf{I}_i = I. \tag{2}$$

where $\|\mathbf{XI}_i\|_{S_p}^p$ is the Schatten p-norm of $\mathbf{XI}_i$. The clustering indicators of each subspace can be obtained from the optimized diagonal matrix $\mathbf{I}_i|_{i=1}^c$ directly.

However, equation (2) is sensitive to data outliers in practical problems since it does not consider the noise in data. To address the robustness problem, we represent the gene expression data $\mathbf{X} \in \mathbb{R}^{m \times n}$ with m genes and n samples as the addition of low-rank representation matrix $\mathbf{A} \in \mathbb{R}^{m \times n}$ and the noise matrix $\mathbf{E} \in \mathbb{R}^{m \times n}$, i.e., $\mathbf{X} = \mathbf{A} + \mathbf{E}$, which is the same strategy as in RPCA. Our one-step low-rank subspace clustering problem can be written as

$$\min_{\mathbf{A},\mathbf{E},\mathbf{I}_i|_{i=1}^c} \sum_{i=1}^c \left(\|\mathbf{AI}_i\|_{S_p}^k \right)^2 + \lambda \|\mathbf{E}\|_L,$$
$$\text{s.t.} \mathbf{X} = \mathbf{A} + \mathbf{E}, \mathbf{I}_i|_{i=1}^c \subseteq \{0, 1\}^{n \times n}, \sum_{i=1}^c \mathbf{I}_i = \mathbf{I}, \tag{3}$$

where $\lambda > 0$ is a balance parameter and $\|\bullet\|_L$ indicates certain regularization strategy. Note that Schatten p-norm is used to approximate the low-rank problem in equation (3) since it is a better relaxation for the rank constraint problem than nuclear norm [38]. The Schatten p-norm ($0 < p < \infty$) of a matrix $\mathbf{G} \in \mathbb{R}^{m \times n}$ was defined as $\|\mathbf{G}\|_{S_p} = (\sum_{i=1}^{\min(m,n)} \sigma_i^p)^{1/p}$, where σ_i is the i-th singular value of $\mathbf{G}$. In [38], the convergence of Schatten p-norm with $0 < p \leq 2$ is proved. Here, we set $0 < 2k \leq 2$ to guarantee the convergence of first term in equation (3). So, the range of k is $0 < k \leq 1$.

To seek a better robustness strategy for the outliers, we adopt capped norm to regularize the noise matrix $\mathbf{E}$, i.e., $\|\mathbf{E}\|_{\text{Capped}}$. Then, equation (3) becomes

$$\min_{\mathbf{A},\mathbf{E},\mathbf{I}_i|_{i=1}^c} \sum_{i=1}^c \left(\|\mathbf{AI}_i\|_{S_p}^k \right)^2 + \lambda \sum_{i=1}^n \min\{\|\mathbf{E}_i\|_2, \theta\},$$
$$\text{s.t.} \mathbf{X} = \mathbf{A} + \mathbf{E}, \mathbf{I}_i|_{i=1}^c \subseteq \{0, 1\}^{n \times n}, \sum_{i=1}^c \mathbf{I}_i = \mathbf{I}, \tag{4}$$

where $\theta > 0$ is a thresholding parameter for choosing the data outliers. If the data point $\|\mathbf{E}_i\|_2 > \theta$, we consider $\mathbf{E}_i$ as extreme outlier, and it is capped as θ. In this way, the influence of extreme outliers is fixed. For other data point $\|\mathbf{E}_i\|_2 \leq \theta$, equation (4) will minimize $\sum_{i=1}^n \|\mathbf{E}_i\|_2$, i.e., the $L_{2,1}$-norm. That is, if θ is set as ∞, $\|\mathbf{E}\|_{\text{Capped}}$ is equivalent to $\|\mathbf{E}\|_{2,1}$. Thus, the capped norm is a more robust strategy than $L_{2,1}$-norm.

As a result, ORLRS provides a more robust low-rank subspace clustering model by using capped norm. And, the clustering indicators of each subspace can be obtained from the optimized diagonal matrix $\mathbf{I}_i|_{i=1}^c$ directly. We will propose an efficient optimization algorithm to solve equation (4) in Section 2.3.

2.3. Optimization Algorithm.

The objective function equation (4) of the ORLRS is nonconvex, thus jointly optimizing $\mathbf{A}$, $\mathbf{E}$, and $\mathbf{I}_i|_{i=1}^c$ is extremely difficult. The augmented Lagrange multiplier (ALM) algorithm is used to optimize equation (4). The Lagrangian function of equation (4) can be written as

$$\min_{\mathbf{A},\mathbf{E},\mathbf{I}_i|_{i=1}^c} \sum_{i=1}^c \left(\left\| \mathbf{A}\mathbf{I}_i \right\|_{S_p}^k \right)^2 + \lambda \sum_{i=1}^n \min\{ \left\| \mathbf{E}_i \right\|_2, \theta \}$$
$$+ \langle \mathbf{Y}, \mathbf{A} + \mathbf{E} - \mathbf{X} \rangle + \frac{\mu}{2} \| \mathbf{A} + \mathbf{E} - \mathbf{X} \|_F^2 + g\left(\Lambda, \mathbf{I}_i|_{i=1}^c \right), \tag{5}$$

where $\mathbf{Y}$ is a Lagrange multiplier, $\mu > 0$ is a penalty parameter, $\|\bullet\|_F$ is the Frobenius norm, and $g(\Lambda, \mathbf{I}_i|_{i=1}^c)$ encodes the constraints of $\mathbf{I}_i|_{i=1}^c$. We rewrite equation (5) as follows:

$$\min_{\mathbf{A},\mathbf{E},\mathbf{I}_i|_{i=1}^c} \sum_{i=1}^c \left(\left\| \mathbf{A}\mathbf{I}_i \right\|_{S_p}^k \right)^2 + \lambda \sum_{i=1}^n \min\{ \left\| \mathbf{E}_i \right\|_2, \theta \}$$
$$+ \frac{\mu}{2} \left\| \mathbf{A} + \mathbf{E} - \mathbf{X} + \frac{\mathbf{Y}}{\mu} \right\|_F^2 + g\left(\Lambda, \mathbf{I}_i|_{i=1}^c \right). \tag{6}$$

We divide equation (6) into three subproblems: optimizing $\mathbf{A}$ while fixing $\mathbf{E}$ and $\mathbf{I}_i|_{i=1}^c$, optimizing $\mathbf{E}$ while fixing $\mathbf{A}$ and $\mathbf{I}_i|_{i=1}^c$, and optimizing $\mathbf{I}_i|_{i=1}^c$ while fixing $\mathbf{A}$ and $\mathbf{E}$.

2.3.1. Fixing $\mathbf{E}$ and $\mathbf{I}_i|_{i=1}^c$ to Optimize $\mathbf{A}$. Equation (6) can be simplified to

$$\min_{\mathbf{A}} \sum_{i=1}^c \left(\left\| \mathbf{A}\mathbf{I}_i \right\|_{S_p}^k \right)^2 + \frac{\mu}{2} \| \mathbf{A} - \mathbf{B} \|_F^2, \tag{7}$$

where $\mathbf{B} = \mathbf{X} - \mathbf{E} + (\mathbf{Y}/\mu)$.

Lemma 1 (Araki-Lieb-Thirring [40, 41]). *For any positive semidefinite matrices $\mathbf{P}, \mathbf{Q} \in \mathbb{R}^{n \times n}$, $q > 0$, the following inequality holds when $0 \leq h < 1$:*

$$\mathrm{Tr}\left(\mathbf{P}^h \mathbf{Q}^h \mathbf{P}^h \right)^q \leq \mathrm{Tr}\left(\mathbf{P}\mathbf{Q}\mathbf{P} \right)^{hq}. \tag{8}$$

While for $h \geq 1$, the inequality is reversed.

Following $0 < k \leq 1$, $\|\mathbf{G}\|_{S_p}^p = \mathrm{Tr}((\mathbf{G}^T\mathbf{G})^{p/2})$ [28] and Lemma 1, the first term in equation (7) can be denoted as $\sum_{i=1}^c \mathrm{Tr}(\mathbf{I}_i(\mathbf{A}^T\mathbf{A})^k) = \sum_{i=1}^c (\|\mathbf{A}\mathbf{I}_i\|_{S_p}^k)^2$ since $\mathbf{I}_i|_{i=1}^c \subseteq \{0,1\}^{n \times n}$. According to $\sum_{i=1}^c \mathbf{I}_i = \mathbf{I}$, we convert the first term in equation (7) to

$$\sum_{i=1}^c \mathrm{Tr}\left(\mathbf{I}_i \left(\mathbf{A}^T\mathbf{A} \right)^k \right) = \mathrm{Tr}\left(\mathbf{I}_1 \left(\mathbf{A}^T\mathbf{A} \right)^k + \mathbf{I}_2 \left(\mathbf{A}^T\mathbf{A} \right)^k + \cdots + \mathbf{I}_c \left(\mathbf{A}^T\mathbf{A} \right)^k \right) = \mathrm{Tr}\left(\mathbf{A}^T\mathbf{A} \right)^k. \tag{9}$$

Then, equation (7) can be represented as

$$J_1 = \min_{\mathbf{A}} \mathrm{Tr}\left(\mathbf{A}^T\mathbf{A} \right)^k + \frac{\mu}{2} \| \mathbf{A} - \mathbf{B} \|_F^2. \tag{10}$$

Taking derivative w.r.t $\mathbf{A}$ and setting to zero, the above formula becomes

$$\frac{\partial J_1}{\partial \mathbf{A}} = 2\mathbf{A}\mathbf{H} + \mu(\mathbf{A} - \mathbf{B}) = 0, \tag{11}$$

where $\mathbf{H} = k(\mathbf{A}^T\mathbf{A})^{k-1}$. So, we can achieve the optimal $\mathbf{A}$:

$$\mathbf{A} = \mu\mathbf{B}(2\mathbf{H} + \mu\mathbf{I})^{-1}. \tag{12}$$

2.3.2. Fixing $\mathbf{A}$ and $\mathbf{I}_i|_{i=1}^c$ to Optimize $\mathbf{E}$. Here, we can denote equation (6) as

$$\min_{\mathbf{E}} \lambda \sum_{i=1}^n \min\{ \left\| \mathbf{E}_i \right\|_2, \theta \} + \frac{\mu}{2} \| \mathbf{E} - \mathbf{F} \|_F^2, \tag{13}$$

where $\mathbf{F} = \mathbf{X} - \mathbf{A} + (\mathbf{Y}/\mu)$. It can be easily verified that the derivative of equation (13) is equivalent to the derivative of

$$\min_{\mathbf{E}} \lambda \sum_{i=1}^n o_i \left\| \mathbf{E}_i \right\|_2^2 + \frac{\mu}{2} \| \mathbf{E} - \mathbf{F} \|_F^2, \tag{14}$$

where

$$o_i = \begin{cases} \dfrac{1}{2\left\| \mathbf{E}_i \right\|_2}, & \text{if } \left\| \mathbf{E}_i \right\|_2 < \theta; \\ \\ 0, & \text{otherwise.} \end{cases} \tag{15}$$

Equation (14) can be formulated as

$$J_2 = \min_{\mathbf{E}} \lambda \mathrm{Tr}\left(\mathbf{O}\mathbf{E}^T\mathbf{E} \right) + \frac{\mu}{2} \| \mathbf{E} - \mathbf{F} \|_F^2, \tag{16}$$

where $\mathbf{O}$ is a diagonal matrix with $\mathbf{O}_{ii} = o_i$. The problem of equation (16) can be optimized by using the iterative reweighted optimization strategy.

When fixing $\mathbf{O}$, taking derivative w.r.t $\mathbf{E}$ and setting it to zero, the above formulation can be written as

$$\frac{\partial J_2}{\partial \mathbf{E}} = 2\lambda\mathbf{E}\mathbf{O} + \mu(\mathbf{E} - \mathbf{F}) = 0. \tag{17}$$

So, we can obtain the optimal $\mathbf{E}$:

$$\mathbf{E} = \mu\mathbf{F}(2\lambda\mathbf{O} + \mu\mathbf{I})^{-1}. \tag{18}$$

When fixing $\mathbf{E}$, the updating rule for $\mathbf{O}$ is as follows:

$$\mathbf{O}_{ii} = \begin{cases} \dfrac{1}{2\left\| \mathbf{E}_i \right\|_2}, & \text{if } \left\| \mathbf{E}_i \right\|_2 < \theta; \\ \\ 0, & \text{otherwise.} \end{cases} \tag{19}$$

2.3.3. Fixing $\mathbf{A}$ and $\mathbf{E}$ to Optimize $\mathbf{I}_i|_{i=1}^c$. We can rewrite equation (6) as

$$J_3 = \min_{\mathbf{I}_i|_{i=1}^c} \sum_{i=1}^c \left(\left\| \mathbf{A}\mathbf{I}_i \right\|_{S_p}^k \right)^2 + g\left(\Lambda, \mathbf{I}_i|_{i=1}^c \right). \tag{20}$$

Taking derivative w.r.t $\mathbf{I}_i|_{i=1}^c$ and setting to zero, the above formulation can be written as

$$\frac{\partial J_3}{\partial \mathbf{I}_i|_{i=1}^c} = \sum_{i=1}^c 2\mathbf{A}^T \mathbf{L}_i \mathbf{A} \mathbf{I}_i + \frac{\partial g\left(\Lambda, \mathbf{I}_i|_{i=1}^c\right)}{\partial \mathbf{I}_i|_{i=1}^c} = 0, \qquad (21)$$

where $\mathbf{L}_i = k\|\mathbf{A}\mathbf{I}_i\|_{S_p}^k \left(\mathbf{A}\mathbf{I}_i^2\mathbf{A}^T\right)^{(k-2)/2}$.

Since $\mathbf{L}_i$ depends on $\mathbf{I}_i$, an iteration-based algorithm is used to obtain the solution of equation (21). Firstly, we calculate $\mathbf{L}_i$ by using the current solution of $\mathbf{I}_i$. If $\mathbf{L}_i$ is given, the solution of $\mathbf{I}_i$ to the following objective function will satisfy equation (21):

$$\min_{\mathbf{I}_i|_{i=1}^c} \quad \sum_{i=1}^c \mathrm{Tr}\left(\mathbf{I}_i^T \mathbf{A}^T \mathbf{L}_i \mathbf{A} \mathbf{I}_i\right),$$

$$\text{s.t.} \quad \mathbf{I}_i|_{i=1}^c \subseteq \{0,1\}^{n\times n}, \sum_{i=1}^c \mathbf{I}_i = \mathbf{I}. \qquad (22)$$

The current solution of $\mathbf{I}_i$ can be updated according to the optimal solution to equation (22).

Denote that $\mathbf{Z}_i = \mathbf{A}^T \mathbf{L}_i \mathbf{A}$, equation (22) can be written as

$$\min_{\mathbf{I}_i|_{i=1}^c} \sum_{i=1}^c \mathrm{Tr}\left(\mathbf{Z}_i \mathbf{I}_i\right),$$

$$\text{s.t.} \mathbf{I}_i|_{i=1}^c \subseteq \{0,1\}^{n\times n}, \sum_{i=1}^c \mathbf{I}_i = \mathbf{I}. \qquad (23)$$

Due to $\mathbf{I}_i|_{i=1}^c$ are $n\times n$ diagonal matrices, the above formulation becomes

$$\min_{r_{ci}\in\{0,1\}, \sum_{i=1}^c r_{ci}=1} \sum_{i=1}^c \sum_{i=c}^n \left(z_{ci} r_{ci}\right), \qquad (24)$$

where r_{ci} is the c-th diagonal element of matrix $\mathbf{I}_i$ and z_{ci} is the c-th diagonal element of matrix $\mathbf{Z}_i$. We can optimize equation (24) by

$$r_{ci} = \begin{cases} 1, & \text{if } i = \arg\min_l \left(z_{cl}\right), \\ 0, & \text{otherwise.} \end{cases} \qquad (25)$$

The algorithm to solve the problem of ORLRS is summarized in Algorithm 1.

2.4. Convergence Analysis.

In this section, the convergence analysis of the proposed algorithm will be proved.

Theorem 1. *At each iteration, the updating rule in Algorithm 1 for matrix $\mathbf{A}$ while fixing others will monotonically decrease the objective value in equation (4) when $0 < k \leq 1$.*

Proof. It can be verified that equation (12) is the solution to the following problem:

$$\min_{\mathbf{A}} \mathrm{Tr}\left(\mathbf{A}^T \mathbf{A} \mathbf{H}\right).$$

$$\text{s.t.} \mathbf{X} = \mathbf{A} + \mathbf{E}. \qquad (26)$$

Then, at the t iteration

$$\mathbf{A}_{t+1} = \arg\min_{\mathbf{A}} \mathrm{Tr}\left(\mathbf{A}^T \mathbf{A} \mathbf{H}_t\right). \qquad (27)$$

That is,

$$\mathrm{Tr}\left(\mathbf{A}_{t+1}^T \mathbf{A}_{t+1} \mathbf{H}_t\right)) \leq \mathrm{Tr}\left(\mathbf{A}_t^T \mathbf{A}_t \mathbf{H}_t\right). \qquad (28)$$

Equation (28) can be converted to

$$k\mathrm{Tr}\left(\mathbf{A}_{t+1}^T \mathbf{A}_{t+1}\left(\mathbf{A}_t^T \mathbf{A}_t\right)^{k-1}\right) \leq k\mathrm{Tr}\left(\mathbf{A}_t^T \mathbf{A}_t\left(\mathbf{A}_t^T \mathbf{A}_t\right)^{k-1}\right)), \qquad (29)$$

according to Lemma 2 in [38]. $\qquad\square$

Lemma 2. *For any positive definite matrices $\mathbf{P}, \mathbf{P}_t \in \mathbb{R}^{m\times m}$, the following inequality holds when $0 < p \leq 2$.*

$$\mathrm{Tr}\left(\mathbf{P}^{p/2}\right) - \frac{p}{2}\mathrm{Tr}\left(\mathbf{P}\mathbf{P}_t^{(p-2)/2}\right) \leq \mathrm{Tr}\left(\mathbf{P}_t^{p/2}\right) - \frac{p}{2}\mathrm{Tr}\left(\mathbf{P}_t\mathbf{P}_t^{(p-2)/2}\right), \qquad (30)$$

Note that, here, we set $0 < k \leq 1$, so equation (30) is equivalent to

$$\mathrm{Tr}\left(\mathbf{P}^k\right) - k\mathrm{Tr}\left(\mathbf{P}\mathbf{P}_t^{k-1}\right) \leq \mathrm{Tr}\left(\mathbf{P}_t^k\right) - k\mathrm{Tr}\left(\mathbf{P}_t\mathbf{P}_t^{k-1}\right), \qquad (31)$$

Then, we have

$$\mathrm{Tr}\left(\left(\mathbf{A}^T\mathbf{A}\right)^k\right) - k\mathrm{Tr}\left(\mathbf{A}^T\mathbf{A}\left(\mathbf{A}_t^T\mathbf{A}_t\right)^{k-1}\right)$$

$$\leq \mathrm{Tr}\left(\left(\mathbf{A}_t^T\mathbf{A}_t\right)^k\right) - k\mathrm{Tr}\left(\mathbf{A}_t^T\mathbf{A}_t\left(\mathbf{A}_t^T\mathbf{A}_t\right)^{k-1}\right)). \qquad (32)$$

Combining equations (29) and (32), we have

$$\mathrm{Tr}\left(\left(\mathbf{A}^T\mathbf{A}\right)^k\right) \leq \mathrm{Tr}\left(\left(\mathbf{A}_t^T\mathbf{A}_t\right)^k\right). \qquad (33)$$

That is to say,

$$\left(\|\mathbf{A}_{t+1}\|_{S_p}^k\right)^2 \leq \left(\|\mathbf{A}_t\|_{S_p}^k\right)^2. \qquad (34)$$

Thus, the updating rule for matrix $\mathbf{A}$ in Algorithm 1 will not increase the objective value of the problem in equation (10) at each iteration t when $0 < k \leq 1$.

Theorem 2. *At each iteration, the updating rule in Algorithm 1 for matrix $\mathbf{E}$ while fixing others will monotonically decrease the objective value in equation (4).*

Proof. We fist prepare the following lemma in [37]. $\qquad\square$

Lemma 3. *Given $s = \begin{cases} (1/2|e|), & \text{if } |e| < \theta, \\ 0, & \text{otherwise,} \end{cases}$, we have the following inequality:*

$$\min\left(|\tilde{e}|, \theta\right) - s\tilde{e}^2 \leq \min\left(|e|, \theta\right) - se^2. \qquad (35)$$

It can be verified that equation (18) is the solution to the following problem:

$$\min_{\mathbf{E}} \quad \mathrm{Tr}\left(\mathbf{O}\mathbf{E}^T\mathbf{E}\right),$$

$$\text{s.t.} \quad \mathbf{X} = \mathbf{A} + \mathbf{E}. \qquad (36)$$

Suppose the updated $\mathbf{E}$ in Algorithm 1 is $\tilde{\mathbf{E}}$ while fixing others. Since $\tilde{\mathbf{E}}$ is the optimal solution to equation (4), we have

---

Input: data matrix: $\mathbf{X} \in \mathbb{R}^{m \times n}$, number of subspace c, the low-rank constraint parameter k, balance parameter λ, threshold parameter θ.

Initialize: $\mathbf{A} = \mathbf{X}$, $\mathbf{E} = 0$, $\mathbf{Y} = 0$, $\mu = 10^{-6}$, $\rho = 1.1$, $\max_\mu = 10^{10}$, $\varepsilon = 10^{-8}$, $\mathbf{O}$ as the identity matrix, $\mathbf{I}_i|_{i=1}^c$ such that the discrete constraints in equation (16) are satisfied.

Output: the optimal $\mathbf{I}_i|_{i=1}^c$ for the i-th cluster.

while not converge **do**

(1) Fix the others and update $\mathbf{A}$: ① calculate $\mathbf{B} = \mathbf{X} - \mathbf{E} + (\mathbf{Y}/\mu)$, ② calculate $\mathbf{H} = k(\mathbf{A}^T\mathbf{A})^{k-1}$, ③update $\mathbf{A}$ by $\mathbf{A} = \mu\mathbf{B}(2\mathbf{H} + \mu\mathbf{I})^{-1}$.

(2) Fix the others and update $\mathbf{E}$: ① calculate $\mathbf{F} = \mathbf{X} - \mathbf{A} + (\mathbf{Y}/\mu)$, ② update $\mathbf{E}$ by $\mathbf{E} = \mu\mathbf{F}(2\lambda\mathbf{O} + \mu\mathbf{I})^{-1}$, ③ calculate
$\mathbf{O}_{ii} = \begin{cases} (1/2\|\mathbf{E}_i\|_2), & \text{if } \|\mathbf{E}_i\|_2 < \theta; \\ 0, & \text{otherwise.} \end{cases}$

(3) Fix the others and update $\mathbf{I}_i|_{i=1}^c$: ① calculate $\mathbf{L}_i = k\|\mathbf{A}\mathbf{I}_i\|_{S_p}^k (\mathbf{A}\mathbf{I}_i^2\mathbf{A}^T)^{(k-2)/2}$, ② calculate $\mathbf{Z}_i = \mathbf{A}^T\mathbf{L}_i\mathbf{A}$, ③ update $\mathbf{I}_i|_{i=1}^c$, where the c-th diagonal element r_{ci} of matrix $\mathbf{I}_i$ is updated by equation (33).

(4) Update the multiplier: $\mathbf{Y} = \mathbf{Y} + \mu(\mathbf{X} - \mathbf{A} - \mathbf{E})$.

(5) Update the parameter μ by $\mu = \min(\rho\mu, \max_\mu)$.

(6) Check the convergence condition
$\|\mathbf{X} - \mathbf{A} - \mathbf{E}\|_\infty < \varepsilon$.

end while

ALGORITHM 1: An efficient algorithm to solve the problem in equation (4).

$$\lambda\text{Tr}\left(\mathbf{O}\widetilde{\mathbf{E}}^T\widetilde{\mathbf{E}}\right) \le \lambda\text{Tr}\left(\mathbf{O}\mathbf{E}^T\mathbf{E}\right). \quad (37)$$

According to the definition of $\mathbf{O}_{ii}$ in equation (19) and Lemma 3, we have

$$\lambda\sum_{i=1}^n \min\left(\|\widetilde{\mathbf{E}}_i\|_2, \theta\right) - \lambda\sum_{i=1}^n \mathbf{O}_{ii}\|\widetilde{\mathbf{E}}_i\|_2^2$$
$$\le \lambda\sum_{i=1}^n \min\left(\|\mathbf{E}_i\|_2, \theta\right) - \lambda\sum_{i=1}^n \mathbf{O}_{ii}\|\mathbf{E}_i\|_2^2. \quad (38)$$

Summing over equations (37) and (38) at both sides, we can obtain

$$\lambda\sum_{i=1}^n \min\left(\|\widetilde{\mathbf{E}}_i\|_2, \theta\right) \le \lambda\sum_{i=1}^n \min\left(\|\mathbf{E}_i\|_2, \theta\right). \quad (39)$$

Therefore, at each iteration, the updating rule in Algorithm 1 for matrix $\mathbf{E}$ while fixing others will monotonically decrease the objective value in equation (4).

Theorem 3. *At each iteration, the updating rule in Algorithm 1 for $\mathbf{I}_i|_{i=1}^c$ while fixing others will monotonically decrease the objective value in equation (4) when $k = 1$.*

Proof. It can be easily verified that equation (25) is the solution to the following problem:

$$\min_{\mathbf{I}_i|_{i=1}^c} \sum_{i=1}^c \left(\|\mathbf{A}\mathbf{I}_i\|_{S_p}^k\right)^2,$$
$$\text{s.t.} \mathbf{I}_i|_{i=1}^c \subseteq \{0,1\}^{n\times n}, \sum_{i=1}^c \mathbf{I}_i = \mathbf{I}. \quad (40)$$

Assume the updated $\mathbf{I}_i$ in Algorithm 1 is $\widetilde{\mathbf{I}}_i$. Since $\widetilde{\mathbf{I}}_i$ is the optimal solution to the equation (22), we can have

$$\sum_{i=1}^c \text{Tr}\left(\widetilde{\mathbf{I}}_i^T \mathbf{A}^T\mathbf{L}_i\mathbf{A}\widetilde{\mathbf{I}}_i\right) \le \sum_{i=1}^c \text{Tr}\left(\mathbf{I}_i^T\mathbf{A}^T\mathbf{L}_i\mathbf{A}\mathbf{I}_i\right). \quad (41)$$

According to the definition of $\mathbf{L}_i$ in Algorithm 1, equation (41) can be written as

$$\sum_{i=1}^c \|\mathbf{A}\mathbf{I}_i\|_{S_p}^k \text{Tr}\left(\left(\mathbf{A}\mathbf{I}_i^2\mathbf{A}^T\right)^{(k-2)/2}\mathbf{A}^T\left(\widetilde{\mathbf{I}}_i^T\right)^2\mathbf{A}\right) \le \sum_{i=1}^c \left(\|\mathbf{A}\mathbf{I}_i\|_{S_p}^k\right)^2. \quad (42)$$

According to the Cauchy–Schwarz inequality, it can be proved that, when $p = 1$, we have

$$\sum_{i=1}^c \left(\|\mathbf{A}\widetilde{\mathbf{I}}_i\|_{S_p}^k\right)^2$$
$$\le \sum_{i=1}^c \text{Tr}\left(\left(\mathbf{A}\mathbf{I}_i^2\mathbf{A}^T\right)^{k/2}\right)\text{Tr}\left(\left(\mathbf{A}\mathbf{I}_i^2\mathbf{A}^T\right)^{(k-2)/2}\mathbf{A}^T\left(\widetilde{\mathbf{I}}_i^T\right)^2\mathbf{A}\right). \quad (43)$$

Thus, combining inequations (42) and (43), we can obtain

$$\sum_{i=1}^c \left(\|\mathbf{A}\widetilde{\mathbf{I}}_i\|_{S_p}^k\right)^2 \le \sum_{i=1}^c \left(\|\mathbf{A}\mathbf{I}_i\|_{S_p}^k\right)^2. \quad (44)$$

Equation (44) indicates that the updating rule in Algorithm 1 for $\mathbf{I}_i|_{i=1}^c$ while fixing others will monotonically decrease the objective value in equation (4) during the iteration until the algorithm converges when $k = 1$. In practice, the algorithm is also converged when $0 < k < 1$. If the objective function of equation (40) is changed to $\sum_{i=1}^c (\|\mathbf{A}\mathbf{I}_i\|_{S_p}^k)^d$, $(d > 2)$ ($\mathbf{L}_i$ in Algorithm 1 becomes $(dk/2)(\|\mathbf{A}\mathbf{I}_i\|_{S_p}^k)^{d-1}(\mathbf{A}\mathbf{I}_i^2\mathbf{A}^T)^{(k-2)/2}$), the convergence is also observed [28].

As a result, the objective of equation (4) is nonincreasing under the updates of $\mathbf{A}$, $\mathbf{E}$, and $\mathbf{I}_i|_{i=1}^c$ according to Theorems 1–3, respectively. Therefore, the iteratively updating Algorithm 1 converges to a local optimal. □

2.5. Complexity Analysis. In Algorithm 1, the most complicated calculations are $\mathbf{L}_i = k\|\mathbf{AI}_i\|_{S_p}^k (\mathbf{AI}_i^2 \mathbf{A}^T)^{(k-2)/2}$ and $\mathbf{Z}_i = \mathbf{A}^T \mathbf{L}_i \mathbf{A}$ in Step 3. We suppose $m > n$ in the low-rank representation matrix $\mathbf{A} \in \mathbb{R}^{m \times n}$. Firstly, $\mathbf{L}_i$ needs to be computed. Denoting the SVD of $\mathbf{AI}_i$ is $\mathbf{U\Sigma V}^T$. Computing $\|\mathbf{AI}_i\|_{S_p}^k$ needs SVD of $\mathbf{AI}_i$, which takes $O(mn^2)$. $(\mathbf{AI}_i^2 \mathbf{A}^T)^{(k-2)/2}$ can be decomposed as $\mathbf{U\Sigma}^{k-2} \mathbf{U}^T$ by SVD, which takes $O(m^3)$. So, computing $\mathbf{L}_i$ takes $O(m^3)$ and computing $\mathbf{L}_i|_{I=1}^c$ takes $O(m^3 c)$, where c is the number of clusters. For $\mathbf{Z}_i$, we only need to compute the diagonal elements, which takes $O(m^3)$. And, computing $\mathbf{Z}_i|_{I=1}^c$ takes $O(m^3 c)$. In summary, the computational complexity of Algorithm 1 is $O(m^3 ct)$, where t is the iteration number.

3. Results and Discussion

We test ORLRS on six publicly available gene expression data sets, i.e., Leukemia [42], DLBCL [43], Colon cancer [44], Brain_Tumor1 [43], Brain_Tumor2 [43], and 9_Tumors [43].

Following [28, 45–47], clustering accuracy (ACC) is a widely used evaluation method for tumor clustering. Given a data point $\mathbf{x}_i$, suppose N_i as the target label and T_i as the truth label. ACC can be denoted as [45]

$$\text{ACC} = \frac{\sum_{i=1}^n \varphi[T_i, \text{map}(N_i)]}{n}, \qquad (45)$$

where $\varphi(x, y) = 0$ if $x \neq y$ and $\varphi(x, y) = 1$ if $x = y$, $\text{map}(N_i)$ maps N_i to the equivalent label from the raw data and n is the number of tumor samples.

We also evaluate the clustering performance by normalized mutual information (NMI) [48]. NMI is defined as

$$\text{NMI} = \frac{M(S, C)}{((H(S) + H(C))/2)}, \qquad (46)$$

where $M(S, C)$ is the mutual information function between the true class label C and the clustering label S and $H(\cdot)$ is the entropy function. The larger the NMI value is, the better the clustering result is.

3.1. Gene Expression Data Sets. A brief introduction of six gene expression data sets is presented, and the detailed information of these datasets is summarized in Table 1.

Leukemia data contain 25 cases of AML and 47 cases of ALL. It is packaged into a 7129×72 matrix [42].

DLBCL data consist of 5469 genes and 77 samples. These samples include 58 patients of diffuse large B-cell lymphoma (DLBCL) and 19 patients of follicular lymphomas (FL) [43].

The colon cancer data [44] consists of a matrix that includes 2000 genes and 62 tissues. These tissues are divided into 22 normal and 40 colon tumor samples

Brain_Tumor1 data set consists of 5920 genes in 90 patient samples. These samples contain 5 types of histological diagnoses, i.e., 60 cases of

medulloblastoma, 10 cases of malignant glioma, 10 cases of atypical teratoid/rhabdoid tumors (AT/RTs), 4 cases of normal cerebellum, and 6 cases of primitive neuroectodermal tumors (PNETs).

The Brain_Tumor2 data set contains 10367 genes in 50 samples. It contains 4 types of malignant glioma, i.e., classic glioblastomas (CG), classic anaplastic oligodendrogliomas (CAO), nonclassic glioblastomas (NCG), and nonclassic anaplastic oligodendrogliomas (NCAO) [34].

9_Tumors data set integrates 9 tumor types to develop a genomics-based approach to the prediction of drug response. It contains 5726 genes in 60 samples. The number of samples of 9 tumor types is shown as follows: 9 samples of non-small-cell carcinoma (NSCLC), 7 samples of colon cancer, 8 samples of breast cancer, 6 samples of ovary cancer, 6 samples of leukemia, 8 samples of renal cancer, 8 samples of melanoma, 2 samples of prostate cancer, and 6 samples central nervous system cancer (CNS).

3.2. Comparison Algorithms. We compare LRS [28], Ext-LRR [32], RPCA [3], PLRR [47], robust LRR [33], LatLRR [49], robust NMF [50], and K-means [51] with the proposed method for tumor clustering. In these methods, LRS is the basic version of our method to implement the one-step clustering, and Ext-LRR is a simpler and more effective extension work compared with LRS; RPCA is a classic robust learning algorithm; PLRR (projection LRR) is one of the latest subspace clustering methods for tumor sample clustering; Robust LRR and LatLRR are the best state-of-art low-rank subspace segmentation algorithms; robust NMF is a classic NMF-based method and is widely used for tumor clustering. *K*-means is the most commonly used clustering method and is embedded into many methods including PLRR, robust LRR, and LatLRR to achieve better performance. Since our proposed method is a novel one-step robust low-rank subspace clustering model, we choose these methods as our comparison algorithms.

3.3. Parameter Setting. Since gene expression data have the characteristics of high-dimensional and small samples, we use PCA to perform dimensionality reduction. And, we use the K-means method to initialize $\mathbf{I}_i|_{i=1}^c$ in the proposed ORLRS. Here, three parameters, i.e., threshold parameter θ, balance parameter λ, and low-rank constraint parameter k, need to be determined. In the experiment, we investigated one parameter by fixing the other two parameters. Since the initialization of $\mathbf{I}_i|_{i=1}^c$ will bring some uncertainty, the proposed ORLRS method run 100 times, and the average of the accuracies of 100 times is reported. The choices of parameters in the following are heuristic and might not be the best for tumor clustering.

3.3.1. Determination of Threshold Parameter θ. In the ORLRS model, data outliers are not heuristically determined based on the magnitude. They are selected during the

TABLE 1: Summary of tumor gene expression data sets.

| Data set | Diagnostic task | Number of | | |
		Genes	Samples	Classes
Leukemia	AML, ALL	7129	72	2
DLBCL	DLBCL, FL	5469	77	2
Colon cancer	Tumor, normal	2000	62	2
Brain_Tumor1	Medulloblastoma, malignant glioma, normal cerebellum, AT/RTs, PNETs	5920	90	5
Brain_Tumor2	CG, CAO, NCG, NCAO	10367	50	4
9_Tumors	NSCLC, colon, breast, ovary, renal, leukemia, melanoma, prostate, CNS	5726	60	9
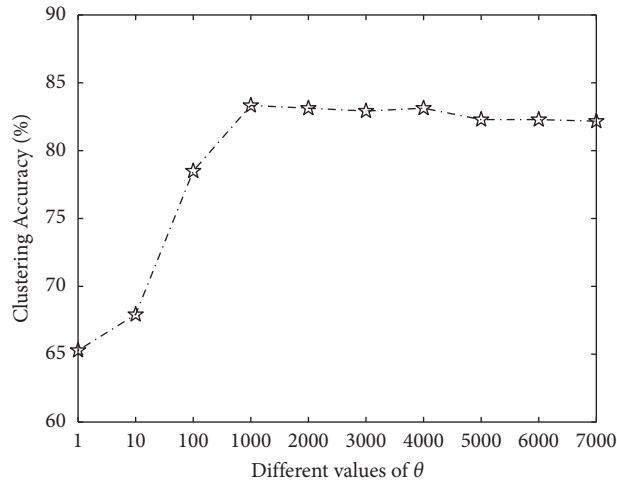
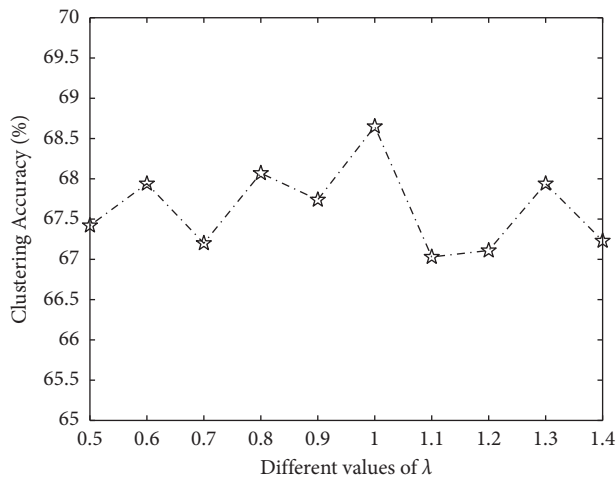

(a)
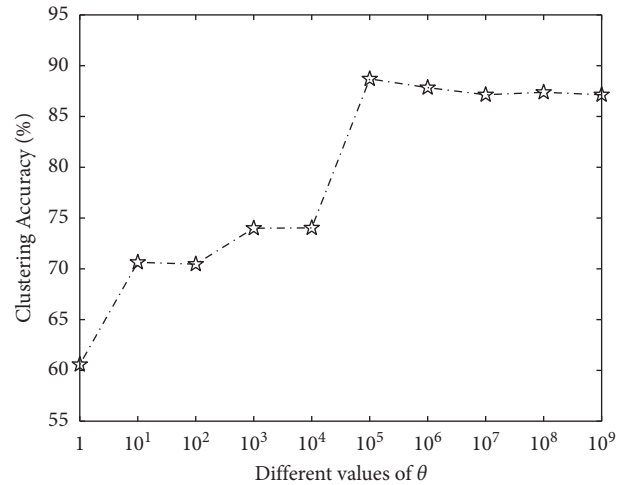
(b)
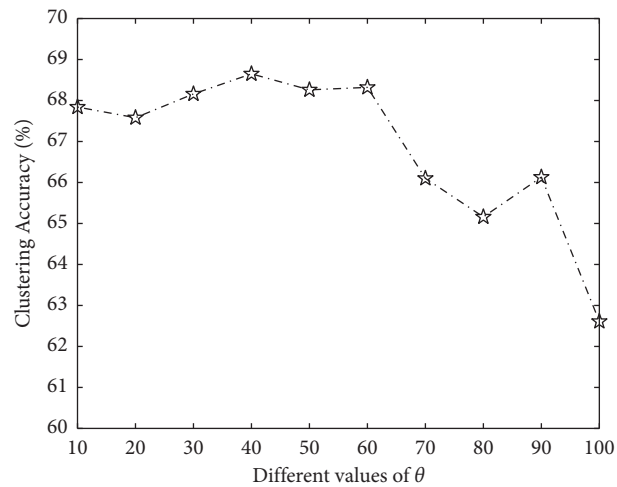
(c)
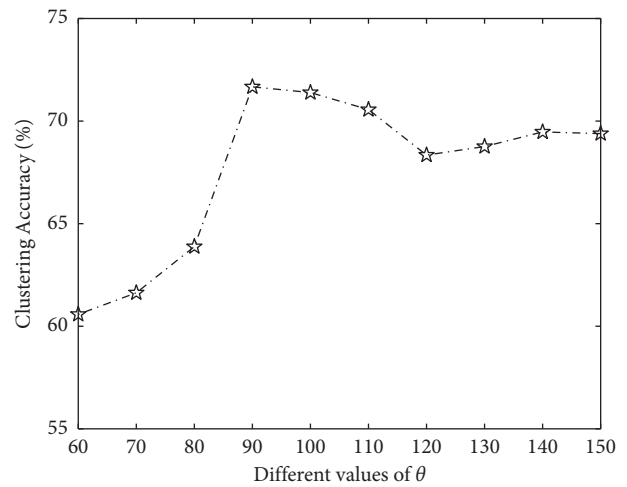
(d)

FIGURE 1: Continued.
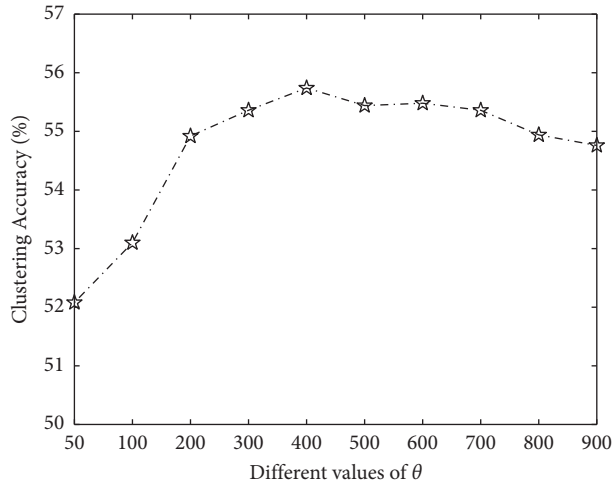
(e)
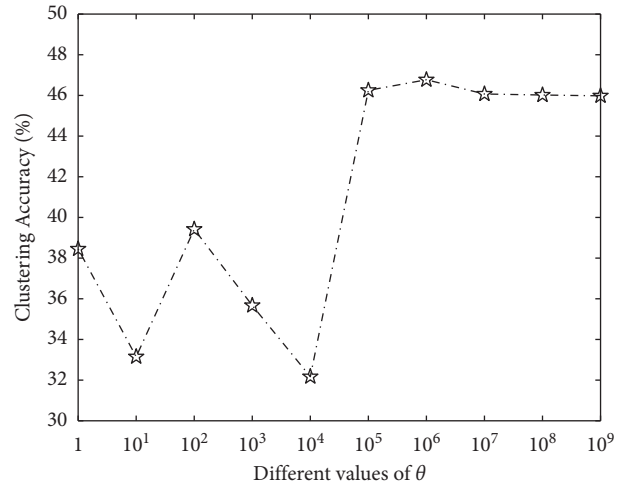
(f)

Figure 1: Clustering accuracies of the proposed ORLRS on six data sets with different values of θ. (a) Leukemia. (b) DLBCL. (c) Colon. (d) Brain_Tumor1. (e) Brain_Tumor2. (f) 9_Tumors.
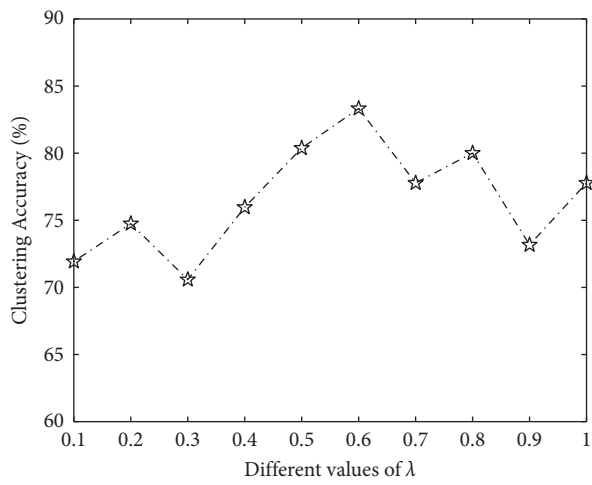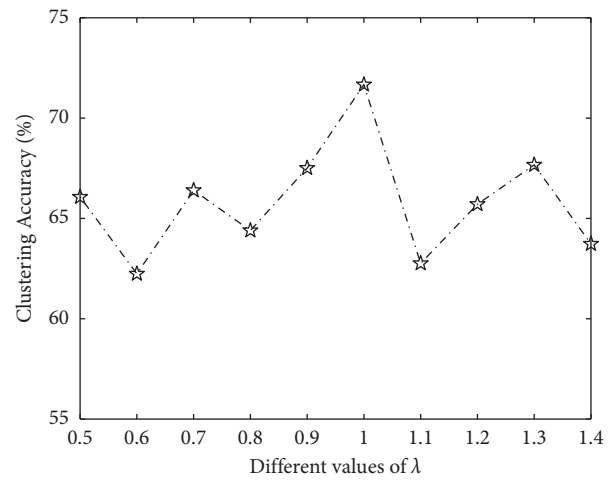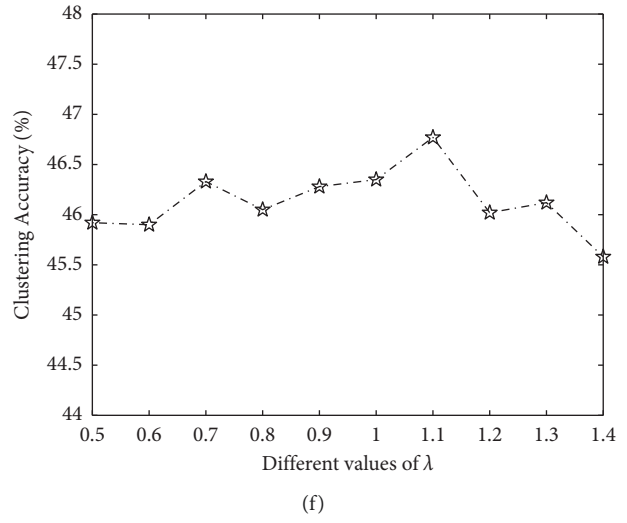

(a)
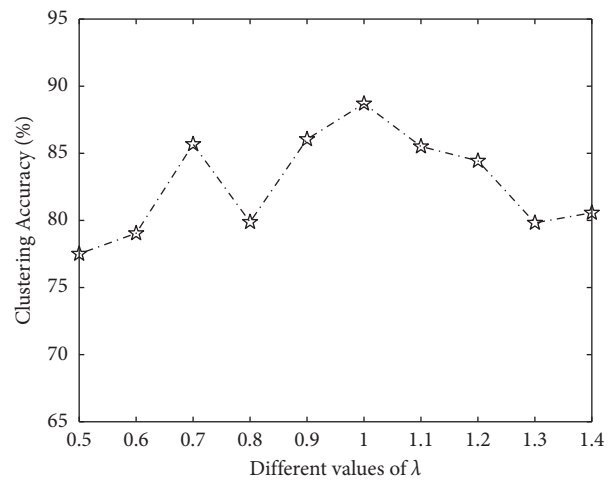
(b)

(c)

(d)

Figure 2: Continued.

(e)

(f)

FIGURE 2: Clustering accuracies of the proposed ORLRS on six data sets with different values of λ. (a) Leukemia. (b) DLBCL. (c) Colon. (d) Brain_Tumor1. (e) Brain_Tumor2. (f) 9_Tumors.
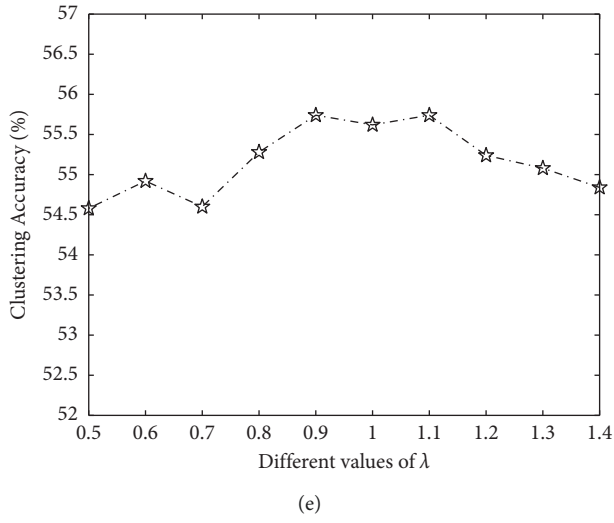

(a)

(b)

(c)

(d)

FIGURE 3: Continued.

(e)

(f)

FIGURE 3: Clustering accuracies of the proposed ORLRS on six data sets with different values of k. (a) Leukemia. (b) DLBCL. (c) Colon. (d) Brain_Tumor1. (e) Brain_Tumor2. (f) 9_Tumors.
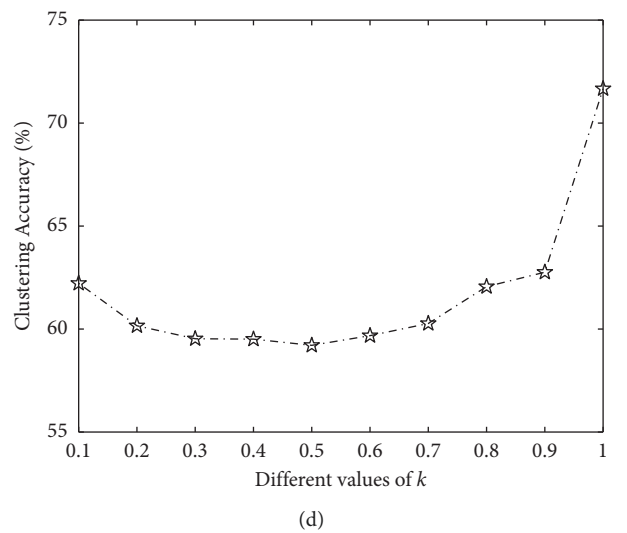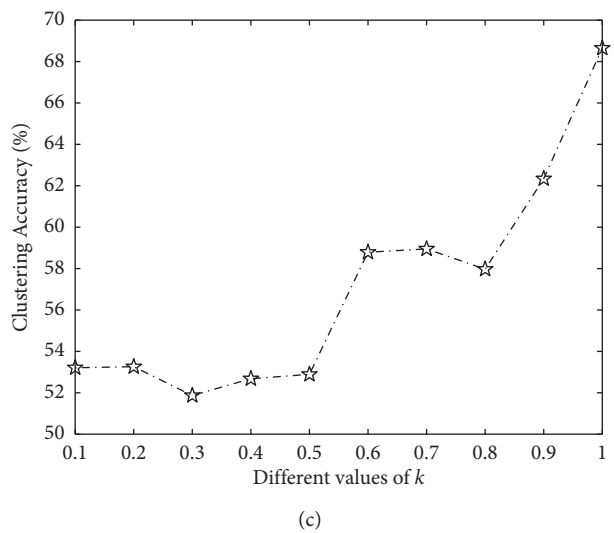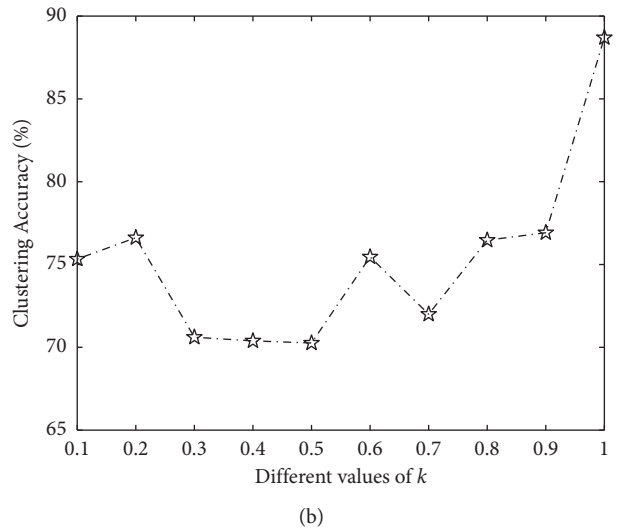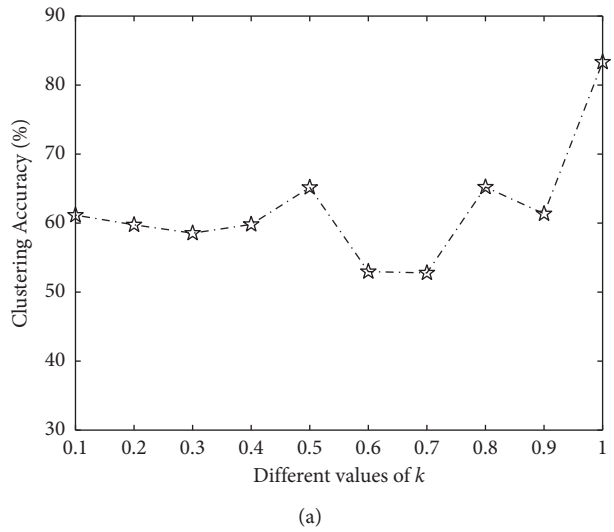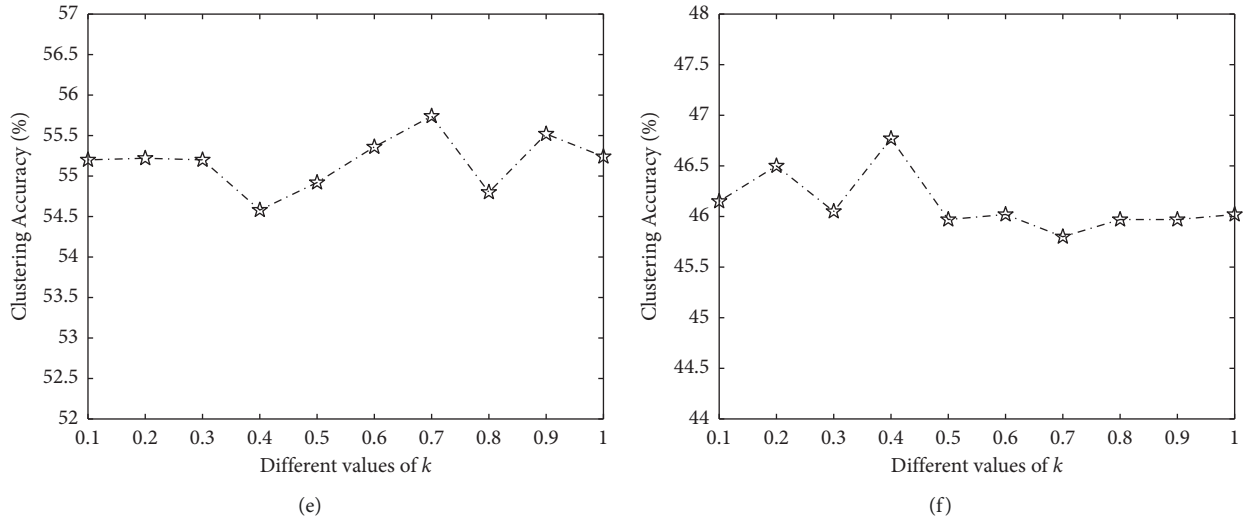
TABLE 2: Clustering accuracy (%) and standard error (%) of different methods on six gene expression data sets.

Data sets	Methods								
	ORLRS	LRS	PLRR	Robust LRR	LatLRR	Robust NMF	K-means	RPCA	Ext-LRR
Leukemia	**83.33 ± 0.00**	74.63 ± 2.14	74.57 ± 2.56	63.89 ± 0.00	66.67 ± 0.00	66.25 ± 5.03	68.06 ± 2.93	46.69 ± 0.00	65.28 ± 0.00
DLBCL	**88.70 ± 3.50**	76.62 ± 0.00	79.35 ± 0.41	76.62 ± 0.00	76.62 ± 0.00	68.96 ± 7.71	68.83 ± 0.00	57.49 ± 0.00	73.64 ± 2.87
Colon	**74.63 ± 2.14**	55.16 ± 1.59	72.79 ± 2.87	60.81 ± 1.09	58.06 ± 0.00	70.65 ± 7.25	53.71 ± 2.12	39.71 ± 0.00	64.52 ± 0.00
Brain_Tumor1	**71.67 ± 2.30**	60.22 ± 4.53	69.22 ± 10.32	68.89 ± 0.00	69.69 ± 2.19	66.11 ± 2.58	43.78 ± 1.83	57.67 ± 0.00	67.00 ± 4.86
Brain_Tumor2	**55.74 ± 2.93**	55.36 ± 2.51	54.20 ± 2.36	44.00 ± 0.00	50.00 ± 0.00	34.00 ± 2.98	39.20 ± 1.03	46.00 ± 0.00	48.40 ± 4.60
9_Tumors	46.77 ± 4.62	45.85 ± 2.90	41.67 ± 1.05	38.67 ± 1.05	43.33 ± 0.00	35.50 ± 3.60	45.52 ± 3.16	33.33 ± 0.00	31.67 ± 2.83

optimization process. The data outliers may be distinct at different iterations (with the same thresholding parameter), while we iteratively optimize the objective function of ORLRS method. When the algorithm converges, likely correct extreme data outliers can be found. So, we just need to determine one value of θ for each data set.

Figure 1 presents the results of ORLRS with different θ. Since the gene expression levels in different data are very different, the values of extreme data outliers are also very different. So, the value of θ has a large range in six data sets. From Figure 1, we can observe that ORLRS can obtain the best performance in the case of $\theta = \{10^3, 10^5, 40, 90, 400, 10^6\}$ in Leukemia data, DLBCL data, Colon cancer data, Brain_Tumor1 data, Brain_Tumor2 data, and 9_Tumors data, respectively. The results indicate that the value of θ should be determined appropriately. If the value of θ is too large, we will miss some extreme outliers. If the value of θ is too small, some important information may be removed, thereby affecting the clustering performance.

3.3.2. Determination of Balance Parameter λ. Figure 2 presents the results of ORLRS with different λ. ORLRS can obtain the best results in the case of $\lambda = \{0.6, 1, 1, 1, 0.9, 1.1\}$ in Leukemia data, DLBCL data, Colon cancer data, Brain_Tumor1 data, Brain_Tumor2 data,

and 9_Tumors data, respectively. According to the experimental results in each data set, before reaching the best results, the clustering accuracies showed an overall upward trend when λ increases; after achieving the best results, the clustering accuracies showed an overall downward trend when λ increases. So, we suggest a rough range $0.1 \leq \lambda \leq 2$ on the choice of λ.

3.3.3. Determination of Schatten P-Norm Parameter k. Since the algorithm is converged for the Schatten p-norm parameter $0 < k \leq 1$, we determine the value of k in this range. Figure 3 presents the results of ORLRS with different k. ORLRS can achieve the best performance in the case of $k = \{1, 1, 1, 1, 0.7, 0.4\}$ in Leukemia data, DLBCL data, Colon cancer data, Brain_Tumor1 data, Brain_Tumor2 data, and 9_Tumors data, respectively. So, a general guidance is given $0 < k \leq 1$ on the choice of k.

3.4. Experimental Results. In this section, experimental results of our proposed method and six comparison algorithms, i.e., LRS, Ext-LRR, RPCA, PLRR, robust LRR, LatLRR, robust NMF, and K-means, are reported. ORLRS, LRS, and Ext-LRR use K-means to initialize the indicator matrix $\mathbf{I}_i|_{i=1}^c$. PLRR, robust LRR, and LatLRR use the normalized cuts method to segment data, which cluster data

Table 3: NMI and standard error of different methods on five gene expression data sets.

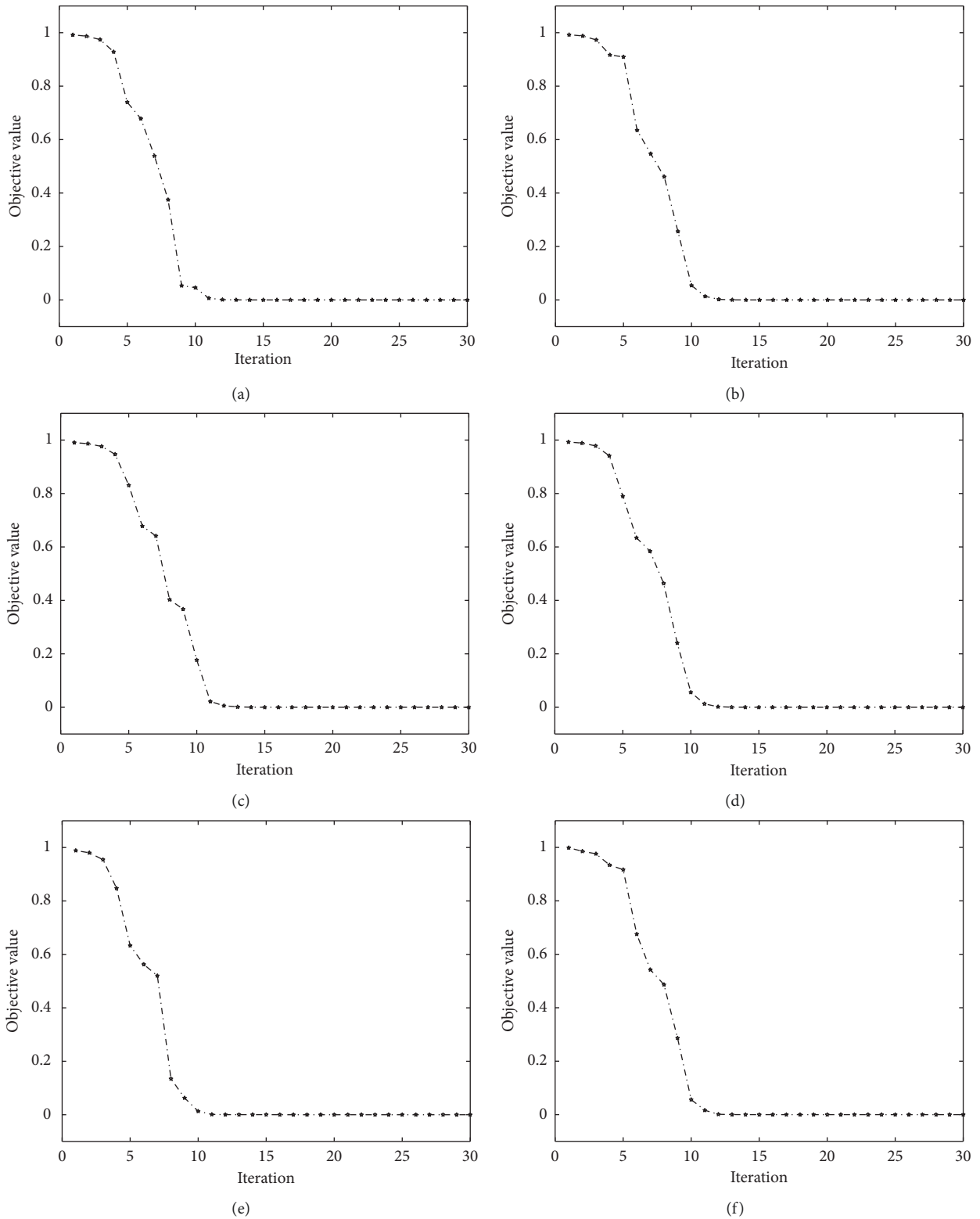| Data sets	Methods								
	ORLRS	LRS	PLRR	Robust LRR	LatLRR	Robust NMF	K-means	RPCA	Ext-LRR
Leukemia	0.1347 ± 0.0928	0.1025 ± 0.0466	0.1229 ± 0.0586	0.0684 ± 0.0000	0.0684 ± 0.0000	0.0920 ± 0.0529	0.1250 ± 0.0534	0.0275 ± 0.0000	0.1042 ± 0.0778
DLBCL	0.1387 ± 0.1113	0.1104 ± 0.0209	0.1208 ± 0.1020	0.0937 ± 0.0000	0.0937 ± 0.0000	0.0188 ± 0.0000	0.1150 ± 0.0141	0.0188 ± 0.0000	0.0853 ± 0.1617
Brain_Tumor1	0.3854 ± 0.0743	0.3045 ± 0.0538	0.3689 ± 0.0773	0.2914 ± 0.0146	0.2864 ± 0.0464	0.3254 ± 0.0786	0.2854 ± 0.0534	0.0362 ± 0.0000	0.2846 ± 0.0675
Brain_Tumor2	0.5486 ± 0.0431	0.5407 ± 0.0461	0.5415 ± 0.0451	0.3668 ± 0.0257	0.2645 ± 0.0323	0.2935 ± 0.0565	0.5208 ± 0.0370	0.1296 ± 0.0000	0.2913 ± 0.0867
9_Tumors	0.4082 ± 0.0463	0.4294 ± 0.0421	0.3885 ± 0.0664	0.3832 ± 0.0424	0.3854 ± 0.0497	0.2593 ± 0.0487	0.4204 ± 0.0222	0.2839 ± 0.000	0.2923 ± 0.0246

Figure 4: The convergence of ORLRS on six data sets. (a) Leukemia, (b) DLBCL, (c) Colon, (d) Brain_Tumor1, (e) Brain_Tumor2, (f) 9_Tumors.

points by using the K-means method. For robust NMF method, we initial the coefficient matrix and basis matrix randomly. To avoid randomness, we run all methods 100 times, and the mean and standard error results of the clustering accuracies of 100 times are shown in Table 2. The best result of each data is indicated in bold.

Based on the results reported in Table 2, we have the following observations and discussions. ORLRS extends

TABLE 4: The running time of ORLRS on six gene expression data sets.

Data sets	Leukemia	DLBCL	Colon	Brain_Tumor1	Brain_Tumor2	9_Tumors
Running time (second)	652.35	358.36	121.73	385.64	990.87	370.34

LRS by adding a noise matrix into the objective function to enhance the robustness, which contributes to the observation that ORLRS outperforms LRS. From the results shown in Table 2, it can be observed that ORLRS achieves generally 8%–19% higher performances than LRS in terms of the clustering accuracy on four data sets, i.e., Leukemia data, DLBCL data, Colon cancer data, and Brain_Tumor1 data. On Brain_Tumor2 and 9_Tumors data sets, ORLRS has a slightly better performance than LRS. ORLRS has better results than Ext-LRR on all datasets. Compared to the three classical low-rank based methods, PLRR, robust LRR, and LatLRR, the clustering accuracy of ORLRS is 1%–9% higher on all six data. The main reason is that we use capped norm to remove the extreme outliers in the noise matrix and Schatten p-norm to better approximate the low-rank representation. Compared with traditional clustering methods, RPCA, robust NMF, and K-means, ORLRS achieves outstanding results on all of the six data sets.

The NMI results on five gene expression data sets are shown in Table 3. The best result of each data is indicated in bold. Due to the NMI results of all the methods on colon data is less than 0.1, we only reported the results of remain five data sets. From Table 3, we can observe that ORLRS has better results on all the five data sets than PLRR, robust LRR, LatLRR, robust NMF, RPCA, and Ext-LRR. Except on 9_Tumors data set, our method outperforms LRS and K-means on the other four data sets.

3.5. Convergence Curves and Running Time. We plotted the convergence curves of our ORLRS on different datasets. The convergence curves can be found in Figure 4. It shows that our method can converge around the 10-th iteration on all six data sets. In Table 4, we also reported the running time of ORLRS on six gene expression data sets without dimensionality reduction by PCA. We implement our experiment with MATLAB R2020b on an ordinary computer, which is configured with Intel i9-10900 KF (up to 3.70 GHz) cores, 8 GB RAM, and Windows 10 operating system.

4. Conclusions

In this paper, a novel one-step robust low-rank subspace clustering method (ORLRS) is proposed for tumor clustering, where the gene expression data set is represented by a low-rank matrix and a noise matrix. By using the Schatten p-norm and discrete constraint, low-rank representation of each subspace can be well obtained. Different from traditional low-rank-based methods, such as LRR and LatLRR, ORLRS learns indicators directly and perform clustering process in one step by using the discrete constraint. Capped norm is used to improve the robustness of ORLRS since it can effectively remove the extreme data outliers in the noise

matrix. Furthermore, we propose an efficient algorithm to solve the proposed subspace clustering model, and the convergence of the proposed algorithm is proved. We thus can discover the clusters of tumor data depending on the optimal cluster indicators. We tested the proposed ORLRS method on six tumor data. The results are proved that ORLRS is an excellent method for clustering tumor sample.

There remain several interesting directions for future work. First, it might be better to learn a dictionary for ORLRS since some low-rank subspace segmentation methods achieve significant improvements by learning a dictionary. Second, ORLRS may be extended to solve other problems, such as matrix recovery and classification. Third, ORLRS may be employed in other applications, such as gene clustering and coclustering.

Data Availability

The data used to support the findings of this study are available from the first author upon request.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Acknowledgments

References

[1] M. J. Heller, "DNA microarray technology: devices, systems, and applications," *Annual Review of Biomedical Engineering*, vol. 4, no. 1, pp. 129–153, 2002.

[2] J.-P. Brunet, P. Tamayo, T. R. Golub, and J. P. Mesirov, "Metagenes and molecular pattern discovery using matrix factorization," *Proceedings of the National Academy of Sciences*, vol. 101, no. 12, pp. 4164–4169, 2004.

[3] E. J. Candes, X. Li, Y. Ma, and J. Wright, "Robust principal component analysis?" *Journal of the ACM*, vol. 58, no. 3, 2009.

[4] W. Hua and L. Mo, "Clustering ensemble model based on self-organizing map network," *Computational Intelligence and Neuroscience*, vol. 2020, Article ID 2971565, 11 pages, 2020.

[5] G. Liu, Z. Lin, and Y. Yu, "Robust subspace segmentation by low-rank representation," *Proceedings of the 27th International Conference on Machine Learning*, vol. 3, pp. 663–670, 2010.

[6] J. Zhang and Z. Ma, "Hybrid fuzzy clustering method based on FCM and enhanced logarithmical PSO (ELPSO)," *Computational Intelligence and Neuroscience*, vol. 2020, Article ID 1386839, 12 pages, 2020.

[7] X. Chang, F. Nie, S. Wang, Y. Yang, X. Zhou, and C. Zhang, "Compound rank- k projections for bilinear analysis," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 27, no. 7, pp. 1502–1513, 2016.

[8] D. Yuan, X. Chang, P.-Y. Huang, Q. Liu, and Z. He, "Self-supervised deep correlation tracking," *IEEE Transactions on Image Processing*, vol. 30, pp. 976–985, 2021.

[9] H. Wang, Z. Li, Y. Li, B. B. Gupta, and C. Choi, "Visual saliency guided complex image retrieval," *Pattern Recognition Letters*, vol. 130, pp. 64–72, 2020.

[10] C. H. Chun-Hou Zheng, L. Lei Zhang, V. T. Ng, S. C. Chi Keung Shiu, and D. S. De-Shuang Huang, "Molecular pattern discovery based on penalized matrix decomposition," *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, vol. 8, no. 6, pp. 1592–1603, 2011.

[11] J. Qiang, W. Ding, M. Kuijjer, J. Quackenbush, and P. Chen, "Clustering sparse data with feature correlation with application to discover subtypes in cancer," *IEEE Access*, vol. 8, pp. 67775–67789, 2020.

[12] M. B. Eisen, P. T. Spellman, P. O. Brown, and D. Botstein, "Cluster analysis and display of genome-wide expression patterns," *Proceedings of the National Academy of Sciences*, vol. 95, no. 25, pp. 14863–14868, 1998.

[13] C. M. Perou, T. Sørlie, M. B. Eisen et al., "Molecular portraits of human breast tumours," *Nature*, vol. 490, no. 7418, pp. 61–70, 2000.

[14] C. Scheidegger, L. Sigg, and R. Behra, "Molecular classification of cancer: class discovery and class prediction by gene expression monitoring," *Brain Research*, vol. 501, no. 2, pp. 205–214, 1999.

[15] C. H. Chun-Hou Zheng, D. S. De-Shuang Huang, L. Lei Zhang, and X. Z. Xiang-Zhen Kong, "Tumor clustering using nonnegative matrix factorization with gene selection," *IEEE Transactions on Information Technology in Biomedicine*, vol. 13, no. 4, pp. 599–607, 2009.

[16] Y. Gao and G. Church, "Improving molecular cancer class discovery through sparse non-negative matrix factorization," *Bioinformatics*, vol. 21, no. 21, pp. 3970–3975, 2005.

[17] I. T. Jolliffe, "Principal component analysis," *Journal of Marketing Research*, vol. 87, no. 100, p. 513, 2002.

[18] K. Y. Yeung and W. L. Ruzzo, "Principal component analysis for clustering gene expression data," *Bioinformatics*, vol. 17, no. 9, pp. 763–774, 2001.

[19] H.-Q. Wang, D.-S. Huang, X.-M. Zhao, and X. Huang, "A novel clustering analysis based on PCA and SOMs for gene expression patterns," *Lecture Notes in Computer Science*, vol. 3174, pp. 476–481, 2004.

[20] C. Meng, O. A. Zeleznik, G. G. Thallinger, B. Kuster, A. M. Gholami, and A. C. Culhane, "Dimension reduction techniques for the integrative analysis of multi-omics data," *Briefings in Bioinformatics*, vol. 17, no. 4, pp. 628–641, 2016.

[21] N. Yu, M.-J. Wu, J.-X. Liu, C.-H. Zheng, and Y. Xu, "Correntropy-based hypergraph regularized nmf for clustering and feature selection on multi-cancer integrated data," *IEEE Transactions on Cybernetics*, vol. 51, no. 8, pp. 3952–3963, 2021.

[22] C.-N. Jiao, Y.-L. Gao, N. Yu, J.-X. Liu, and L.-Y. Qi, "Hypergraph regularized constrained nmf for selecting differentially expressed genes and tumor classification," *IEEE Journal of Biomedical and Health Informatics*, vol. 24, no. 10, pp. 3002–3011, 2020.

[23] C. Wang, Y.-L. Gao, X.-Z. Kong, J.-X. Liu, and C.-H. Zheng, "Unsupervised cluster analysis and gene marker extraction of scrna-seq data based on non-negative matrix factorization,"

[24] R. Vidal, "Subspace clustering," *IEEE Signal Processing Magazine*, vol. 28, no. 2, pp. 52–68, 2011.

[25] J. Ho, M.-H. Yang, J. Lim, K.-C. Lee, and D. Kriegman, "Clustering appearances of objects under varying illumination conditions," *IEEE Computer Society Conference on Computer Vision & Pattern Recognition*, vol. 1, p. I, 2003.

[26] Y. Cui, C.-H. Zheng, and J. Yang, "Identifying subspace gene clusters from microarray data using low-rank representation," *PLoS One*, vol. 8, no. 3, Article ID e59377, 2013.

[27] E. Elhamifar and R. Vidal, "Sparse subspace clustering," *Proceedings of IEEE Conference on Computer Vision & Pattern Recognition*, vol. 35, no. 11, pp. 2790–2797, 2009.

[28] F. Nie and H. Huang, "Subspace clustering via new low-rank model with discrete group structure constraint," in *Proceedings of the International Joint Conference on Artificial Intelligence*, pp. 1874–1880, New York, NY, USA, July 2016.

[29] B. Nasihatkon and R. Hartley, "Graph connectivity in sparse subspace clustering," in *Proceedings of the IEEE Conference on Computer Vision & Pattern Recognition*, pp. 2137–2144, Colorado Springs, CO, USA, June 2011.

[30] C. Tang, X. Zhu, X. Liu et al., "Learning a joint affinity graph for multiview subspace clustering," *IEEE Transactions on Multimedia*, vol. 21, no. 7, pp. 1724–1736, 2019.

[31] C. Tang, X. Liu, X. Zhu et al., "Feature selective projection with low-rank embedding and dual laplacian regularization," *IEEE Transactions on Knowledge and Data Engineering*, vol. 32, no. 9, 2020.

[32] F. Nie, W. Chang, Z. Hu, and X. Li, "Robust subspace clustering with low-rank structure constraint," *IEEE Transactions on Knowledge and Data Engineering*, p. 1. In press, 2020.

[33] G. Liu, Z. Lin, S. Yan, J. Sun, Y. Yu, and Y. Ma, "Robust recovery of subspace structures by low-rank representation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 1, pp. 171–184, 2013.

[34] H. Xu, C. Caramanis, and S. Sanghavi, "Robust PCA via outlier pursuit," *IEEE Transactions on Information Theory*, vol. 58, no. 5, pp. 3047–3064, 2010.

[35] W. Siming and L. Zhouchen, "Analysis and improvement of low rank representation for subspace segmentation," 2011, https://arxiv.org/abs/1107.1561.

[36] H. Gao, F. Nie, W. Cai, and H. Huang, "Robust capped norm nonnegative matrix factorization: capped norm NMF," in *Proceedings of the 24th ACM International on Conference on Information and Knowledge Management*, pp. 871–880, Toronto, Canada, October 2015.

[37] F. Nie, Z. Huo, and H. Huang, "Joint capped norms minimization for robust matrix recovery," in *Proceedings of the 26th International Joint Conference on Artificial Intelligence*, pp. 2557–2563, Melbourne, Australia, August 2017.

[38] F. Nie, H. Huang, and C. H. Ding, "Low-rank matrix recovery via efficient Schatten p-norm minimization," in *Proceedings of the 26th AAAI Conference on Artificial Intelligence*, pp. 655–661, Toronto, Canada, July 2012.

[39] Z. Zhang, X. Liu, and L. Wang, "Spectral clustering algorithm based on improved Gaussian kernel function and beetle antennae search with damping factor," *Computational Intelligence and Neuroscience*, vol. 2020, Article ID 1648573, 9 pages, 2020.

[40] E. H. Lieb and W. E. Thirring, *Inequalities for the Moments of the Eigenvalues of the Schrödinger Hamiltonian and Their*

IEEE Journal of Biomedical and Health Informatics, 2021, In press.

Telation to Sobolev Inequalities, pp. 205–239, Springer, Berlin, Germany, 2005.

[41] H. Araki, "On an inequality of lieb and thirring," *Letters in Mathematical Physics*, vol. 19, no. 2, pp. 167–170, 1990.

[42] L. Yu, C. Ding, and S. Loscalzo, "Stable feature selection via dense feature groups," *Proceedings of ACM Sigkdd International Conference on Knowledge Discovery & Data Mining*, vol. 40, no. 1, pp. 803–811, 2008.

[43] A. Statnikov, I. Tsamardinos, Y. Dosbayev, and C. F. Aliferis, "GEMS: A system for automated cancer diagnosis and biomarker discovery from microarray gene expression data," *International Journal of Medical Informatics*, vol. 74, no. 7, pp. 491–503, 2005.

[44] G. Getz, H. Gal, I. Kela, D. A. Notterman, and E. Domany, "Coupled two-way clustering analysis of breast cancer and colon cancer gene expression data," *Bioinformatics*, vol. 19, no. 9, pp. 1079–1089, 2003.

[45] D. Cai, X. He, X. Wu, and J. Han, "Non-negative matrix factorization on manifold," in *Proceedings of the 8th IEEE International Conference on Data Mining*, pp. 63–72, Pisa, Italy, December 2008.

[46] X. Chen and C. Jian, "Gene expression data clustering based on graph regularized subspace segmentation," *Neurocomputing*, vol. 143, no. 16, pp. 44–50, 2014.

[47] X. Chen, M. Liao, and X. Ye, "Projection subspace clustering," *Journal of Algorithms & Computational Technology*, vol. 11, no. 3, pp. 224–233, 2017.

[48] P. A Estévez, M Tesmer, C. A Perez, and J. M Zurada, "Normalized mutual information feature selection," *IEEE Transactions on Neural Networks*, vol. 20, no. 2, pp. 189–201, 2009.

[49] G. Liu and S. Yan, "Latent low-rank representation for subspace segmentation and feature extraction," in *Proceedings of the International Conference on Computer Vision*, no. 4, pp. 1615–1622, Barcelona, Spain, November 2011.

[50] D. Kong, H. Huang, and H. Huang, "Robust nonnegative matrix factorization using L21-norm," in *Proceedings of the ACM International Conference on Information and Knowledge Management*, pp. 673–682, Scotland, UK, October 2011.

[51] J. A. Hartigan, "Direct clustering of a data matrix," *Journal of the American Statistical Association*, vol. 67, no. 337, pp. 123–129, 1972.