IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 27 November 2023, accepted 22 December 2023, date of publication 25 December 2023,
date of current version 12 January 2024.

Digital Object Identifier 10.1109/ACCESS.2023.3346946

RESEARCH ARTICLE

Joint Learning of Spectral Clustering and Low-Rank Representation Based on Precise Segmentation Cost for Cancer Subtype Identification

YANMING CAO[1], DAN LI[1], ZEYUAN WANG[1], AND JIA WANG[2]
[1]School of Control Science and Engineering, Dalian University of Technology, Dalian, Liaoning 116024, China
[2]Department of Breast Surgery, The Second Hospital of Dalian Medical University, Dalian, Liaoning 116014, China

Corresponding author: Jia Wang (wangjia77@hotmail.com)

This work was supported by the "1+X" Program Cross-Disciplinary Innovation Project under Grant 2022JCXKYB07, and in part by Wu Jieping Medical Foundation under Grant 320.6750.2022-19-81.

ABSTRACT Tumor samples clustering based on subspace segmentation is an effective method to discover cancer subtypes. Accurate and reliable identifications of cancer subtypes are crucial for understanding cancer pathogenesis as well as clinical diagnosis and treatment. Joint learning-based subspace clustering methods utilize the high correlation between data affinity and data segmentation for better clustering results. However, existing joint learning-based methods only provide an approximation of the segmentation cost in the joint optimization term, which could lead to sub-optimal results. To address this problem, we propose an algorithm named joint learning of spectral clustering and low-rank representation based on precise segmentation cost (JLSLPS) for cancer subtype identification. In our method, we impose non-negative and symmetric constraints on the low-rank representation matrix so that the representation coefficient can be equivalent to data affinity for the precise representation of segmentation cost. Therefore, the spectral clustering objective can be represented precisely to guide the learning of data affinity and segmentation more effectively. Finally, we solve the optimization problem of JLSLPS by using the linearized alternating direction method with adaptive penalty. We run JLSLPS on 8 cancer gene expression datasets and used 9 state-of-the-art clustering methods for comparison. Experimental results show that our method can increase ACC by 1.00-7.95% and NMI by 2.7-12.22% compared with the other methods, which proves the superiority of the proposed method.

INDEX TERMS Low-rank representation, spectral clustering, joint learning, gene expression data, cancer subtype identification.

I. INTRODUCTION

Cancer, as one of the most malignant diseases since the 21st century, has been a key concern and research challenge for the medical community [1]. Several studies have shown that tumor samples clustering can be used to discover the pathological characteristics of the samples and thus identify biologically significant cancer subtypes. Therefore, it is important for understanding the pathogenesis of cancer and promoting personalized cancer therapy [2], [3], [4], [5].

The associate editor coordinating the review of this manuscript and approving it for publication was Wentao Fan.

Cancer gene expression data are characterized by a small sample size and a high feature dimensionality [6], which makes it challenging to cluster efficiently and accurately. In recent years, several scholars have found that it is promising to explore the cluster structure of high-dimensional data in the sub-feature spaces [7], [8], [9], [10], [11]. Generally, there are two most representative methods among the popular subspace clustering methods, which are sparse subspace clustering (SSC) [12] and low-rank representation subspace clustering (LRR) [13], [14]. Both methods apply the self-representation property of the data, i.e., using the data itself as a dictionary. Specifically, SSC learns a sparse

self-representation matrix by imposing l_1 norm on the representation coefficients, while LRR adopts the nuclear norm as a convex approximation of the rank function to construct a low-rank representation matrix. SSC only considers the sparse representation of the data in each subspace, lacking the ability to capture the global structure of the data [15]. Besides, LRR is derived from a strict mathematical proof [16]. Therefore, LRR has become a research focus in subspace clustering and has been widely used in image processing and tumor samples clustering [8], [15], [16], [17], [18], [19].

One of the challenges of LRR is to construct a representation model that can reveal the true subspace structure of high-dimensional data [11]. A couple of methods focus on proposing various constraints. For example, in literatures [20], [21], and [22], the sparse, non-negative, symmetric constraints are used respectively to enhance the discriminability of the representation matrix. Besides, several methods consider introducing different regularizations into LRR, such as [8], [19], [23], [24], and [25]. In [23], the compound schatten p-norm is used to replace the nuclear norm to induce adaptive affinity matrix. The rest of the methods utilize graph Laplacian regularization based on manifold learning to capture local information of data. Noting that the above methods fail to utilize the high correlation between data affinity and data segmentation, the structured sparse subspace clustering (S^3C) [26] is presented to co-optimize the representation matrix and segmentation matrix of spectral clustering. Consequently, a series of subspace clustering methods based on the joint learning framework of S^3C have been proposed [27], [28], [29]. However, the subspace structured norm causes the approximate representation of the segmentation cost of data in the joint learning framework. Therefore, these methods fail to precisely measure the disagreement between the representation matrix and segmentation matrix, which directly affects the clustering performance.

In view of the mentioned discussions, we attempt to design an algorithm that could precisely represent the segmentation cost of data for joint learning. Inspired by the subspace structured norm in S^3C, we consider improving the joint optimization term so that it can be an equivalent replacement for the objective function for spectral clustering. To this end, we propose a method called joint learning of spectral clustering and low-rank representation based on precise segmentation cost (JLSLPS). Specifically, we impose non-negative and symmetric constraints on the representation matrix. This idea is mainly motivated by the fact that, when both constraints are imposed, the representation coefficients can be equivalently transformed into the data affinity. Thus, the spectral clustering objective can be represented precisely in the joint learning framework. It greatly differs from the subspace structured norm in S^3C. The improved joint optimization term contains precise rather than approximate segmentation cost of data, thereby enhancing the clustering performance. For simplicity, we give the abbreviations and their corresponding full names in Table 1. The specific contributions of our work are as follows:

TABLE 1. Abbreviations and their full names.

Abbreviation	Full Name
SSC	sparse subspace clustering
LRR	low-rank representation
LLRR	Laplacian regularized low-rank representation
sgLRR	graph regularized low-rank representation under symmetric and sparse constraints
NSLRG	non-negative symmetric low-rank representation with graph regularization
S^3C	structured sparse subspace clustering
LRS^3C	low-rank and structured sparse subspace clustering
SGRLRSC	structured graph regularized low-rank subspace clustering
S3C2	structured sparse subspace clustering and completion
JLSLPS	joint learning of spectral clustering and low-rank representation based on precise segmentation cost
LADMAP	linearized alternating direction method with adaptive penalty
ALM	augmented Lagrange multiplier
SVD	skinny singular value decomposition
SVT	singular value thresholding
AC	Adrenocortical Carcinoma
CHOL	Cholangiocarcinoma
COAD	Colon Adenocarcinoma
ESCA	Esophageal Carcinoma
HNSC	Head and Neck Squamous Cell Carcinoma
PAAD	Pancreatic Adenocarcinoma
READ	Rectum Adenocarcinoma
ACC	accuracy
NMI	Normalized mutual information

(1) The proposed algorithm makes full use of the high correlation between the representation matrix and segmentation matrix to learn them jointly under a unified framework. On this foundation, we utilize non-negative and symmetric constraints to improve the joint optimization term, which leads it to be equivalent to the spectral clustering objective for precisely describing the cost of segmentation.

(2) We design a loop iteration solution for the optimization of the proposed algorithm. We introduce manifold learning into LRR and solve for a better representation matrix by LADMAP. Then, we use spectral clustering to obtain a segmentation matrix and utilize it to update the representation matrix for the next iteration.

(3) We construct several multi-cancer integrated datasets using data from The Cancer Genome Atlas (TCGA) to perform comparative experiments. We also perform comparative experiments with clustering visualization at different numbers of clusters. The experimental results demonstrate the practical significance of our work in the field of cancer subtype identification.

II. RELATED WORK

A. LOW-RANK REPRESENTATION (LRR)

Low-rank representation (LRR) is a subspace clustering method widely used for data mining. Its main idea is to represent data samples as linear combinations of dictionary data and find the lowest-rank representation of the data [14], [23]. LRR assumes that the data $X = [x_1, x_2, \ldots, x_n] \in \mathbb{R}^{m\times n}$ are sampled from a union of d orthogonal subspaces $\cup_{j=1}^{d} S_j$ in the data space, where X contains m features and n samples as its rows and columns. It uses the rank of a matrix to measure its sparsity, and the model is represented as follows [14]:

$$\min_{Z,E} \|Z\|_* + \lambda\|E\|_{2,1} \quad s.t. \quad X = XZ + E \tag{1}$$

where $Z \in \mathbb{R}^{n\times n}$ is the low-rank representation matrix and z_i is the coefficient vector of datum x_i represented by other data. $\|Z\|_*$ is the nuclear norm of Z as a convex approximation of its rank, i.e., the sum of all singular values of the matrix [14]. $E \in \mathbb{R}^{m\times n}$ is the error matrix and $\|E\|_{2,1}$ is the $l_{2,1}$ norm of E to describe sample-specific corruptions. $\lambda > 0$ is the penalty parameter used to balance the role of each part in Eq. (1). The constraint is a self-representation of the data choosing data matrix X itself as the dictionary. Then, the affinity matrix can be constructed by the representation matrix as the input for spectral clustering.

Ideally, the representation matrix Z learned by LRR with a block diagonal structure can describe the global structure of the data well, which can lay a solid foundation for the subsequent clustering analysis [15], [16], [17]. However, LRR ignores the local geometric information implicit in the data and their neighbors [19], [24], [25]. Therefore, we attempt to solve this problem to achieve better clustering results.

B. SUBSPACE CLUSTERING BASED ON JOINT LEARNING

The subspace clustering methods based on joint learning utilize the intrinsic connection between the data affinity and data segmentation to learn the representation matrix and segmentation matrix in a joint optimization framework. The representative method of these methods [26], [27], [28], [29] is structured sparse subspace clustering (S^3C) [26].

The S^3C algorithm first defines a binary segmentation matrix $Q \in \mathbb{R}^{n\times d}$, in which the element q_{ij} denotes the membership of sample x_i to subspace S_j. If x_i belongs to S_j, then $q_{ij} = 1$; otherwise, $q_{ij} = 0$. This segmentation matrix should satisfy $Q \in \Omega$, where $\Omega = \{Q \in \{0,1\}^{n\times d}: Q\mathbf{1} = \mathbf{1}, \text{rank}(Q) = d\}$. After learning the representation matrix Z, the affinity matrix can be calculated based on Eq. (2):

$$A = \frac{1}{2}\left(|Z| + \left|Z^\top\right|\right) \tag{2}$$

where $|Z|$ indicates the absolute value matrix.

It can be seen that the affinity matrix A is symmetric with all elements positive. Besides, spectral clustering [30] relaxes the constraint on Q to $Q^\top Q = I$, thus, the optimization objective of spectral clustering can be expressed as:

$$\min_{Q} \frac{1}{2}\sum_{i,j} a_{ij}\left\|\boldsymbol{Q}_{i,:} - \boldsymbol{Q}_{j,:}\right\|_2^2 \quad s.t.\ Q^\top\boldsymbol{Q} = \boldsymbol{I_d} \tag{3}$$

where $Q_{i,:}$ is the i-th row of Q, I_d is the identity matrix with size $d \times d$.

To reflect the connection between data affinity and segmentation, S^3C defines the following subspace structured norm:

$$\|\boldsymbol{Z}\|_Q \doteq \sum_{i,j} |z_{ij}|\left(\frac{1}{2}\left\|\boldsymbol{Q}_{i,:} - \boldsymbol{Q}_{j,:}\right\|_2^2\right) = \|\Theta \odot \boldsymbol{Z}\|_1 \tag{4}$$

where $\Theta \in \mathbb{R}^{n\times n}$ is the subspace structure matrix with element $\theta_{i,j} = (\|Q_{i,:} - Q_{j,:}\|_2^2)/2$, $\odot$ denotes the Hadamard product. Then, Li et al. constructed the S^3C model based on SSC with the following objective function [26]:

$$\begin{aligned}&\min_{\boldsymbol{Z},E,Q} \|\boldsymbol{Z}\|_1 + \omega\|\boldsymbol{Z}\|_Q + \lambda\|\boldsymbol{E}\|_1\\ &s.t. \quad \boldsymbol{X} = \boldsymbol{XZ} + \boldsymbol{E}, diag(\boldsymbol{Z}) = 0, Q^\top\boldsymbol{Q} = \boldsymbol{I}\end{aligned} \tag{5}$$

where ω and λ are non-negative penalty parameters. S^3C is a generalization of the standard SSC by adding $\|Z\|_Q$ as the joint optimization term. The method integrates the optimization of the representation matrix and spectral clustering into a unified optimization framework, utilizing these two to guide their iterations of each other.

III. PROPOSED METHOD

Despite the success of the existing subspace clustering methods based on joint learning [26], [27], [28], [29], they have a common drawback, that is, they fail to precisely incorporate the segmentation cost into the joint learning framework. It can be seen that the subspace structured norm $\|Z\|_Q$ replaces a_{ij} in Eq. (3) with $|z_{ij}|$, so $\|Z\|_Q$ is only an approximation of the segmentation cost. This problem undoubtedly affects the performance of joint learning. As pointed out in the literature [12] and [22], when using $|Z|$ as the pairwise affinity relationships directly, some data in the same subspace may not necessarily be chosen to represent each other, this inevitably leads to loss of intrinsic correlation of data. Therefore, in this paper, a joint learning of spectral clustering and low-rank representation based on precise segmentation cost (JLSLPS) is proposed for the cluster analysis of cancer gene expression data. By imposing non-negative and symmetric constraints on the representation matrix, JLSLPS enables the representation coefficients to equivalently replace the data affinity in the joint optimization term. Therefore, the segmentation cost is precisely described in the joint learning framework, which can effectively overcome the above-mentioned limitation and further enhance the clustering performance.

A. PROPOSED OBJECTIVE FUNCTION

In this paper, from the perspective of improving the precision of the segmentation cost in joint learning, the proposed

JLSLPS algorithm imposes non-negative and symmetric constraints on the low-rank representation matrix Z. In this case, the representation coefficients in Z can be directly used as the affinity between data. Therefore, the spectral clustering objective Eq. (3) can be rewritten in the following form:

$$\min_{\boldsymbol{Q}} \frac{1}{2}\sum_{i,j} z_{ij}\left\|\boldsymbol{Q}_{i,:}-\boldsymbol{Q}_{j,:}\right\|_2^2$$
$$s.t.\ \boldsymbol{Z}=Z^\top, \boldsymbol{Z}\geq 0, Q^\top \boldsymbol{Q}=\boldsymbol{I} \tag{6}$$

Eq. (6) is the spectral clustering objective, aiming to find the segmentation matrix Q that minimizes the cost of data partition. Moreover, it precisely quantifies the interdependence between matrices Z and Q. It can be seen that if Q is correct and Z reflects the true subspace structure of data, then Eq. (6) is equal to zero; otherwise, Eq. (6) is positive, in which case the data partition with the minimum cost can be found. The proposed JLSLPS algorithm imposes the non-negative and symmetric constraints on Z for the following two reasons:

1) The non-negative and symmetric constraints guarantee the elements in Z strictly satisfy $|z_{ij}| = z_{ij}$ and $z_{ij} = z_{ji}$. Therefore, according to Eq. (2), the data affinity a_{ij} in Eq. (3) can be replaced equivalently by the representation coefficient z_{ij} to describe the segmentation cost. Compared with the subspace structure norm defined in Eq. (4), Eq. (6) is fully equivalent to the spectral clustering objective, which can precisely reflect the interdependence between representation coefficients and segmentation.

2) Applying non-negative constraint on Z is beneficial to obtain a more interpretative representation matrix and prevent negative representation coefficients from affecting data partition [21]. Besides, the symmetric low-rank representation can guarantee the consistency of the affinity of data pairs. Thus, the symmetric constraint imposed on Z is favorable for preserving the subspace structure of high-dimensional data and learning a more discriminative representation matrix [22].

In this paper, we introduce Eq. (6) into the low-rank representation subspace clustering as a joint optimization term, which is used to measure the correlation of representation coefficients and segmentation. The objective function of the proposed JLSLPS algorithm is as follows:

$$\min_{\boldsymbol{Z},\boldsymbol{E},\boldsymbol{Q}} \|\boldsymbol{Z}\|_* + \lambda_1\|\boldsymbol{Z}\|_1 + \frac{\alpha}{2}\sum_{i,j} z_{ij}\left\|\boldsymbol{Q}_{i,:}-\boldsymbol{Q}_{j,:}\right\|_2^2 + \lambda_2 tr\left(\boldsymbol{Z}\boldsymbol{L}\boldsymbol{Z}^\top\right) + \lambda_3\|\boldsymbol{E}\|_1$$
$$s.t.\ \boldsymbol{X}=\boldsymbol{X}\boldsymbol{Z}+\boldsymbol{E}, \boldsymbol{Z}=Z^\top, \boldsymbol{Z}\geq 0, Q^\top\boldsymbol{Q}=\boldsymbol{I} \tag{7}$$

where λ_1, λ_2, λ_3, and α are positive penalty parameters. The first term of Eq. (7) is the same as LRR, which plays a key role in capturing the global low-rank structure of the high-dimensional data. The second term is a sparse constraint based on l_1 norm, whose aim is to weaken the affinity of inter-class data and partition similar data into the same subspace. The third term is the joint optimization term based on Eq. (6), which exploits the row information of Q to guide the learning of Z and achieve the co-optimization of Z and Q. The fourth term is the graph regularization based on manifold learning to capture the local manifold structure of data [8], [19], [24], [25]. Finally, the error term is adjusted from $\|E\|_{2,1}$ to $\|E\|_1$ as suggested by literatures [8] and [25], which is used to portray the sparse random noise in cancer gene expression data.

Based on the objective function, the proposed JLSLPS algorithm establishes a joint learning framework for representation matrix Z and segmentation matrix Q. By imposing non-negative and symmetric constraints on Z, the proposed joint learning framework enables Z to be more diagonal, which helps to learn Q closer to the real partition. Additionally, spectral clustering can be regarded as a process of obtaining the partition by denoising [7], [26], so feeding back Q to Z can correct some of the errors in Z. Therefore, the partition information from the previous step can be used as a better initialization to enhance the discriminability of Z. When the termination criterion is met, we impose K-means clustering on Q to obtain the final clustering results.

B. OPTIMIZATION OF OBJECTIVE FUNCTION

For the JLSLPS model (7) proposed in this paper, the following two sub-problems are solved by using an alternating optimization strategy to achieve joint learning:

1) Fix Q and solve the model (7) to learn Z and E.

2) Fix Z and E and solve the spectral clustering problem to learn Q.

For sub-problem 1), the linearized alternating direction method with adaptive penalty (LADMAP) [31] is used for solving the problem. First, the second and third terms of the model (7) are combined to simplify the optimization process:

$$\lambda_1\|\boldsymbol{Z}\|_1 + \frac{\alpha}{2}\sum_{i,j} z_{ij}\left\|\boldsymbol{Q}_{i,:}-\boldsymbol{Q}_{j,:}\right\|_2^2 = \|(\lambda_1\boldsymbol{I}+\alpha\Theta)\odot\boldsymbol{Z}\|_1 \tag{8}$$

Second, we introduce the auxiliary variable J to make the objective function Eq. (7) separable, and consider the following equivalent optimization problem:

$$\min_{\boldsymbol{Z},\boldsymbol{E},\boldsymbol{J}} \|\boldsymbol{Z}\|_* + \|(\lambda_1\boldsymbol{I}+\alpha\Theta)\odot\boldsymbol{J}\|_1 + \lambda_2 tr\left(\boldsymbol{Z}\boldsymbol{L}\boldsymbol{Z}^\top\right) + \lambda_3\|\boldsymbol{E}\|_1$$
$$s.t.\ \ \boldsymbol{X}=\boldsymbol{X}\boldsymbol{Z}+\boldsymbol{E}, \boldsymbol{Z}=Z^\top, \boldsymbol{Z}\geq 0, \boldsymbol{Z}=\boldsymbol{J} \tag{9}$$

Then, Eq. (9) can be converted into an unconstrained optimization problem by using the augmented Lagrange multiplier (ALM) [32]. The augmented Lagrangian function is:

$$\begin{aligned}\mathcal{L}(\boldsymbol{Z},\boldsymbol{E},\boldsymbol{J},\boldsymbol{Y}_1,\boldsymbol{Y}_2) = {} & \|\boldsymbol{Z}\|_* + \|(\lambda_1\boldsymbol{I}+\alpha\Theta)\odot\boldsymbol{J}\|_1 \\ & + \lambda_2 tr\left(\boldsymbol{Z}\boldsymbol{L}\boldsymbol{Z}^\top\right) + \lambda_3\|\boldsymbol{E}\|_1 + <\boldsymbol{Y}_1, \boldsymbol{X} \\ & - \boldsymbol{X}\boldsymbol{Z}-\boldsymbol{E}> + <\boldsymbol{Y}_2, \boldsymbol{Z}-\boldsymbol{J}> \\ & + \frac{\mu}{2}\|\boldsymbol{X}-\boldsymbol{X}\boldsymbol{Z}-\boldsymbol{E}\|_F^2 + \frac{\mu}{2}\|\boldsymbol{Z}-\boldsymbol{J}\|_F^2\end{aligned} \tag{10}$$

where $<A, B> = tr(A^\top B)$ is the inner product of the matrices, Y_1 and Y_2 are Lagrange multipliers, $\mu > 0$ is the penalty parameter. Further, Eq. (10) is divided into three optimization problems to be solved separately.

Update Z: Fix the other variables in Eq. (10), and update Z by the following problem:

$$\begin{aligned}\min_{\boldsymbol{Z}=Z^\top} \|\boldsymbol{Z}\|_* + \lambda_2 tr\left(\boldsymbol{Z}\boldsymbol{L}\boldsymbol{Z}^\top\right) + <\boldsymbol{Y}_1, \boldsymbol{X}-\boldsymbol{X}\boldsymbol{Z}-\boldsymbol{E}> \\ + <\boldsymbol{Y}_2, \boldsymbol{Z}-\boldsymbol{J}> + \frac{\mu}{2}\|\boldsymbol{X}-\boldsymbol{X}\boldsymbol{Z}-\boldsymbol{E}\|_F^2 \\ + \frac{\mu}{2}\|\boldsymbol{Z}-\boldsymbol{J}\|_F^2 \end{aligned} \tag{11}$$

First, define the following formula:

$$\begin{aligned} f(\boldsymbol{Z}) = \lambda_2 tr\left(\boldsymbol{Z}\boldsymbol{L}\boldsymbol{Z}^\top\right) + \frac{\mu}{2}\left\|\boldsymbol{X}-\boldsymbol{X}\boldsymbol{Z}-\boldsymbol{E}+\frac{\boldsymbol{Y}_1}{\mu}\right\|_F^2 \\ + \frac{\mu}{2}\left\|\boldsymbol{Z}-\boldsymbol{J}+\frac{\boldsymbol{Y}_2}{\mu}\right\|_F^2 \end{aligned} \tag{12}$$

Let $Z^{(k)}$ be the representation matrix obtained at the k-th iteration, we derive the first derivative of $f(Z)$ with respect to $Z^{(k)}$ as:

$$\begin{aligned}\frac{\partial f(\boldsymbol{Z})^{(k)}}{\partial \boldsymbol{Z}^{(k)}} = \lambda_2\left(\boldsymbol{Z}^{(k)}\boldsymbol{L}+\boldsymbol{Z}^{(k)}L^\top\right) \\ + \mu^{(k)}X^\top\left(\boldsymbol{X}\boldsymbol{Z}^{(k)}-\boldsymbol{X}+\boldsymbol{E}^{(k)}-\frac{\boldsymbol{Y}_1^{(k)}}{\mu^{(k)}}\right) \\ + \mu^{(k)}\left(\boldsymbol{Z}^{(k)}-\boldsymbol{J}^{(k)}+\frac{\boldsymbol{Y}_2^{(k)}}{\mu^{(k)}}\right)\end{aligned} \tag{13}$$

According to LADMAP, Eq. (11) can be replaced by solving the following problem:

$$\min_{\boldsymbol{Z}=Z^\top}\|\boldsymbol{Z}\|_* + <\frac{\partial f(\boldsymbol{Z})^{(k)}}{\partial \boldsymbol{Z}^{(k)}}, \boldsymbol{Z}-\boldsymbol{Z}^{(k)}> + \frac{\eta}{2}\left\|\boldsymbol{Z}-\boldsymbol{Z}^{(k)}\right\|_F^2 \tag{14}$$

where $\eta = 2\lambda_2\|L\|_2 + \mu^{(k)}(1 + \|X\|_2^2)$. Then, Eq. (14) is transformed as follows:

$$\min_{\boldsymbol{Z}=Z^\top}\frac{1}{\eta}\|\boldsymbol{Z}\|_* + \frac{1}{2}\left\|\boldsymbol{Z}-\left(\boldsymbol{Z}^{(k)}-\frac{\partial f(\boldsymbol{Z})^{(k)}}{\partial \boldsymbol{Z}^{(k)}}/\eta\right)\right\|_F^2 \tag{15}$$

Equation (15) is solved by the following lemma:

Lemma [22]: Given a square matrix $H \in \mathbb{R}^{n\times n}$, the unique closed-form solution to the optimization problem:

$$\arg\min_{\boldsymbol{G}}\frac{1}{\varphi}\|\boldsymbol{G}\|_* + \frac{1}{2}\|\boldsymbol{G}-\boldsymbol{H}\|_F^2 \quad s.t.\ \boldsymbol{G}=G^\top \tag{16}$$

takes this form:

$$\boldsymbol{G}^* = \boldsymbol{U}_r\left(\Sigma_r - \frac{1}{\varphi}\boldsymbol{I}_r\right)V_r^\top, \tag{17}$$

We set $\tilde{H} = (H + H^\top)/2$, $\tilde{H} = U_r\Sigma_r V_r^\top$ is the skinny singular value decomposition (SVD) of the matrix $\tilde{H}$. $\Sigma_r = diag(\delta_1, \delta_2, \ldots, \delta_r)$, δ_r is a singular value larger than $1/\varphi$, U_r and V_r are the corresponding singular vectors of $\tilde{H}$.

According to the above lemma, we set

$$\widetilde{\boldsymbol{Z}}^{(k+1)} = \frac{\left(\left(Z^{(k)}-\frac{\partial f(Z)^{(k)}}{\partial Z^{(k)}}/\eta\right)+\left(Z^{(k)}-\frac{\partial f(Z)^{(k)}}{\partial Z^{(k)}}/\eta\right)^\top\right)}{2} \tag{18}$$

and perform the SVD on $\tilde{Z}^{(k+1)}$ as $\tilde{Z}^{(k+1)} = U\Sigma V$. The diagonal elements of Σ are singular values larger than $1/\eta$, then the solution of $Z^{(k+1)}$ is:

$$\boldsymbol{Z}^{(k+1)} = \boldsymbol{U}\left(\Sigma - \frac{1}{\eta}\boldsymbol{I}\right)V^\top \tag{19}$$

We then impose a non-negative constraint on $Z^{(k+1)}$ as follows:

$$Z_{i,j}^{(k+1)} = \begin{cases}\boldsymbol{Z}_{i,j}^{(k+1)}, & Z_{i,j}^{(k+1)} \geq 0 \\ 0, & \text{otherwise}\end{cases} \tag{20}$$

Update J: Fix the other variables in Eq. (10), and update J using the following problem:

$$\min_{\boldsymbol{J}}\|(\lambda_1\boldsymbol{I}+\alpha\Theta)\odot\boldsymbol{J}\|_1 + <\boldsymbol{Y}_2, \boldsymbol{Z}-\boldsymbol{J}> + \frac{\mu}{2}\|\boldsymbol{Z}-\boldsymbol{J}\|_F^2 \tag{21}$$

Then, we transform Eq. (21) into the following formula:

$$\begin{aligned}\boldsymbol{J}^{(k+1)} = \arg\min_{\boldsymbol{J}}\frac{1}{\mu^{(k)}}\left\|(\lambda_1\boldsymbol{I}+\alpha\Theta)\odot\boldsymbol{J}^{(k)}\right\|_1 \\ + \frac{1}{2}\left\|\boldsymbol{J}^{(k)}-\left(\boldsymbol{Z}^{(k+1)}+\frac{\boldsymbol{Y}_2^{(k)}}{\mu^{(k)}}\right)\right\|_F^2\end{aligned} \tag{22}$$

We use the singular value thresholding (SVT) [33] to solve Eq. (22). Let $P^{(k+1)} = Z^{(k+1)} + Y_2{}^{(k)}/\mu^{(k)}$, the closed-form solution of $J^{(k+1)}$ can be found as:

$$\boldsymbol{J}_{i,j}^{(k+1)} = \mathcal{S}_{\frac{\lambda_1+\alpha\theta_{ij}}{\mu^{(k)}}}P_{i,j}^{(k+1)} \tag{23}$$

where $\mathcal{S}_\tau(u)$ is the shrinkage thresholding operator defined as $\mathcal{S}_\tau(u) = \text{sgn}(u)\cdot\max(|u| - \tau, 0)$.

Update E: Fix the other variables in Eq. (10), and update E using the following problem:

$$\min_{\boldsymbol{E}}\lambda_3\|\boldsymbol{E}\|_1 + <\boldsymbol{Y}_1, \boldsymbol{X}-\boldsymbol{X}\boldsymbol{Z}-\boldsymbol{E}> + \frac{\mu}{2}\|\boldsymbol{X}-\boldsymbol{X}\boldsymbol{Z}-\boldsymbol{E}\|_F^2 \tag{24}$$

Similar to the optimization of J, Eq. (24) is firstly transformed as:

$$\begin{aligned}E^{(k+1)} = \arg\min_{\boldsymbol{E}}\frac{1}{\mu^{(k)}}\left\|\boldsymbol{E}^{(k)}\right\|_1 \\ + \frac{1}{2}\left\|\boldsymbol{E}^{(k)}-\left(\boldsymbol{X}-\boldsymbol{X}\boldsymbol{Z}^{(k+1)}+\frac{\boldsymbol{Y}_1^{(k)}}{\mu^{(k)}}\right)\right\|_F^2\end{aligned} \tag{25}$$

The closed-form solution of $E^{(k+1)}$ can be found as:

$$\boldsymbol{E}^{(k+1)} = \mathcal{S}_{\frac{\lambda_3}{\mu^{(k)}}}\left(\boldsymbol{X}-\boldsymbol{X}\boldsymbol{Z}^{(k+1)}+\frac{\boldsymbol{Y}_1^{(k)}}{\mu^{(k)}}\right) \tag{26}$$

Algorithm 1 Solving Problem (9) by LADMAP

Input:
Data matrix X, parameters λ_1, λ_2, λ_3, and α
Initialize:
$Z^{(0)} = J^{(0)} = E^{(0)} = \Theta^{(0)} = Y_1{}^{(0)} = Y_2{}^{(0)} = 0$, $\mu^{(0)} = 10^{-2}$, $\mu_{max} = 10^6$, $\rho_0 = 2$, $\varepsilon_1 = 10^{-6}$, $\varepsilon_2 = 10^{-2}$, the number of iteration $k = 0$
while not converged **do**
(1) Update $Z^{(k+1)}$, $J^{(k+1)}$, $E^{(k+1)}$ by (19), (23) and (26);
(2) Update $Y_1{}^{(k+1)}$ and $Y_2{}^{(k+1)}$:

$$\boldsymbol{Y}_1^{(k+1)} = \boldsymbol{Y}_1^{(k)} + \mu^{(k)}\left(\boldsymbol{X} - \boldsymbol{X}\boldsymbol{Z}^{(k+1)} - \boldsymbol{E}^{(k+1)}\right)$$
$$\boldsymbol{Y}_2^{(k+1)} = \boldsymbol{Y}_2^{(k)} + \mu^{(k)}\left(\boldsymbol{Z}^{(k+1)} - \boldsymbol{J}^{(k+1)}\right)$$

(3) Update

$$\mu^{(k+1)} \leftarrow \min\left(\mu_{\max}, \quad \rho^{(k)}\mu^{(k)}\right)$$

where

$$\rho^{(k)} = \begin{cases} \rho_0, & \text{if } b \le \varepsilon_2 \\ 1, & \text{otherwise} \end{cases}$$
$$b = \max\left\{\eta l_Z, \mu^{(k)} l_J, \mu^{(k)} l_E\right\}$$
$$[3pt] l_Z = \left\|Z^{(k+1)} - Z^{(k)}\right\|_1$$
$$l_J = \left\|J^{(k+1)} - J^{(k)}\right\|_1$$
$$l_E = \left\|E^{(k+1)} - E^{(k)}\right\|_1$$

(4) Check the convergence condition

$$\frac{\left\|X - XZ^{(k+1)} - E^{(k+1)}\right\|_1}{\|X\|_1} < \varepsilon_1$$

or $b < \varepsilon_2$, if not converged,then set $k \leftarrow k + 1$
end while
Output: $Z^{(k+1)}$ and $E^{(k+1)}$

The procedure for solving the matrices Z and E is outlined in Algorithm 1.

For sub-problem 2), after obtaining the matrices Z and E, the JLSLPS model (7) can be transformed into:

$$\min_{\boldsymbol{Q}} \frac{1}{2}\sum_{i,j} z_{ij}\left\|\boldsymbol{Q}_{i,:} - \boldsymbol{Q}_{j,:}\right\|_2^2 \qquad s.t.\ Q^\top \boldsymbol{Q} = \boldsymbol{I} \tag{27}$$

Since the non-negative and symmetric constraints make $z_{ij} = a_{ij}$, the problem is transformed into a spectral clustering problem:

$$\min_{\boldsymbol{Q}} tr\left(Q^\top \boldsymbol{L}\boldsymbol{Q}\right) \qquad s.t.\ Q^\top \boldsymbol{Q} = \boldsymbol{I} \tag{28}$$

Thus, the segmentation matrix Q can be learned by computing the d eigenvectors corresponding to the d smallest eigenvalues of L.

Algorithm 2 JLSLPS

Input:
Data matrix X, parameters λ_1, λ_2, λ_3, and α, number of subspaces d
Initialize:
$Q^{(0)} = 0$
while not converged **do**
(1) Given Q, update Z and E by **Algorithm 1**;
(2) Given Z and E, update Q by solving the spectral clustering problem (33);
end while
Perform K-means on Q to segment X into d subspaces
Output: The clustering results

C. CONVERGENCE CONDITION AND COMPUTATIONAL COMPLEXITY

The problem solved by Algorithm 1 is a convex optimization problem, and the convergence analysis of Algorithm 1 is concisely verified in the Appendix. However, it cannot be guaranteed that Algorithm 2 can converge to a global or local optimum due to the relaxed constraint on Q. In spite of this, the experiment results show that Algorithm 2 can converge if the parameters are set properly. According to the suggestion of literature [26], we terminate Algorithm 2 if a predetermined number of iterations is reached, or if the relative change of the representation matrix Z and the subspace structure matrix Θ in two consecutive iterations is less than a certain threshold value, i.e.,

$$\frac{\left\|Z^{(k)} - Z^{(k+1)}\right\|_1}{\left\|Z^{(k)}\right\|_1} < \varepsilon_3 \tag{29}$$
$$\frac{\left\|\Theta^{(k)} - \Theta^{(k+1)}\right\|_1}{\left\|\Theta^{(k)}\right\|_1} < \varepsilon_3 \tag{30}$$

where $Z^{(k)}$ and $\Theta^{(k)}$ are the matrices Z and Θ obtained at the k-th iteration, respectively. ε_3 is a small positive constant (empirically set as $\varepsilon_3 = 10^{-2}$).

Then, we further measure the computational cost of JLSLPS. According to literature [32], the LADMAP method can reduce the time complexity of LRR to $O(rn^2)$, where r is the rank of matrix Z. Besides, it takes $O(mn^2)$ matrix multiplications to compute XZ. Hence, the computational cost is $O(T_1T_2(rn^2+mn^2))$, where T_1 is the number of iterations in solving Eq. (7) and T_2 is the number of outer iterations in Algorithm 2.

IV. EXPERIMENTS AND ANALYSIS

A. EXPERIMENT SETUP

In this paper, comparative experiments are conducted in the publicly available cancer gene expression datasets from TCGA [34]. For better evaluation, we test our JLSLPS algorithm using integrated datasets and available benchmark datasets. The details of the datasets are listed in Table 2.

(1) Integrated Datasets: We first construct five multi-cancer integrated datasets as suggested by literature [24]. Each of

TABLE 2. Detailed information of the eight cancer gene expression datasets.

Dataset	Number of			
	Classes (cancers or subtypes)	Samples per class	Total samples	Genes
CH-ES-PA	3	15-30-30	75	1800
AC-CH-ES	3	50-20-80	150	10000
ES-HN-CH	3	70-160-36	266	10000
CH-PA-RE-HN	4	36-185-98-200	519	20000
AC-ES-CO-CH	4	80-100-297-36	513	20000
Leukemia	6	15-27-64-20-43-79	248	985
Lung Cancer	3	139-17-41	197	1000
Novartis BPLC	4	26-26-28-23	103	1000

TABLE 3. Parameter selection of JLSLPS.

Dataset	α	λ_1	λ_2	λ_3
CH-ES-PA	2	10^{-3}	10^{-2}	1
AC-CH-ES	6	10^{-2}	10^{-1}	10^{-1}
ES-HN-CH	5	10^{-2}	10^{-1}	10^{-1}
CH-PA-RE-HN	5.5	10^{-1}	10^{-2}	10^{-2}
AC-ES-CO-CH	5	10^{-1}	10^{-1}	10^{-2}
Leukemia	5	10^{-2}	1	1
Lung Cancer	5	10^{-2}	10^{-2}	1
Novartis BPLC	6	10^{-3}	10^{-1}	2

them is composed of the tumor samples of multiple original datasets. We select seven cancer datasets, including Adrenocortical Carcinoma (AC), Cholangiocarcinoma (CHOL), Colon Adenocarcinoma (COAD), Esophageal Carcinoma (ESCA), Head and Neck Squamous Cell Carcinoma (HNSC), Pancreatic Adenocarcinoma (PAAD), and Rectum Adenocarcinoma (READ). The five integrated datasets are: CH-ES-PA, which contains 75 tumor samples (15 from CHOL, 30 from ESCA, 30 from PAAD), AC-CH-ES, which contains 150 tumor samples (50 from AC, 20 from CHOL, 80 from ESCA), ES-HN-CH, which contains 266 tumor samples (70 from ESCA, 160 from HNSC, 36 from CHOL), CH-PA-RE-HN, which contains 519 tumor samples (36 from CHOL, 185 from PAAD, 98 from READ, 200 from HNSC), and AC-ES-CO-CH, which contains 513 tumor samples (80 from AC, 100 from ESCA, 297 from COAD, 36 from CHOL).

(2) Available benchmark datasets: Three single cancer datasets with multiple subtypes are also considered: St. Jude leukemia dataset [35] contains 248 tumor samples, Lung Cancer dataset [36] contains 197 tumor samples, and Novartis BPLC dataset [37] contains 103 tumor samples.

To verify the effectiveness of the JLSLPS algorithm proposed in this paper, JLSLPS is compared with nine state-of-the-art subspace clustering algorithms, including SSC [12], LRR [14], Laplacian regularized low-rank representation (LLRR, [8]), graph regularized low-rank representation under symmetric and sparse constraints (sgLRR, [24]), non-negative symmetric low-rank representation with graph regularization (NSLRG, [25]), S^3C [26], low-rank and structured sparse subspace clustering (LRS^3C, [27]), structured graph regularized low-rank subspace clustering (SGRLRSC, [28]), and structured sparse subspace clustering and completion (S3C2, [29]). Among the comparison methods, SSC, LRR, LLRR, sgLRR, and NSLRG are two-stage methods, in which the affinity matrix is first learned and then spectral clustering is performed on the affinity graph. S^3C, LRS^3C, SGRLRSC, and S3C2 are joint learning-based algorithms that synergistically optimize the affinity and the segmentation.

In terms of parameter settings, for JLSLPS, the four parameters α, λ_1, λ_2, and λ_3 balance the importance of joint optimization term, sparse constraint term, manifold learning-based graph regularization term, and error term, respectively. In our empirical studies, the parameter values corresponding to the optimal result of different datasets are determined by the conventional grid search method based on the suggestion of literature [25] and [26]. Specifically, the parameter settings for the eight cancer data sets are listed in Table 3. The parameter analysis of the proposed method will be further discussed in the following section. For the comparison methods, as suggested by literature [12], [14], [25], and [26], we tune the penalty parameter λ of SSC and LRR within $[2^{-3}, 2^3]$, the parameter ω of joint optimization term of S^3C, LRS^3C, SGRLRSC, and S3C2 within $[1^{-3}, 20]$. Then, we set the graph regularization term parameter $\beta = 0.1$ for LLRR, sgLRR, NSLRG, and SGRLRSC, and the other parameters in comparison methods are set as suggested in the original studies. All algorithms are run 20 times, and the algorithm performance is evaluated using the mean value of the results.

We implemented JLSLPS in Matlab. All experiments were conducted on a Windows machine with an Intel Core i7 CPU at 3.6GHz and 32GB memory. Our method can be employed in real-time. The source code can be downloaded at: https://github.com/gglsx/ptk.git.

B. EVALUATION METRICS

In this paper, two evaluation metrics, accuracy (ACC) [7] and Normalized mutual information (NMI) [38], are used to assess the partition performance. These two metrics are widely used to evaluate the performance of clustering algorithms in data mining and machine learning. ACC is calculated as follows:

$$ACC = \frac{\sum_{i=1}^{n} \delta\left(c_i, map\left(\hat{c}_i\right)\right)}{n} \tag{31}$$

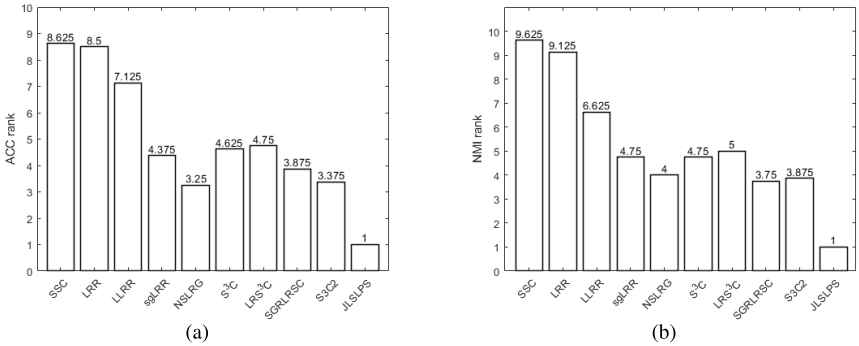


FIGURE 1. Average sort indexes of the nine methods. (a) ACC, and (b) NMI. The x-axis and y-axis represent different methods and their ranking of ACC or NMI, respectively.

TABLE 4. Clustering results (ACC) of each algorithm on eight cancer gene expression datasets. The optimal and suboptimal results are marked with bold and italics, respectively.

Datasets	SSC	LRR	LLRR	sgLRR	NSLRG	S^3C	LRS^3C	SGRLRSC	S3C2	JLSLPS
CH-ES-PA	0.9733	0.9867	0.9867	0.9867	0.9867	0.9867	0.9867	0.9867	0.9733	**1.0000**
AC-CH-ES	0.9333	0.9400	0.9733	0.9800	0.9867	0.9600	0.9867	0.9867	0.9867	**0.9933**
ES-HN-CH	0.8759	0.8872	0.9774	0.9849	0.9849	0.9398	0.9812	0.9812	0.9849	**0.9887**
CH-PA-RE-HN	0.9210	0.9345	0.9711	0.9769	0.9788	0.9653	0.9730	0.9750	0.9769	**0.9827**
AC-ES-CO-CH	0.8655	0.9025	0.9454	0.9727	0.9766	0.9610	0.9610	0.9727	0.9747	**0.9883**
Leukemia	0.9516	0.8185	0.9395	0.9435	0.9476	0.9758	0.9597	0.9677	0.9718	**0.9798**
Lung Cancer	0.9492	0.9137	0.9239	0.9289	0.9340	**0.9594**	0.9492	0.9543	0.9543	**0.9594**
Novartis BPLC	0.9320	0.8540	0.9417	0.9709	0.9709	0.9709	0.9515	0.9612	0.9709	**0.9806**

TABLE 5. Clustering results (NMI) of each algorithm on eight cancer gene expression datasets. The optimal and suboptimal results are marked with bold and italics, respectively.

Datasets	SSC	LRR	LLRR	sgLRR	NSLRG	S^3C	LRS^3C	SGRLRSC	S3C2	JLSLPS
CH-ES-PA	0.9153	0.9444	0.9489	0.9489	0.9489	0.9489	0.9489	0.9489	0.9026	**1.0000**
AC-CH-ES	0.8170	0.8246	0.9159	0.9164	0.9454	0.8576	0.9390	0.9443	0.9484	**0.9694**
ES-HN-CH	0.7678	0.8029	0.9092	0.9427	0.9427	0.8082	0.9231	0.9253	0.9429	**0.9443**
CH-PA-RE-HN	0.8545	0.8866	0.9168	0.9223	0.9278	0.9053	0.9278	0.9310	0.9330	**0.9380**
AC-ES-CO-CH	0.7983	0.8164	0.8524	0.9187	0.9268	0.8881	0.9043	0.9236	0.9133	**0.9606**
Leukemia	0.8790	0.7282	0.8895	0.8911	0.8872	0.9354	0.8958	0.9235	0.9330	**0.9472**
Lung Cancer	0.8196	0.7486	0.7672	0.7672	0.8037	**0.8451**	0.8196	0.8215	0.8201	**0.8451**
Novartis BPLC	0.8046	0.8249	0.8623	0.9199	0.9199	0.9327	0.8712	0.8909	0.9200	**0.9495**

where c_i and $\hat{c}_i$ denote the true label and predicted label of x_i, respectively; map $(\hat{c}_i)$ denotes the mapping match between the true and predicted labels, and $\delta(c_i, \text{map}\ (\hat{c}_i)) = 1$ when $c_i = \text{map}\ (\hat{c}_i)$, otherwise $\delta(c_i, \text{map}\ (\hat{c}_i)) = 0$.

NMI is calculated by:

$$NMI = \frac{2I(M,N)}{H(M)+H(N)} \tag{32}$$

where M and N denote the vectors consisting of the true and predicted labels, respectively; $I(M,N)$ denotes the mutual information measure, $H(M)$ and $H(N)$ denote the entropy of M and N, respectively. Both the ACC and NMI values range from 0 to 1, and for the two metrics, larger values indicate a better clustering performance.

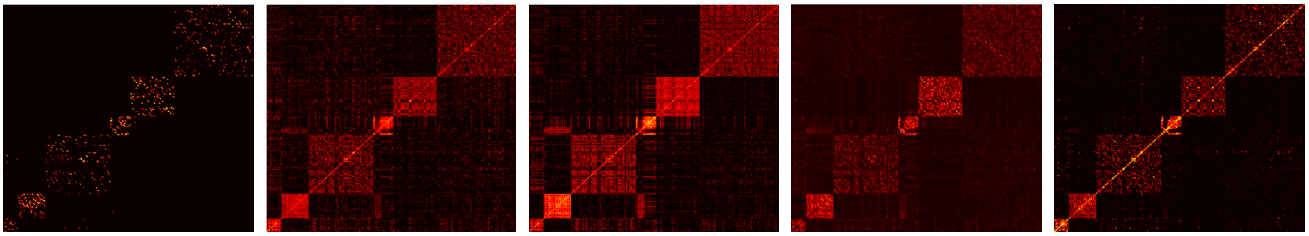
C. EXPERIMENTAL RESULTS AND ANALYSIS

Tables 4 and 5 show the ACC and NMI results of ten algorithms on eight cancer gene expression datasets. To give an overall comparison, we sort the average ACC and NMI values in descending order and obtain the sort index. Therefore, the method with sort index 1 has the highest ACC or NMI value. We gather statistics on the sort indexes of the methods. Fig. 1 gives the average sort indexes of the ten methods, and a smaller average sort index indicates a better clustering result.From the results in Table 4, Table 5, and Fig. 1, it can be seen that:

(1) For all datasets, the ACC and NMI results of the proposed JLSLPS algorithm are significantly superior to the comparison algorithms. For the two metrics, JLSLPS ranks first in all cases, which validates the solid performance of the proposed JLSLPS algorithm in sample clustering. Additionally, to test the statistical difference among JLSLPS and other methods, a one-way analysis of variance (ANOVA) test is applied in this experiment. The results for the one-way ANOVA test of evaluation metrics have been presented in Table 6, where DF, SS, MS, F, and P denote degrees of freedom, adjusted sum of squares, adjusted mean square, F-value, and probability value, respectively. According to

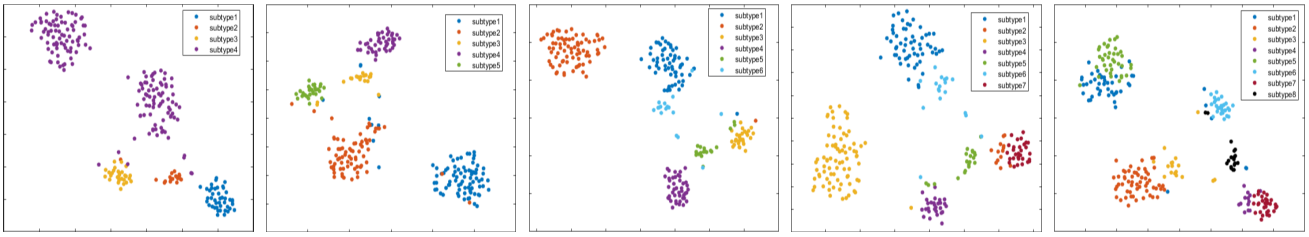
TABLE 6. ANOVA test results of the proposed method.

ACC Values						NMI Values					
Source	SS	DF	MS	F-Value	P-Value	Source	SS	DF	MS	F-Value	P-Value
Model	0.04407	9	0.0049	7.82	7.11e-08	Model	0.10538	9	0.01171	4.49	0.0001
Error	0.04385	70	0.00063			Error	0.18248	70	0.00261		
Total	0.08792	79				Total	0.28787	79			



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

FIGURE 2. Visualizations of the representation matrices on Leukemia. (a) S^3C, (b) LRS^3C, (c) SGRLRSC, (d) S3C2, and (e) JLSLPS. The x-axis and y-axis represent the corresponding representation coefficients of matrix Z.



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

FIGURE 3. Visualizations of different numbers of clusters d on Leukemia. (a) d=4, (b) d=5, (c) d=6, (d) d=7, and (e) d=8.

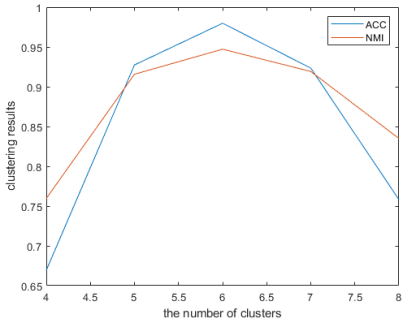


FIGURE 4. Clustering results of different numbers of d. The x-axis and y-axis represent the different values of d and the values of ACC and NMI, respectively.

the ANOVA test results, there are statistically meaningful differences between the proposed JLSLPS algorithm and other methods ($p \leq 0.0001$).

(2) In most cases, the LRR-based algorithms show better performance than the SSC-based algorithms. Specifically, in two-stage methods, the ACC and NMI results of LRR, LLRR, sgLRR, and NSLRG are higher than those of SSC. In the joint optimization algorithms, the metric values of LRS^3C, SGRLRSC, S3C2, and the proposed JLSLPS are higher than those of S^3C. This phenomenon is particularly evident in high-dimensional datasets AC-CH-ES, ES-HN-CH, CH-PA-RE-HN and AC-ES-CO-CH. As pointed out in the literature [8], [18], [19], [24], and [25], when dealing with cancer gene expression data with a large amount of noise, the global structure of data can more effectively capture the subspace structure compared to the sparse structure. The above results indicate that LRR is more suitable for analyzing cancer gene expression data and thereby validates the superiority of JLSLPS in performing sample clustering based on LRR.

(3) Among the LRR-based methods, algorithms with graph regularization, i.e., LLRR, sgLRR, NSLRG, SGRLRSC, and the proposed JLSLPS, outperform LRR. Compared with LRR, these algorithms increase the average ACC values by 5.27%, 6.34%, 6.61%, 6.85%, and 7.95%, and the average NMI values by 6.07%, 8.13%, 9.52%, 9.15%, and 12.21%, respectively. Besides, JLSLPS outperforms S3C2, which is without the graph regularization term. The reason is that the graph regularization can maintain the local manifold structure of high-dimensional data in low-dimensional space, allowing local geometric information between data to be preserved in the representation matrix [8], [24], [25]. This enhances the separability between different cancer subtypes [24]. Furthermore, the joint learning-based LRS^3C,

SGRLRSC, S3C2, and the proposed JLSLPS outperform the two-stage methods LRR, LLRR, sgLRR, and NSLRG. Specifically, NSLRG can be regarded as the two-stage version of JLSLPS without the joint optimization term. Compared with NSLRG, JLSLPS improves the average ACC and NMI values by 0.98% and 2.70%, respectively. For the joint learning-based methods, the representation matrix and segmentation matrix can be optimized alternately with the guidance of each other. Therefore, joint learning can synergistically improve the performance of both tasks by fully considering their inherent relation. The above experimental results validate the advantage of the joint learning-based algorithms in performing high-quality sample clustering.

(4) Compared with the joint learning-based algorithms S^3C, LRS^3C, SGRLRSC, S3C2, and the proposed JLSLPS algorithm has significant advantages, with an improvement of 1.92%, 1.55%, 1.09%, 1.00% and 5.41%, 4.06%, 3.06%, 3.01% in ACC and NMI, respectively. To illustrate the advantage of the learned structure of JLSLPS, visualization of the representation matrices of S^3C, LRS^3C, SGRLRSC, S3C2, and JLSLPS are plotted on the Leukemia dataset, as shown in Fig. 2. The diagonal-block structure of Fig. 2(e) is more obvious and thus it can provide more discriminative information for cancer samples clustering. The improved joint optimization term is completely equivalent to the spectral clustering objective and precisely represents the segmentation cost, which can more effectively guide the co-optimization of the representation matrix and segmentation matrix in the joint framework. Comparing Fig. 2(a) with (e), it can be seen that the representation matrix learned by S^3C is sparser. However, S^3C only enforces the sparse constraint on the representation matrix, lacking guidance on the global structure of the cancer gene expression data. Unlike S^3C, JLSLPS gains the global information of data based on LRR while adding a sparse constraint as a regularization. Therefore, JLSLPS can take into account both the low-rankness and sparsity of the representation matrix, thereby achieving a better performance.

(5) From a realistic biological perspective, the specific number of clusters d is usually unknown when performing cancer samples clustering or subtype identification. A common approach in this case is to determine the variation range of d, and the d corresponding to the optimal value of an index can be chosen as the optimal number of clusters. Based on this approach, we compared the clustering results at different numbers of clusters to validate the biological significance of JLSLPS on the Leukemia dataset. Fig. 3 and Fig. 4 show visualizations and the clustering results (ACC and NMI) of different numbers of clusters, respectively. As can be seen in Fig. 3 and Fig. 4, in the case where the true number of clusters is unknown, the optimal values of ACC and NMI are obtained when $d = 6$. Moreover, the visualization effect of Fig. 3(c) achieves a more discriminative result. It is consistent with the true number of subtypes on the Leukemia dataset and further demonstrates that JLSLPS has a high clustering accuracy. The above results verify the practical relevance of our method in the field of cancer subtype identification.

D. PARAMETER SELECTION AND CONVERGENCE ANALYSIS

In JLSLPS, the four parameters α, λ_1, λ_2, and λ_3 are involved and affect the efficacy of the joint optimization term, sparse constraint term, graph regularization term, and the error term, respectively. This section performs grid search [25] on the parameters to evaluate their sensitivities and select the optimal values.

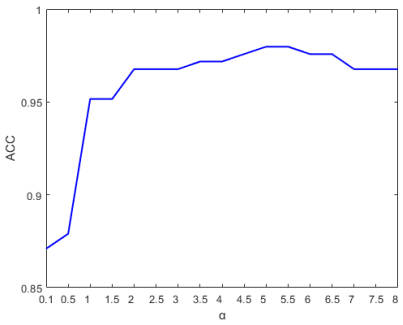
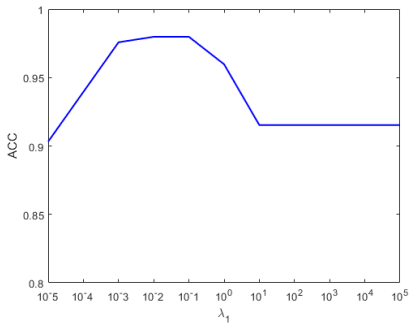
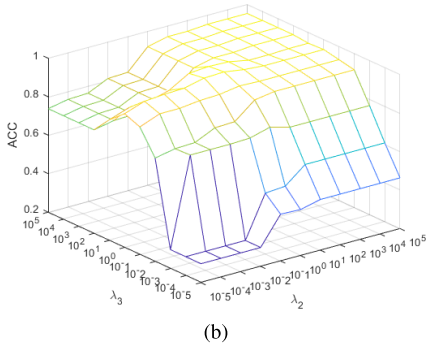


FIGURE 5. ACC of JLSLPS on the Leukemia dataset with different values of α. The x-axis and y-axis represent the different values of α and the ACC values, respectively.



(a)



(b)

FIGURE 6. ACC of JLSLPS on the Leukemia dataset with different parameter settings. (a) λ_1, (b) λ_2 and λ_3. The x-axis and y-axis in Fig. 4(a) represent the different values of λ_1 and the ACC values, respectively. The x-axis, y-axis, and z-axis in Fig. 4(b) represent the different values of λ_2, the different values of λ_3, and the ACC values, respectively.

TABLE 7. Clustering results (ACC) of JLSC and JLSLPS on eight cancer gene expression datasets. The optimal results are marked with bold.

	CH-ES-PA	AC-CH-ES	ES-HN-CH	CH-PA-RE-HN	AC-ES-CO-CH	Leukemia	Lung Cancer	Novartis BPLC
JLSC	0.9867	0.9887	0.9812	0.9788	0.9747	0.9677	0.9492	0.9709
JLSLPS	**1.0000**	**0.9933**	**0.9887**	**0.9827**	**0.9883**	**0.9798**	**0.9594**	**0.9806**

TABLE 8. Clustering results (NMI) of JLSC and JLSLPS on eight cancer gene expression datasets. The optimal results are marked with bold.

	CH-ES-PA	AC-CH-ES	ES-HN-CH	CH-PA-RE-HN	AC-ES-CO-CH	Leukemia	Lung Cancer	Novartis BPLC
JLSC	0.9486	0.9454	0.9253	0.9223	0.9237	0.9354	0.8196	0.9109
JLSLPS	**1.0000**	**0.9694**	**0.9443**	**0.9380**	**0.9606**	**0.9472**	**0.8451**	**0.9495**

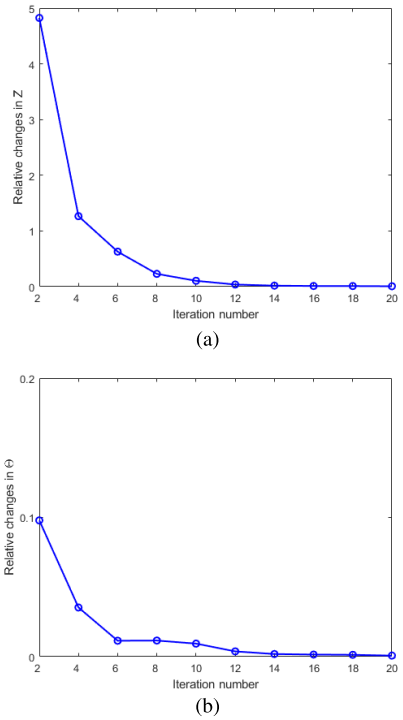


FIGURE 7. Convergence curves of matrices (a) Z and (b) Θ of JLSLPS on the Leukemia dataset. The x-axis and y-axis represent the iteration number and the relative change in Z and Θ, respectively.

Taking the Leukemia dataset as an example, we first fix the other parameters and determine α with a candidate set of $\{0.1, 0.5, 1, 1.5, 2, 2.5, 3, 3.5, 4, 4.5, 5, 5.5, 6, 6.5, 7, 7.5, 8\}$. Fig. 5 shows the ACC results of JLSLPS with different values of α on the Leukemia dataset. α is the parameter to balance the importance of the joint optimization term, which is the key to realizing joint learning. As can be observed, with the increase of α, ACC shows an increasing trend. Specifically, the clustering accuracy remains stable and satisfactory when $\alpha \in [2, 8]$. Particularly, the accuracy arrives at a peak when $\alpha \in [5, 5.5]$. If $\alpha < 2$, the contribution of the joint optimization term will be weakened and the clustering accuracy will significantly decrease. Thus, the joint optimization term can effectively improve model performance.

For the grid search of λ_1, λ_2, and λ_3, we set the range of the three parameters within $[10^{-5}, 10^5]$ with step size 10^{-1}. We perform two experiments to evaluate their sensitivities. First, fixing λ_2 and λ_3, we analyze the effect of λ_1 on clustering accuracy, then we fix λ_1 to analyze λ_2 and λ_3. The effects of the parameters on the performance of JLSLPS on the Leukemia dataset are given in Fig. 6. It can be seen that the optimal performance is achieved when λ_1 changes within the range of $[10^{-3}, 10^{-1}]$. This is because if λ_1 is quite high, the global information of the representation matrix may be affected. In addition, we find that the best performance is achieved at the range where $\lambda_2 > 10^{-1}$ and $\lambda_3 > 1$. The clustering accuracy is relatively stable within this range and declines significantly in other ranges.

To show the convergence behavior of the proposed JLSLPS algorithm, we demonstrate the relative changes of representation matrix Z and subspace structure matrix Θ defined in Eqs. (29) and (30). Fig. 7 gives the convergence curves on the Leukemia dataset.

As can be seen in Fig. 7, the relative changes of Z and Θ decrease sharply at the beginning and become smooth with increasing iteration numbers. The two matrices basically stop changing in approximately 20 iterations and the algorithm finally converges. The experimental results confirm the convergence of JLSLPS, and the optimal solution can be obtained after a few iterations.

E. ABLATION EXPERIMENT

To explore the impact of the precise representation of segmentation cost in joint learning on the performance of JLSLPS, an ablation experiment is conducted in this paper. The ablation algorithm, i.e., JLSLPS without non-negative and symmetric constraints, is referred to as JLSC. Similar to S^3C, the joint optimization term of JLSC degenerates to the subspace structure norm defined in Eq. (4). Therefore, z_{ij} of the joint optimization term in Eq. (7) is replaced with $|z_{ij}|$ to approximate the spectral clustering objective by the absolute value of the representation coefficients.

As shown in Tables 7 and 8, the experimental results show that the proposed JLSLPS algorithm achieves better clustering performance over JLSC, which confirms the positive effect of non-negative and symmetric constraints in the joint learning framework. By introducing the non-negative and symmetric constraints, the discriminability and interpretability of the representation matrix are improved. Additionally,

these two constraints enable the segmentation cost to be precisely represented, which can reflect a more informative subspace structure for the joint learning framework.

V. CONCLUSION

In this article, we reconsider the methods for designing the joint optimization term in existing joint learning methods and propose a method for a precise representation of segmentation cost in the joint learning framework, called JLSLPS. Specifically, by imposing non-negative and symmetric constraints on the representation matrix, the representation coefficients can be directly applied to the joint learning as an equivalent replacement of the affinity. The improved joint optimization term can precisely represent the segmentation cost, which contributes to the joint learning of data affinity and segmentation. We analyze the differences between the existing related methods and the proposed method to point out the superiorities of JLSLPS. Finally, plenty of comparative experiments on cancer gene expression datasets demonstrate the effectiveness of the proposed method.

There are many potential improvements in our future works. For instance, we will consider improving the nuclear norm in LRR to enhance robustness to the noise [39] and further capture the non-linear structure of data [23]. Moreover, we will consider extending the proposed method to multi-view clustering of cancer gene expression data to further improve the clustering performance.

APPENDIX

In this section, we describe a theoretical convergence proof of the proposed Algorithm 1.

Proposition 1: Algorithm 1 is convergent, and the sequence $\{Z^{(k)}, J^{(k)}, E^{(k)}\}$generated by Algorithm 1 would convergent to a stationary point of Eq (9).

Proof: Algorithm 1 aims to minimize the Lagrangian function of Eq. (9) by alternately updating the variables $\{Z, J, E\}$. Based on the previous optimization process, we have

$$\begin{aligned} Z^{(k+1)} &= arg\min_{Z} \mathcal{L}\left(\boldsymbol{Z}^{(k)}, \boldsymbol{J}^{(k)}, \boldsymbol{E}^{(k)}\right) \\ &= arg\min_{Z} \|Z\|_* + \frac{\eta}{2}\left\|\boldsymbol{Z} - \left(\boldsymbol{Z}^{(k)}\right.\right. \\ &\quad \left.\left. - \frac{\partial f(\boldsymbol{Z})^{(k)}}{\partial \boldsymbol{Z}^{(k)}}/\eta\right)\right\|_F^2 \end{aligned} \tag{33}$$

Note that $\mathcal{L}\left(\boldsymbol{Z}^{(k)}, \boldsymbol{J}^{(k)}, \boldsymbol{E}^{(k)}\right)$ is a λ-strongly convex with respect to Z. The following inequality holds:

$$\begin{aligned} \mathcal{L}\left(\boldsymbol{Z}^{(k+1)}, \boldsymbol{J}^{(k)}, \boldsymbol{E}^{(k)}\right) &\leq \mathcal{L}\left(\boldsymbol{Z}^{(k)}, \boldsymbol{J}^{(k)}, \boldsymbol{E}^{(k)}\right) \\ &\quad - \frac{\lambda}{2}\left\|\boldsymbol{Z}^{(k+1)} - \boldsymbol{Z}^{(k)}\right\|_F^2 \end{aligned} \tag{34}$$

According to updating schemes for the rest variables, it can be seen that J and E have the similar properties to Z. Therefore, the corresponding inequalities of them similar to Eq. (34) would hold. By accumulating these inequalities, we can obtain

$$\begin{aligned} &\mathcal{L}\left(\boldsymbol{Z}^{(k+1)}, \boldsymbol{J}^{(k+1)}, \boldsymbol{E}^{(k+1)}\right) \\ &\quad \leq \mathcal{L}\left(\boldsymbol{Z}^{(k)}, \boldsymbol{J}^{(k)}, \boldsymbol{E}^{(k)}\right) \\ &\qquad - \left(\frac{\lambda}{2}\left\|Z^{(k+1)} - \boldsymbol{Z}^{(k)}\right\|_F^2 + \left\|J^{(k+1)} - \boldsymbol{J}^{(k)}\right\|_F^2\right. \\ &\qquad \left. + \left\|E^{(k+1)} - \boldsymbol{E}^{(k)}\right\|_F^2\right) \end{aligned} \tag{35}$$

Accordingly, $\mathcal{L}\left(\boldsymbol{Z}^{(k)}, \boldsymbol{J}^{(k)}, \boldsymbol{E}^{(k)}\right)$ is monotonically decreasing, which further indicates it is upper bounded. This shows that $Z^{(k)}$, $J^{(k)}$, and $E^{(k)}$ are also bounded. Then, summing Eq. (35) over $k = 0, 1, \ldots$, we can obtain

$$\begin{aligned} &\sum_{k=1}^{+\infty}\frac{\lambda}{2}\left(\left\|Z^{(k+1)} - \boldsymbol{Z}^{(k)}\right\|_F^2 + \left\|J^{(k+1)} - \boldsymbol{J}^{(k)}\right\|_F^2\right. \\ &\qquad \left. + \left\|E^{(k+1)} - \boldsymbol{E}^{(k)}\right\|_F^2\right) \leq \mathcal{L}\left(\boldsymbol{Z}^{(0)}, \boldsymbol{J}^{(0)}, \boldsymbol{E}^{(0)}\right) \end{aligned} \tag{36}$$

When $k \to +\infty$, $Z^{(k+1)} - Z^{(k)} \to 0$, $J^{(k+1)} - J^{(k)} \to 0$, and $E^{(k+1)} - E^{(k)} \to 0$. According to the definition, obviously, $\mathcal{L}\left(\boldsymbol{Z}^{(k)}, \boldsymbol{J}^{(k)}, \boldsymbol{E}^{(k)}\right)$ is non-negative. To sum up, the sequence $\{Z^{(k)}, J^{(k)}, E^{(k)}\}$ would convergent to a stationary point, and the convergence of Algorithm 1 is guaranteed.

REFERENCES

[1] F. Bray, J. Ferlay, I. Soerjomataram, R. L. Siegel, L. A. Torre, and A. Jemal, "Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries," *CA, Cancer J. Clinicians*, vol. 68, no. 6, pp. 394–424, Nov. 2018.

[2] F. Yue, "State-of-the-art of cluster analysis of gene expression data," *Acta Autom. Sinica*, vol. 34, no. 2, pp. 113–120, Feb. 2008.

[3] D. Jiang, C. Tang, and A. Zhang, "Cluster analysis for gene expression data: A survey," *IEEE Trans. Knowl. Data Eng.*, vol. 16, no. 11, pp. 1370–1386, Nov. 2004.

[4] M. C. de Souto, I. G. Costa, D. S. de Araujo, T. B. Ludermir, and A. Schliep, "Clustering cancer gene expression data: A comparative study," *BMC Bioinf.*, vol. 9, no. 1, p. 497, Dec. 2008.

[5] 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, Mar. 2013, Art. no. e59377.

[6] X. Chen, L. Lin, and X. Ye, "Gene expression data subspace clustering based on smooth neighbor representation," *J. Control. Decis.*, vol. 32, no. 7, pp. 1235–1240, 2017.

[7] M. Yin, J. Gao, and Z. Lin, "Laplacian regularized low-rank representation and its applications," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 38, no. 3, pp. 504–517, Mar. 2016.

[8] J. Wang, J.-X. Liu, X.-Z. Kong, S.-S. Yuan, and L.-Y. Dai, "Laplacian regularized low-rank representation for cancer samples clustering," *Comput. Biol. Chem.*, vol. 78, pp. 504–509, Feb. 2019.

[9] M. Yin, H. Wu, S. Xie, and Q. Yang, "Self-attention adversarial based deep subspace clustering," *Acta Autom. Sinica*, vol. 48, no. 1, pp. 271–281, Jan. 2022.

[10] N. Pang, J.-F. Zhang, and X. Qin, "A subspace clustering algorithm of categorical data using multiple attribute weights," *Acta Autom. Sinica*, vol. 44, no. 3, pp. 517–532, Mar. 2018.

[11] W. Wang, X. Li, X. Feng, and W. Siqi, "A survey on sparse subspace clustering," *Acta Autom. Sinica*, vol. 41, no. 8, pp. 1373–1384, Aug. 2015.

[12] 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.

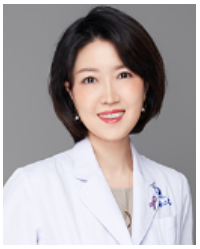
[13] G. Liu, Z. Lin, and Y. Yu, "Robust subspace segmentation by low-rank representation," in *Proc. 27th Int. Conf. Mach. Learn.*, 2010, pp. 663–670.

[14] 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.
[15] Z. Kong, D. Chang, Z. Fu, J. Wang, Y. Wang, and Y. Zhao, "Projection-preserving block-diagonal low-rank representation for subspace clustering," *Neurocomputing*, vol. 526, pp. 19–29, Mar. 2023.
[16] Z. Fu, Y. Zhao, D. Chang, Y. Wang, and J. Wen, "Latent low-rank representation with weighted distance penalty for clustering," *IEEE Trans. Cybern.*, vol. 53, no. 11, pp. 6870–6882, May 2022.
[17] J. Wang, J.-X. Liu, C.-H. Zheng, C.-H. Lu, L.-Y. Dai, and X.-Z. Kong, "Block-constraint Laplacian-regularized low-rank representation and its application for cancer sample clustering based on integrated TCGA data," *Complexity*, vol. 2020, pp. 1–13, Jan. 2020.
[18] T. Wang, J. Zhang, and K. Huang, "Generalized gene co-expression analysis via subspace clustering using low-rank representation," *BMC Bioinf.*, vol. 20, no. S7, pp. 17–27, May 2019.
[19] J. Wang, J.-X. Liu, C.-H. Zheng, Y.-X. Wang, X.-Z. Kong, and C.-G. Wen, "A mixed-norm Laplacian regularized low-rank representation method for tumor samples clustering," *IEEE/ACM Trans. Comput. Biol. Bioinf.*, vol. 16, no. 1, pp. 172–182, Jan. 2019.
[20] D. Luo, F. Nie, C. Ding, and H. Huang, "Multi-subspace representation and discovery," in *Proc. Joint Eur. Conf. Mach. Learn. Knowl. Discovery Databases*, 2011, pp. 405–420.
[21] L. Zhuang, H. Gao, Z. Lin, Y. Ma, X. Zhang, and N. Yu, "Non-negative low rank and sparse graph for semi-supervised learning," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2012, pp. 2328–2335.
[22] J. Chen, H. Mao, Y. Sang, and Z. Yi, "Subspace clustering using a symmetric low-rank representation," *Knowl.-Based Syst.*, vol. 127, pp. 46–57, Jul. 2017.
[23] T. Wang, X. Zhang, L. Lan, Q. Liao, C. Xu, and Z. Luo, "Correlation self-expression shrunk for subspace clustering," *IEEE Access*, vol. 8, pp. 16595–16605, 2020.
[24] J. Wang, C.-H. Lu, J.-X. Liu, L.-Y. Dai, and X.-Z. Kong, "Multi-cancer samples clustering via graph regularized low-rank representation method under sparse and symmetric constraints," *BMC Bioinf.*, vol. 20, no. S22, pp. 1–15, Dec. 2019.
[25] C. Lu, J. Wang, J. Liu, C. Zheng, X. Kong, and X. Zhang, "Non-negative symmetric low-rank representation graph regularized method for cancer clustering based on score function," *Frontiers Genet.*, vol. 10, p. 1353, Jan. 2020.
[26] C.-G. Li, C. You, and R. Vidal, "Structured sparse subspace clustering: A joint affinity learning and subspace clustering framework," *IEEE Trans. Image Process.*, vol. 26, no. 6, pp. 2988–3001, Jun. 2017.
[27] J. Zhang, C. Li, H. Zhang, and J. Guo, "Low-rank and structured sparse subspace clustering," in *Proc. IEEE Vis. Commun. Image Process.*, Mar. 2016, pp. 1–4.
[28] J. Liu and S. Ma, "Structured graph regularized low-rank subspace clustering," *Comput. Appl. Eng. Educ.*, vol. 54, no. 18, pp. 1–7, 2018.
[29] J. Zhuang, L. Cui, T. Qu, C. Ren, J. Xu, T. Li, G. Tian, and J. Yang, "A streamlined scRNA-seq data analysis framework based on improved sparse subspace clustering," *IEEE Access*, vol. 9, pp. 9719–9727, 2021.
[30] U. von Luxburg, "A tutorial on spectral clustering," *Statist. Comput.*, vol. 17, no. 4, pp. 395–416, Aug. 2007.
[31] Z. Lin, R. Liu, and Z. Su, "Linearized alternating direction method with adaptive penalty for low-rank representation," in *Proc. 24th Int. Conf. Neural Inf. Process. Syst.*, 2011, pp. 612–620.
[32] Z. Lin, M. Chen, and Y. Ma, "The augmented Lagrange multiplier method for exact recovery of corrupted low-rank matrices," 2010, *arXiv:1009.5055*.
[33] 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, Mar. 2010.
[34] K. Tomczak, P. Czerwińska, and M. Wiznerowicz, "Review the cancer genome atlas (TCGA): An immeasurable source of knowledge," *Contemp. Oncol.*, vol. 19, no. 1, pp. 68–77, Jan. 2015.
[35] E.-J. Yeoh et al., "Classification, subtype discovery, and prediction of outcome in pediatric acute lymphoblastic leukemia by gene expression profiling," *Cancer Cell*, vol. 1, no. 2, pp. 133–143, Mar. 2002.
[36] A. Bhattacharjee, W. G. Richards, J. Staunton, C. Li, S. Monti, P. Vasa, C. Ladd, J. Beheshti, R. Bueno, M. Gillette, M. Loda, G. Weber, E. J. Mark, E. S. Lander, W. Wong, B. E. Johnson, T. R. Golub, D. J. Sugarbaker, and M. Meyerson, "Classification of human lung carcinomas by mRNA expression profiling reveals distinct adenocarcinoma subclasses," *Proc. Nat. Acad. Sci. USA*, vol. 98, no. 24, pp. 13790–13795, Nov. 2001.
[37] A. I. Su, M. P. Cooke, K. A. Ching, Y. Hakak, J. R. Walker, T. Wiltshire, A. P. Orth, R. G. Vega, L. M. Sapinoso, A. Moqrich, A. Patapoutian, G. M. Hampton, P. G. Schultz, and J. B. Hogenesch, "Large-scale analysis of the human and mouse transcriptomes," *Proc. Nat. Acad. Sci. USA*, vol. 99, no. 7, pp. 4465–4470, 2002.
[38] W. Xu, X. Liu, and Y. Gong, "Document clustering based on non-negative matrix factorization," in *Proc. 26th Annu. Int. ACM SIGIR Conf. Res. Develop. Inf. Retr.*, Jul. 2003, pp. 267–273.
[39] X. Zhi, L. Bi, and J. Fan, "$l_{2,p}$-norm based discriminant subspace clustering algorithm," *IEEE Access*, vol. 8, pp. 76043–76055, 2020.

YANMING CAO was born in Jilin, Jilin, China, in 1999. He received the B.S. degree in control engineering from the Dalian University of Technology, Liaoning, China, where he is currently pursuing the M.S. degree under the guidance of Prof. Dan Li. His research interests include subspace clustering and bioinformatics.

DAN LI received the B.S., M.S., and Ph.D. degrees from the Dalian University of Technology, Dalian, China, in 2000, 2003, and 2011, respectively. She is currently an Associate Professor with the Faculty of Electronic Information and Electrical Engineering, Dalian University of Technology. Her research interests include pattern recognition, machine learning, and bioinformatics.

ZEYUAN WANG received the M.S. degree from the Dalian University of Technology, Dalian, China, where he is currently pursuing the Ph.D. degree with the Faculty of Electronic Information and Electrical Engineering. His research interests include bioinformatics, machine learning, and statistical analysis.



JIA WANG received the B.S. and M.S. degrees from China Medical University, Shenyang, China, and the Ph.D. degree from Kagoshima University, Japan. She is currently an Associate Chief Physician with The Second Hospital of Dalian Medical University. Her research interests include the diagnosis and treatment of breast cancer.