Multi-Kernel Tensor Fusion on Grassmann Manifold for Genomic Data Clustering

1st Fei Qi
South China University of Technology
Guizhou Minzu University
Guangdong, China
fqiscut@foxmail.com

2nd Junyu Li
South China University of Technology
Guangdong, China
chunyulee018@foxmail.com

3rd Yi Liao
South China University of Technology
Guangdong, China
liaoy@mail.scut.edu.cn

4th Wenxiong Liao
South China University of Technology
Guangdong, China
cswxliao@mail.scut.edu.cn

5th Jiazhou Chen
South China University of Technology
Guangdong, China
csjzchen@scut.edu.cn

6th Hongmin Cai
South China University of Technology
Guangdong, China
hmcai@scut.edu.cn

2023 IEEE International Conference on Bioinformatics and Biomedicine (BIBM) | 979-8-3503-3748-8/23/$31.00 ©2023 IEEE | DOI: 10.1109/BIBM58861.2023.10385751

***Abstract*—Due to the inherent high-dimensional characteristics of genomic data, traditional single metric/kernel-based clustering methods fail to accurately perform data analysis. To address this issue, we propose a multi-kernel clustering with tensor fusion on the Grassmann manifold (MKCTM). Specifically, multiple kernel functions are employed to map data into different kernel spaces and utilize tensor representations to capture their high-order relationships. By introducing a tensor low-rank constraint, we maximize the correlation among kernels while separating the noise and redundancy information from kernel tensor. Finally, the learned kernel tensor is fused on the Grassmann manifold to obtain the final k ernel m atrix f or e nhancing c lustering. We integrate tensor learning and tensor fusion steps into a unified optimization model and propose an efficient iterative optimization algorithm to solve it. Our proposed method is evaluated on six high-dimensional gene expression datasets against eight popular baseline methods. The remarkable experimental performance demonstrates the exceptional effectiveness of our approach. Our code is available at https://github.com/foureverfei/MKCTM.git**

***Index Terms*—Genomic data clustering, tensor, grassmann manifold**

I. INTRODUCTION

Precise analysis of genomic data has become a hot topic in the field o f b ioinformatics i n r ecent y ears [1]. Genomic data provide a multifaceted characterization for types and markers at the molecular level [2], therefore, analyzing genomic data through clustering methods enables the grouping of gene expression patterns, facilitating disease classification and the discovery of novel subtypes, which holds significant importance in personalized treatment and precise medicine.

K-means [3] and Spectral Clustering (SC) [4] are two fundamental clustering strategies. K-means iteratively updates cluster centers by solving the distance between each sample and the cluster centers. SC, as a benchmark strategy in graph-based methods, constructs an affinity graph of samples based on distance metrics and transforms the problem into a graph minimum cut problem. Both methods have been widely studied and have generated numerous variants due to their interpretability and ease of implementation [5]. However, these methods heavily rely on accurate measurements of data relationships, leading to a significant challenge when dealing with genomic data with high dimensionality.

Many efforts have been made to address the decrease in clustering performance caused by inaccurate initial measurements. Kernel methods [6] aim to map inseparable data in the original space to a high-dimensional Reproducing Kernel Hilbert Space (RKHS) to obtain clear separation boundaries for clustering. Meanwhile, graph-based methods [7] learn the constructed affinity matrix more suitable clustering objectives to achieve a more accurate representation of relationships. For example, Nie et al. [8] utilized adaptive neighbors for each sample and dynamically learned an affinity matrix. Lu et al. [9] designed a diagonal norm to constrain the affinity graph for maintaining a block structure that is more appropriate for clustering. The kernel matrix constructed using the Kernel trick can serve as the initialized affinity graph in graph-based methods due to its positive semidefinite. Previous studies [10] have also demonstrated the connection between Kernel k-means and spectral clustering. However, the effectiveness of learning the graph in graph-based methods highly depends on the construction of initial graph, and different kernel function choices yield different learning results. Furthermore, when dealing with high-dimensional data, the noise in the data itself and redundant information generated during the mapping process both affect the effectiveness of clustering.

Recently, many studies have proposed to construct multiple affinity graphs by a set of base kernels and integrate them into a consensus graph for enhanced clustering. For example, Liu et al. [11] proposed an approach to gener-

This work was supported in part by the National Key Research and Development Program of China (2022YFE0112200), the National Natural Science Foundation of China (U21A20520, 62325204, 62102153, 62272326), Science and Technology Project of Guangdong Province (2022A0505050014), the Key-Area Research and Development Program of Guangzhou City (202206030009), the Natural Science Foundation of Guangdong Province of China (2022A1515011162), and the China Postdoctoral Science Foundation (2021M691062, 2023T160226). (Corresponding author: Jiazhou Chen.)

ate robust representations from multiple kernel matrices and then perform co-training subspace clustering. Liu et al. [12] designed a multi-kernel-based strategy to construct adaptive local kernels for clustering by considering each sample's local density. Nevertheless, these methods have failed to examine the common and unique information among kernels from a high-order perspective simultaneously during the generation of the final consensus kernel. Besides, they have been ineffective in removing noise and redundancy during the learning process, leading to unsatisfactory results when dealing with high-dimensional genomic data.

To address the challenge of accurate clustering on high-dimensional genomic data, we propose a multi-kernel clustering with tensor fusion on Grassman manifold (MKCTM). Specifically, a set of affinity matrices is generated by multiple kernel functions to observe sample relationships from different spaces. These affinity matrices are then stacked into a kernel tensor. By applying a tensor low-rank constraint, we maximize the consensus information among the kernel matrices while separating noise and redundancy information from kernel tensor. Then, a manifold fusion strategy is designed to fuse the affinity matrices from different kernels on the Grassmann manifold. We further develop an efficient iterative solving strategy for this problem and demonstrate the effectiveness of our proposed tensor-based multi-kernel manifold fusion clustering framework on high-dimensional genomic data.

The main contributions of our method are summarized as follows:

- We utilize multiple kernel functions to map high-dimensional data into different kernel spaces and use tensor representation to comprehensively characterize the relationships among samples from a high-order perspective. By introducing tensor low-rank constraint, the kernel tensor is decomposed into a consensus kernel tensor and a sparse noise term. The learned kernel tensor can align the consensus information between kernels for clustering.
- In contrast to traditional weighted fusion methods for tensors, we propose to perform kernel space fusion on the Grassmann Manifold to learn a common kernel matrix for clustering.
- Tensor learning and tensor fusion step are integrated into a unified framework and an efficient iterative minimization strategy is proposed to decompose the problem into several subproblems for solving. Each subproblem has a corresponding closed-form solution, making this strategy highly efficient.
- Experimental results on multiple high-dimensional genomic datasets demonstrate the effectiveness of our proposed method. Our approach achieves remarkable performance compared to baseline methods across different evaluation metrics.

II. RELATED WORK

A. Notation and Preliminarise

Throughout the paper, lowercase letters represent scalars, while bold lowercase letters represent vectors, i.e. $\boldsymbol{v} \in \mathbb{R}^m$. Bold uppercase letters denote matrices, i.e., $\boldsymbol{M} \in \mathbb{R}^{m\times n}$, while bold calligraphy letters represent third-order tensors, i.e., $\boldsymbol{\mathcal{T}} \in \mathbb{R}^{m\times n\times l}$. For an third-order tensor $\boldsymbol{\mathcal{T}}$, $\boldsymbol{\mathcal{T}}(:,:,i), \boldsymbol{\mathcal{T}}(:,i,:), \boldsymbol{\mathcal{T}}(i,:,:)$ is used to represent the i-th frontal, lateral and horizontal slices, respectively. In particular, the frontal slices $\boldsymbol{\mathcal{T}}(:,:,i)$ and the lateral slices $\boldsymbol{\mathcal{T}}(:,j,:)$ are abbreviated as $\boldsymbol{T}^{(i)}$ and $\boldsymbol{T}_{(i)}$, respectively. We now introduce elementary operations on tensors.

Definition II.1. ***Tensor singular value decomposition [13]:*** *A tensor $\boldsymbol{\mathcal{T}} \in \mathbb{R}^{n_1\times n_2\times n_3}$ can be factorized into three tensors as follows:*

$$\boldsymbol{\mathcal{T}} = \boldsymbol{\mathcal{U}} * \boldsymbol{\mathcal{S}} * \boldsymbol{\mathcal{V}}^T, \tag{1}$$

where $\boldsymbol{\mathcal{S}} \in \mathbb{R}^{n_1\times n_2\times n_3}$ is the f-diagonal tensor and $\boldsymbol{\mathcal{U}} \in \mathbb{R}^{n_1\times n_1\times n_3}$ and $\boldsymbol{\mathcal{V}} \in \mathbb{R}^{n_2\times n_2\times n_3}$ are orthogonal. We call this decomposition tensor singular value decomposition (t-SVD).

Definition II.2. ***Tensor nuclear norm [13]:*** *For a third-order tensor $\boldsymbol{\mathcal{T}} \in \mathbb{R}^{n_1\times n_2\times n_3}$, its nuclear norm, denoted by $\|\boldsymbol{\mathcal{T}}\|_{\circledast}$, is defined as the sum of all frontal slices of $\boldsymbol{\mathcal{T}}_f$:*

$$\|\boldsymbol{\mathcal{T}}\|_{\circledast} = \sum_{k=1}^{n_3} \|\boldsymbol{T}_f^{(k)}\|_* = \sum_{i=1}^{\min(n_1,n_2)} \sum_{k=1}^{n_3} |\boldsymbol{T}_f^{(k)}(i,i)|. \tag{2}$$

where $\boldsymbol{\mathcal{T}}_f = \mathrm{fft}(\boldsymbol{\mathcal{T}}, [], 3)$ is the fast Fourier transform on $\boldsymbol{\mathcal{T}}$ along the mode-3 dimension. Conversely, the tensor $\boldsymbol{\mathcal{T}}$ can also be obtained by $\boldsymbol{\mathcal{T}} = \mathrm{ifft}(\boldsymbol{\mathcal{T}}_f, [], 3)$.

Definition II.3. *For a third-order tensor $\boldsymbol{\mathcal{T}} \in \mathbb{R}^{n_1\times n_2\times n_3}$, its tensor $L_{2,1}$-norm, denoted by $\|\boldsymbol{\mathcal{T}}\|_{2,1}$, is defined as the sum of all lateral slice's L_{21} norms:*

$$\|\boldsymbol{\mathcal{T}}\|_{2,1} = \sum_{i}^{n_3} \|\boldsymbol{\mathcal{T}}(:,i,:)\|_{2,1}, \tag{3}$$

B. Grassmann Manifold Merging

The Grassmann manifold $\boldsymbol{\mathcal{G}}(d,n)$ can be defined as a set of all the d-dimensional linear subspaces embedded in Euclidean space $\mathbb{R}^n$. An orthonormal matrix $U \in \mathbb{R}^{n\times d}$ is introduced to denote a point on the Grassmann manifold. The column vectors of $\boldsymbol{U}$ can span a subspace $span(\boldsymbol{U})$ [14]. Any two point in $\boldsymbol{\mathcal{G}}(d,n)$ with same linear space $span(\boldsymbol{U})$ are equivalent [15]. The geodesic distance of any two d-dimensional Grassmann subspaces $span(\boldsymbol{U}^{(1)})$ and $span(\boldsymbol{U}^{(2)})$ on Grassman manifold can be measured by a set of principal angles $\{\boldsymbol{\theta}_i\}_{i=1}^d$ between their subspaces [16]. The distance can be defined as:

$$\begin{aligned} dist^2_{proj}(\boldsymbol{U}^{(1)}, \boldsymbol{U}^{(2)}) &= \sum_{i=1}^{d} sin^2\boldsymbol{\theta}_i = d - \sum_{i=1}^{d} cos^2\boldsymbol{\theta}_i \\ &= d - tr(\boldsymbol{U}^{(1)}{\boldsymbol{U}^{(1)}}^T \boldsymbol{U}^{(2)}{\boldsymbol{U}^{(2)}}^T). \end{aligned} \tag{4}$$

Given k subspaces $\{\boldsymbol{U}^{(v)}\}_{v=1}^k$, a common subspace $\boldsymbol{U}$ should close to all these individual subspaces on the Grassmann manifold. the summation of geodesic distance between common subspace $\boldsymbol{U}$ with k subspaces $\boldsymbol{U}^{(v)}$ can be calculated as follows:

$$dist^2_{proj}(\boldsymbol{U}, \{\boldsymbol{U}^{(v)}\}_{v=1}^k) = \sum_{v=1}^{k} dist^2_{proj}(\boldsymbol{U}, \boldsymbol{U}^{(v)}) = kd - \sum_{v=1}^{k} tr(\boldsymbol{U}\boldsymbol{U}^T\boldsymbol{U}^{(v)}{\boldsymbol{U}^{(v)}}^T). \tag{5}$$

III. Multi-kernel Tensor Clustering on Grassman Manifold

In response to the limitation of traditional single metrics/kernels method dealing with genomic data, we propose Multi-Kernel Clustering with Tensor Fusion on the Grassmann manifold. In this section, we introduce our proposed MKCTM in detail, including the model formulation and optimization strategy.

A. Model Formulation

Mathematically, denote $\mathbf{X} \in \mathbb{R}^{n\times m}$ as the data matrix, including n samples with m-dimensional features from k clusters. The process of mapping data to Reproducing Kernel Hilbert Spaces (RKHS) using single kernel introduces noise and redundant information. In order to address this, we propose using multiple kernels to map the data into different spaces. By integrating consensus information from these spaces, the obtained consensus kernel tensor can eliminate noise and redundancy for describing data relationships accurately. Specifically, we first employ l different kernel function to map original data into multiple RKHS, yielding l different kernel matrices $\mathbf{T}^{(i)}$. Then these l kernel matrices are stacked into a multi-kernel tensor $\boldsymbol{\mathcal{T}} \in \mathbb{R}^{n\times l\times n}$. Based on tensor representation, we employ Tensor Robust Principal Component Analysis (TRPCA) [17] to separate the low-rank consensus kernel tensor and the error term from the kernel tensor $\boldsymbol{\mathcal{T}}$, which can be modeled as follows:

$$\min_{\boldsymbol{\mathcal{K}},\boldsymbol{\mathcal{E}}} \|\boldsymbol{\mathcal{K}}\|_{\circledast} + \|\boldsymbol{\mathcal{E}}\|_{2,1} \quad s.t. \boldsymbol{\mathcal{T}} = \boldsymbol{\mathcal{K}} + \boldsymbol{\mathcal{E}}. \tag{6}$$

We use the tensor low-rank constraint to maximize the consensus information among kernel matrices and separate the noise and redundancy. L_{21} norm on $\boldsymbol{E}^{(i)}$ is introduced to quantify the sparse noise on each kernel, which can be viewed from the tensor perspective as the tensor L_{21} norm on $\boldsymbol{\mathcal{E}}$.

Unlike traditional tensors that directly weigh the fusion of different kernels, we propose a fusion approach that minimizes the geodesic distance on the Grassmann manifold. Let the common space be spanned by a set of orthonormal bases $\boldsymbol{U}$, and the i-th kernel space is represented by basis $\boldsymbol{U}^{(i)}$. Since the positive semidefinite of the kernel matrix, we can treat it as an affinity matrix. This property can be ensured by $\boldsymbol{K}^{(i)} + {\boldsymbol{K}^{(i)}}^T$ after learning. Based on this, the affinity matrix in i-th space can be expressed as $\boldsymbol{U}^{(i)}{\boldsymbol{U}^{(i)}}^T = \boldsymbol{K}^{(i)} + {\boldsymbol{K}^{(i)}}^T$ by orthonormal basis $\boldsymbol{U}^{(i)}$, and the fusion affinity matrix in the common space is given by $\boldsymbol{K} = \boldsymbol{U}\boldsymbol{U}^T$. According to Eq. (**??**), we aim to minimize the sum of geodesic distances on the Grassmann manifold between the fused common space and each individual kernel space. We further introduce a sparse regularization term $\|\boldsymbol{K}\|_1$ to make the fused affinity matrix more discriminative. Thus, the problem of fusing kernel spaces on the manifold can be formulated as follows:

$$\min_{\boldsymbol{\alpha},\boldsymbol{K}} -\sum_{i=1}^{l} \boldsymbol{\alpha}_i tr((\boldsymbol{K}^{(i)} + {\boldsymbol{K}^{(i)}}^T)\boldsymbol{K}) + \gamma\|\boldsymbol{K}\|_1 \quad s.t.\ \boldsymbol{K} \in \mathcal{F}^d, \sum_{i=1}^{l} \boldsymbol{\alpha}^2 = 1, \tag{7}$$

where $\mathcal{F}^d := \{\boldsymbol{K} : trace(\boldsymbol{K}) = d \ and \ \boldsymbol{I} \geq \boldsymbol{K} \geq \boldsymbol{0}\}$, and the trade-off parameters γ is used to balance weight of sparse constraint with latent space learning term.

As stated in [18], $\boldsymbol{K}$ lies on a convex body called the Fantope, denoted as $\mathcal{F}^d$. Additionally, we introduce an adaptive fusion weight $\boldsymbol{\alpha}$ with constraints $\sum_{i=1}^{l} \boldsymbol{\alpha}^2 = 1$. $\boldsymbol{\alpha}$ is used to achieve weighted fusion of the kernel matrices in the manifold space. Combining Eq. (7) with the tensor consensus learning model in Eq. (6), the final objective function is constructed as follows:

$$\begin{aligned}\min_{\boldsymbol{\mathcal{K}},\boldsymbol{\mathcal{E}},\boldsymbol{\alpha},\boldsymbol{K}} \ & \|\boldsymbol{\mathcal{K}}\|_{\circledast} + \beta\|\boldsymbol{\mathcal{E}}\|_{2,1} \\ & - \lambda\sum_{i=1}^{l} \boldsymbol{\alpha}_i tr[(\boldsymbol{K}^{(i)} + {\boldsymbol{K}^{(i)}}^T)\boldsymbol{K}] + \gamma\|\boldsymbol{K}\|_1 \\ & s.t.\ \boldsymbol{\mathcal{T}} = \boldsymbol{\mathcal{K}} + \boldsymbol{\mathcal{E}}, \boldsymbol{\mathcal{K}}^{(i)} \geq \boldsymbol{0}, \boldsymbol{K} \in \mathcal{F}^d, \sum_{i=1}^{l} {\boldsymbol{\alpha}_i}^2 = 1,\end{aligned} \tag{8}$$

To ensure that the learned kernel is positive semidefinite, it is typically required that the learned kernel tensor $\boldsymbol{\mathcal{K}} \geq \boldsymbol{0}$. The trade-off parameters β, and λ are used to balance the different terms in the objective function. We integrate the learning of multi-kernel tensor and tensor fusion step into a uniform optimization problem, rather than a two-step learning strategy. By solving this uniform problem in Eq. (8), we obtain the fused kernel matrix $\boldsymbol{K}$, which can be directly utilized for clustering. This approach not only simplifies the solving process but also offers the following advantages: The learning of the fused kernel matrix can guide the consensus kernel tensor learning step, enabling it to more accurately eliminate redundant information. Conversely, the learning of the kernel tensor also can enhance the tensor fusion step, resulting in a more discriminative affinity matrix for enhancing clustering performance.

B. Numerical Scheme

Simultaneously optimizing all variables in model (8) is undoubtedly challenging. Therefore, we propose an efficient alternating optimization strategy. To facilitate the solution process, we introduce two auxiliary variables $\boldsymbol{\mathcal{Z}}$ and $\boldsymbol{B}$, where

$\boldsymbol{\mathcal{Z}} = \boldsymbol{\mathcal{K}}$ and $\boldsymbol{B} = \boldsymbol{K}$. We can write the augmented Lagrangain form of the objective function as follows:

$$\begin{aligned}\min_{\boldsymbol{\mathcal{K}},\boldsymbol{\mathcal{E}},\boldsymbol{\mathcal{Z}},\boldsymbol{\alpha},\boldsymbol{K},\boldsymbol{B}} \ & \|\boldsymbol{\mathcal{K}}\|_{\circledast} + \beta\|\boldsymbol{\mathcal{E}}\|_{2,1} \\ & - \lambda(\sum_{i=1}^{l} \boldsymbol{\alpha}^{(i)} tr[(\boldsymbol{Z}^{(i)} + \boldsymbol{Z}^{(i)^T})\boldsymbol{K}] + \gamma\|\boldsymbol{B}\|_1 \\ & + < \boldsymbol{\mathcal{Y}}_1, \boldsymbol{\mathcal{T}} - \boldsymbol{\mathcal{K}} - \boldsymbol{\mathcal{E}} > + \frac{\mu}{2}\|\boldsymbol{\mathcal{T}} - \boldsymbol{\mathcal{K}} - \boldsymbol{\mathcal{E}}\|_F^2 \\ & + < \boldsymbol{\mathcal{Y}}_2, \boldsymbol{\mathcal{Z}} - \boldsymbol{\mathcal{K}} > + \frac{\mu}{2}\|\boldsymbol{\mathcal{Z}} - \boldsymbol{\mathcal{K}}\|_F^2 \\ & + < \boldsymbol{Y}_3, \boldsymbol{K} - \boldsymbol{B} > + \frac{\mu}{2}\|\boldsymbol{K} - \boldsymbol{B}\|_F^2 \\ s.t. \ & \boldsymbol{\mathcal{K}}^{(i)} \geq \boldsymbol{0}, \boldsymbol{K} \in \mathcal{F}^d, \sum_{i=1}^{l} {\boldsymbol{\alpha}_i}^2 = 1\end{aligned} \tag{9}$$

where $\boldsymbol{\mathcal{Y}}_1$,$\boldsymbol{\mathcal{Y}}_2$,$\boldsymbol{Y}_3$ are augmented Lagrangain multipliers and $\mu > 0$ is the penalty parameters. By solving for each variable with fixing the remaining variables, the original problem is divided into six convex subproblems.

(1) Optimizing $\mathcal{K}$ with fixed irrelevant variables: After rearrangement, the subproblem $\mathcal{K}$ can be rewritten as follow:

$$\min_{\boldsymbol{\mathcal{K}}} \|\boldsymbol{\mathcal{K}}\|_{\circledast} + \frac{\mu}{2}\|\boldsymbol{\mathcal{K}} - (\boldsymbol{\mathcal{T}} - \boldsymbol{\mathcal{E}} + \frac{\boldsymbol{\mathcal{Y}}_1}{\mu})\|_F^2. \tag{10}$$

According to [19], this problem can be solved analytically using a tensor tubal-shrinkage operator based on t-SVD. The closed-form solution is given by:

$$\boldsymbol{\mathcal{K}} = \mathcal{C}_\sigma(\boldsymbol{\mathcal{T}} - \boldsymbol{\mathcal{E}} + \frac{\boldsymbol{\mathcal{Y}}_1}{\mu}) = \boldsymbol{\mathcal{U}} * \mathcal{C}_\sigma(\boldsymbol{\mathcal{S}}) * \boldsymbol{\mathcal{V}}^T, \tag{11}$$

where $\sigma = N \cdot \mu$ and $\boldsymbol{\mathcal{U}}$, $\boldsymbol{\mathcal{S}}$, $\boldsymbol{\mathcal{V}}$ are obtained from the t-SVD decomposition of $\boldsymbol{\mathcal{T}} - \boldsymbol{\mathcal{E}} + \frac{\boldsymbol{\mathcal{Y}}_1}{\mu}$, i.e., $\boldsymbol{\mathcal{T}} - \boldsymbol{\mathcal{E}} + \frac{\boldsymbol{\mathcal{Y}}_1}{\mu} = \boldsymbol{\mathcal{U}} * \boldsymbol{\mathcal{S}} * \boldsymbol{\mathcal{V}}^T$. We have $\mathcal{C}_\sigma(\boldsymbol{\mathcal{S}}) = \boldsymbol{\mathcal{S}} * \boldsymbol{\mathcal{J}}$, where $\boldsymbol{\mathcal{J}}$ is an f-diagonal tensor. After transforming $\boldsymbol{\mathcal{J}}$ into the Fourier domain, the diagonal elements of $\boldsymbol{\mathcal{J}}_f$ are given by:

$$\boldsymbol{\mathcal{J}}_f(i,i,j) = \max(1 - \frac{\sigma}{\boldsymbol{\mathcal{S}}_f^{(j)}}, 0) \tag{12}$$

(2) Optimizing $\mathcal{E}$ with fixed irrelevant variables: This problem can be solved in matrix form based on the lateral slice of $\boldsymbol{\mathcal{E}}$:

$$\min_{\boldsymbol{E}_{(i)}} \beta\|\boldsymbol{E}_{(i)}\|_{2,1} + \frac{\mu}{2}\|\boldsymbol{E}_{(i)} - (\boldsymbol{T}_{(i)} - \boldsymbol{K}_{(i)} + \frac{\boldsymbol{Y}_{1(i)}}{\mu})\|_F^2 \tag{13}$$

For each lateral slice, a quadratic minimization with structural sparsity constraints can be performed. According to [20], the closed-form solution for this problem is as follows:

$$\mathbf{E}_{(i):,j} = \begin{cases} \frac{\|\boldsymbol{O}_{:,j}\|_2 - \frac{\beta}{\mu}}{\|\boldsymbol{O}_{:,j}\|_2}\boldsymbol{O}_{:,j}, & \text{if } \|\boldsymbol{O}_{:,j}\|_2 > \frac{\beta}{\mu} \\ \boldsymbol{0}, & \text{otherwise} \end{cases}, \tag{14}$$

where $\boldsymbol{O} = \boldsymbol{T}_{(i)} - \boldsymbol{K}_{(i)} + \frac{\boldsymbol{Y}_{1(i)}}{\mu}$.

(3) Optimizing $\mathcal{Z}$ with fixed irrelevant variables: Similarly, this problem can also be transformed into matrix form along lateral slice and solved as follows [21]:

$$\boldsymbol{Z}_{(i)} = [\boldsymbol{K}_{(i)} + \frac{\boldsymbol{Y}_{\boldsymbol{2}(i)}}{\mu} + \frac{\lambda\boldsymbol{\alpha}_i}{\mu}(\boldsymbol{K} + \boldsymbol{K}^T)]_+. \tag{15}$$

TABLE I
STATISTICS OF THE SIX GENOMIC DATASETS.

Dataset	Type	Samples	Features	Clusters
GLIOMA	Bioinformatics	50	4,434	4
Colon	Bioinformatics	171	5,748	4
GLI-85	Bioinformatics	85	22,283	2
Leukemia	Bioinformatics	72	7,070	2
CLL-SUB-111	Bioinformatics	111	11,340	3
SMK-CAN-187	Bioinformatics	187	19,993	2

Let the $\boldsymbol{P} = \boldsymbol{K}_{(i)} + \frac{\boldsymbol{Y}_{\boldsymbol{2}(i)}}{\mu} + \frac{\lambda\boldsymbol{\alpha}_i}{\mu}(\boldsymbol{K} + \boldsymbol{K}^T)$. Through eigenvalue decomposition, we have $\boldsymbol{P}_+ = \boldsymbol{U}\boldsymbol{S}_+\boldsymbol{U}^T$, where $\boldsymbol{S}_+ = Diag(\boldsymbol{s}_+)$,$\boldsymbol{s}_+ = \max\{0, \boldsymbol{s}\}$.

(4) Optimizing $\boldsymbol{K}$ with fixed irrelevant variables: By taking the derivative of $\boldsymbol{K}$ and setting it to zero, it can be solved by Fantope projection [18]:

$$\boldsymbol{K} = \mathcal{P}_{\mathcal{F}^d}(\frac{1}{\mu}(\sum_{i=1}^{l} \boldsymbol{\alpha}_i(\boldsymbol{Z}_{(i)} + \boldsymbol{Z}_{(i)}^T) + \mu\boldsymbol{B} - \boldsymbol{Y}_3)), \tag{16}$$

where $\mathcal{P}_{\mathcal{F}^d}(\cdot)$ is the Euclidean projection on $\mathcal{F}^d$.

(5) Optimizing $\boldsymbol{B}$ with fixed irrelevant variables: This subproblem can be solved via soft thresholding operator $\mathcal{S}_\tau$:

$$\boldsymbol{B} = \mathcal{S}_{\frac{\gamma}{\mu}}(\boldsymbol{K} + \frac{\boldsymbol{Y}_3}{\mu}) \tag{17}$$

(6) Optimizing $\boldsymbol{\alpha}$ with fixed irrelevant variables: Let $\eta_i = tr((\boldsymbol{Z}_{(i)} + \boldsymbol{Z}_{(i)}^T)\boldsymbol{K})$, this subproblem is formulated as:

$$\max_{\boldsymbol{\alpha}} \sum_{i=1}^{l} \boldsymbol{\alpha}_i\eta_i, \quad s.t. \ \sum_{i=1}^{l} \boldsymbol{\alpha}^2 = 1, \boldsymbol{\alpha} \geq 0. \tag{18}$$

This subproblem can be solved with closed form as follows:

$$\boldsymbol{\alpha}_i = \frac{\eta_i}{\sqrt{(\sum_{i=1}^{l} \eta_i^2)}} \tag{19}$$

(7) Updating $\boldsymbol{\mathcal{Y}}_1$,$\boldsymbol{\mathcal{Y}}_2$,$\boldsymbol{\mathcal{Y}}_3$ by:

$$\begin{aligned} \boldsymbol{\mathcal{Y}}_1 &= \boldsymbol{\mathcal{Y}}_1 + \mu(\boldsymbol{\mathcal{T}} - \boldsymbol{\mathcal{K}} - \boldsymbol{\mathcal{E}}) \\ \boldsymbol{\mathcal{Y}}_2 &= \boldsymbol{\mathcal{Y}}_2 + \mu(\boldsymbol{\mathcal{Z}} - \boldsymbol{\mathcal{K}}) \\ \boldsymbol{Y}_3 &= \boldsymbol{Y}_3 + \mu(\boldsymbol{K} - \boldsymbol{Z}). \end{aligned} \tag{20}$$

We initialize $\mu = \min(\mu_{max}, \rho\mu)$ by a ratio of ρ until μ reach a pre-defined maximum value μ_{max}.

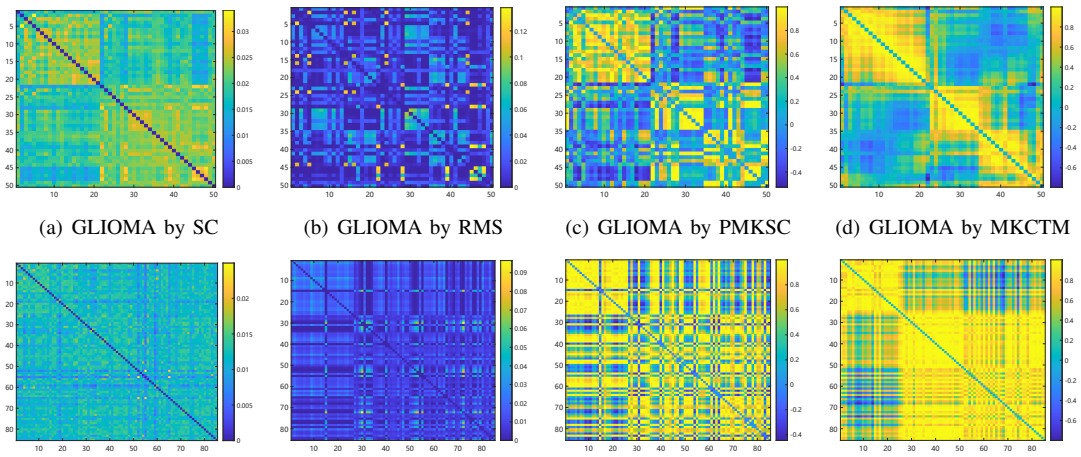
IV. EXPERIMENTS AND RESULTS

In this section, we evaluate the performance of our proposed method by comparing it with eight baseline methods on six real high-dimensional genomic datasets.

A. Datasets and Comparative Algorithms

We test our proposed MKCTM method on six high-dimensional genomic datasets[1] in the entire experiment. These six genomic datasets are GLOMA, Colon, GLI-85, Leukemia, CLL-SUB-111, and SMK-CAN-187. The statistical information on these six datasets is summarized in Tab. I. In the

[1]https://jundongl.github.io/scikit-feature/datasets.html



(a) GLIOMA by SC (b) GLIOMA by RMS (c) GLIOMA by PMKSC (d) GLIOMA by MKCTM

(e) GLI-85 by SC (f) GLI-85 by RMS (g) GLI-85 by PMKSC (h) GLI-85 by MKCTM

Fig. 1. Heatmaps of the affinity matrices obtained by the SC, RMS, PMKSM, and MKCTM approaches on GLIOMA and GLI-85 datasets. (a) GLIOMA by SC; (b) GLIOMA by RMS; (c) GLIOMA by PMKSC; (d) GLIOMA by MKCTM; (e) GLI-85 by SC; (f) GLI-85 by RMS; (g) GLI-85 by PMKSC; (h) GLI-85 by MKCTM.

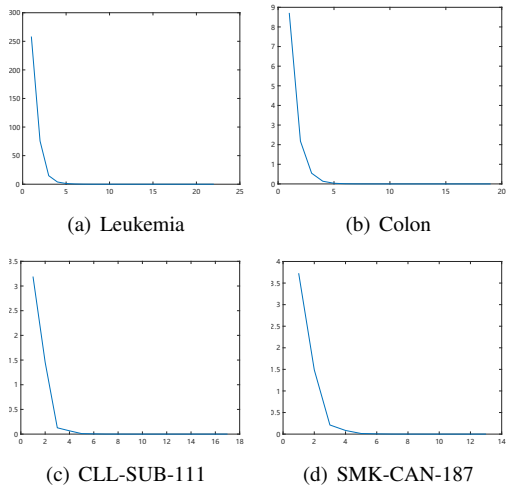
TABLE II
CLUSTERING PERFORMANCE ON SIX HIGH DIMENSIONAL GENE EXPRESSION DATASETES

Datasets	Methods	ACC	NMI	ARI	F-score	Datasets	Methods	ACC	NMI	ARI	F-score
GLIOMA	SC	0.5000	0.2834	0.1511	0.3666	GLI-85	SC	0.6941	0.1781	0.1412	0.5976
	LRR	0.4200	0.1646	0.0834	0.4200		LRR	0.6118	0.1685	0.0241	0.6118
	CAN	0.6000	0.4993	0.3278	0.5118		CAN	0.6588	0.0343	-0.035	0.6970
	RMS	0.5800	0.4796	0.3534	0.5191		RMS	0.6941	0.2391	0.1382	0.6022
	IPS2	0.4000	0.1607	0.0500	0.3091		IPS2	0.7294	0.2750	0.2012	0.6267
	On-ALK	0.7400	0.5242	0.4256	0.5677		On-ALK	0.7529	0.2387	0.2482	0.6465
	PMKSC	0.7600	0.5929	0.4926	0.6218		PMKSC	0.7412	0.2880	0.2241	0.6364
	SMKKM	0.5940	0.4676	0.3350	0.5027		SMKKM	0.7412	0.2524	0.2247	0.6355
	MKCTM	**0.9400**	**0.8695**	**0.8603**	**0.8951**		MKCTM	**0.9882**	**0.9114**	**0.9521**	**0.9792**
Colon	SC	0.5484	0.0062	-0.0058	0.5098	Leukemia	SC	0.6389	0.0694	0.0652	0.5473
	LRR	0.6613	0.0463	0.0263	0.6613		LRR	0.6667	0.0414	0.0243	0.6667
	CAN	0.5484	0.0008	-0.0094	0.5216		CAN	0.7361	0.1547	0.2112	0.6238
	RMS	0.3871	0.0007	-0.0206	0.4000		RMS	0.5556	0.1836	0.0965	0.4825
	IPS2	0.5484	0.0021	-0.0076	0.5158		IPS2	0.5556	0.0259	-0.0035	0.5242
	On-ALK	0.6613	0.0463	0.0263	0.6969		On-ALK	0.7361	0.1547	0.2112	0.6238
	PMKSC	0.6129	0.0322	0.0359	0.5323		PMKSC	0.7083	0.1368	0.1627	0.5951
	SMKKM	0.5323	0.0087	-0.0122	0.5110		SMKKM	0.5139	0.0001	-0.0124	0.5104
	MKCTM	**0.7903**	**0.2255**	**0.3221**	**0.6879**		MKCTM	**0.8750**	**0.4438**	**0.5551**	**0.7904**
CLL-SUB-111	SC	0.5315	0.1873	0.0873	0.4758	SMK-CAN-187	SC	0.5882	0.0221	0.0261	0.5246
	LRR	0.5315	0.2873	0.1537	0.5315		LRR	0.6364	0.0566	0.0694	0.6364
	CAN	0.5405	0.2665	0.1217	0.4885		CAN	0.5080	0.0080	-0.0023	0.6276
	RMS	0.3604	0.2152	0.0721	0.3344		RMS	0.5294	0.0794	0.0941	0.4439
	IPS2	0.5496	0.2676	0.1238	0.4902		IPS2	0.6257	0.0541	0.0583	0.5403
	On-ALK	0.5225	0.1325	0.07632	0.4341		On-ALK	0.6524	0.0727	0.0881	0.5454
	PMKSC	0.5496	0.2198	0.1139	0.4752		PMKSC	0.6578	0.0768	0.0947	0.5481
	SMKKM	0.5320	0.1892	0.1034	0.4657		SMKKM	0.6257	0.0480	0.0581	0.5288
	MKCTM	**0.6306**	**0.3153**	**0.1682**	**0.5713**		MKCTM	**0.7219**	**0.1490**	**0.1928**	**0.6000**

experiments, we compare our proposed MKCTM method with eight other methods, which include both classic baseline methods and recently reported new approaches for clustering. These methods can be divided into three categories: classical spectral clustering algorithms, including Spectral Clustering (SC) [4] and Low-rank representation (LRR) [22]; graph learning methods based on a single metric/kernel, including Clustering with adaptive neighbors (CAN) [8], Robust matrix factorization with spectral embedding (RMS) [23], and Integrating a tensor affinity and pairwise affinity (IPS2) [24]; multi-kernel methods, including Optimal neighborhood MKC with adaptive local kernels (ON-ALK) [12], Projective multiple kernel subspace clustering (PMKSC) [25], and Simple multiple kernel K-means(SMKKM) [26]. We implement the codes of BDR and RMS by MATLAB, while the codes of the others are taken from the authors' websites, and we adopt them directly.

B. Experimental Settings

Four well-established criteria, i.e., clustering accuracy (ACC), normalized mutual information (NMI), adjusted rand index (ARI), and F-score, are employed to evaluate the clustering performance. In the experiment, we applied three kernel functions to map the data into four different spaces. These kernel functions are the Gaussian kernel function $e^{\frac{-(\|x-y\|_2^2)}{2*\sigma^2}}$ with scale parameter σ, polynomial kernel $(x^Ty)^d$ with scale parameter $d = 1, 2$ and polynomial plus kernel $(x^Ty + 1)^d$ with scale parameter $d = 2$. For our proposed MKCTM, three balanced hyperparameters β, λ and γ are chosen from $[10^{-4}, \ldots, 10^0]$ by grid search.



(a) Leukemia (b) Colon

(c) CLL-SUB-111 (d) SMK-CAN-187

Fig. 2. Convergence results on the (a) Leukemia, (b) Colon, (c) CLL-SUB-111, and (d) SMK-CAN-187 datasets.

C. Experimental Results

Tab. IV presents the detailed clustering results of the proposed method and eight related methods on six gene expression datasets. The best results are highlighted in bold. To ensure fairness, the parameters of all compared algorithms were fine-tuned using grid search to obtain optimal performance. Additionally, for a fair comparison, all multi-kernel methods utilized the same kernel function set. Based on the experimental outcomes, the following observations hold: Our proposed MKCTM method outperforms the other comparative methods on all six gene expression datasets, achieving significant performance improvements. In terms of NMI, our method surpasses the baseline method by 57.91%, 73.33%, 21.93%, 37.44%, 12.8% and 12.69% on datasets GLIOMA, GLI-85, Colon, Leukemia, CLL-SUB-111, and SMK-CAN-187, respectively. Compared to the second-best method, our method surpasses by more than 27.66%, 62.34%, 17.92%, 26.02%, 4.77% and 7.22%. It also exhibits superior performance in ACC, ARI, and F-score metrics compared to other methods. These observations strongly demonstrate the superiority of our proposed model over existing methods. The remarkable performance of MKCTM can be attributed to two main factors: (1) Characterizing multiple kernel matrices through tensors and enabling tensor learning under low-rank constraints. This manner maximizes the correlation among different affinity matrices, eliminates noise and redundancy while preserving unique information within each space. (2) Weighted fusion of different kernels on the manifold space, considering the correlations and weights between different kernels from a higher-order perspective. This approach outperforms the traditional tensor fusion method of direct summation.

D. Visualization

In this section, we demonstrate the superiority of our method through visualization. Since the fused kernel matrix can be regarded as the affinity matrix used for clustering, this paper compares the affinity matrices obtained from the baseline method SC, graph-based method RMS, multi-kernel clustering PMKSC, and the proposed method on the GLIOMA and GLI-85 datasets through visualization as shown in Fig. 1. The first three columns are the heatmaps generated by the comparison methods, and the last column is the heatmaps generated by our method. A dark yellow color indicates higher sample affinity. It can be observed from Fig.1 that the heatmap obtained through traditional methods does not exhibit clear block structures between different clusters. Particularly on the GLI-85 dataset, the similarity between samples is completely lost due to the high-dimensional nature of the data. However, the MKCTM method is able to measure higher affinity among samples within the same cluster while having weak associations between different clusters. The heatmap clearly demonstrates distinct block structures, which unequivocally proves that our method can accurately reveal the relationships between samples and achieve superior clustering performance.

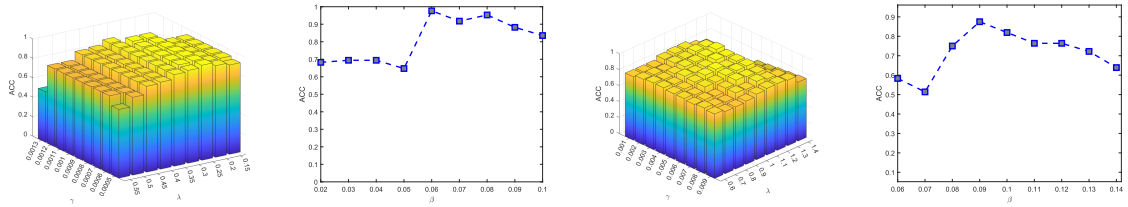
E. Convergence and Parameter Sensitivity

To validate the convergence of our proposed method, we calculate the total error at each iteration on the GLIOMA and GLI-85 datasets, as shown in Fig. 2. The total error is defined as the maximum value of the reconstruction error $\|\boldsymbol{\mathcal{T}} - \boldsymbol{\mathcal{K}} - \boldsymbol{\mathcal{E}}\|_F^2$ and the maximum change in each iteration, $\|\boldsymbol{\mathcal{K}}^{k+1} - \boldsymbol{\mathcal{K}}^k\|_F^2$, $\|\boldsymbol{\mathcal{E}}k+1-\boldsymbol{\mathcal{E}}^k\|_F^2$, and $\|\boldsymbol{K}^{k+1} - \boldsymbol{K}^k\|_F^2$: $Error = \max(\|\boldsymbol{\mathcal{T}} - \boldsymbol{\mathcal{K}} - \boldsymbol{\mathcal{E}}\|_F^2, \|\boldsymbol{\mathcal{K}}^{k+1} - \boldsymbol{\mathcal{K}}^k\|_F^2, \|\boldsymbol{\mathcal{E}}^{k+1} - \boldsymbol{\mathcal{E}}^k\|_F^2, \|\boldsymbol{K}^{k+1} - \boldsymbol{K}^k\|_F^2)$. From Fig. 2, it can be observed that the objective function value sharply decreases and quickly converges. On all datasets, the algorithm reaches a stable state within 25 iterations. Due to the closed-form solutions for each sub-problem in every iteration, our method is highly efficient.

It is worth noting that our proposed MKCTM method introduces three hyperparameters: β, λ, and γ. We divide these parameters into two groups: β and λ, γ, as they correspond to the subproblems of multi-kernel tensor learning and kernel tensor fusion, respectively. To observe the sensitivity of MKCTM to ACC, we fix β and vary λ, γ, as well as fix λ, γ and vary β, on the GLIOMA and Leukemia datasets as shown in Fig. 3. It can be observed from Fig. 3(b) and (d) that when β is too small, indicating insufficient sparse constraints on the separation noise term, a significant amount of useful information is lost, leading to a decrease in learning effectiveness. Conversely, when β is too large, the low-rank constraint on the tensor becomes excessively strong, resulting in the removal of only a small amount of noise during the learning process and thus reducing the clustering performance. Similar conclusions can be drawn from the analysis of the λ and γ parameters in Fig. 3(a) and (c).

V. CONCLUTION

In this paper, we propose a multi-kernel tensor clustering method that operates on the Grassmann manifold. To address the measurement error caused by the high-dimensional nature of genomic datasets, different kernel functions are employed to map the data into multiple RKHS and obtain a set of kernel matrices. By incorporating low-rank constraints in



(a) GLI-85 with fixed β (b) GLI-85 with fixed λ and γ (c) Leukemia with fixed β (d) Leukemia with fixed λ and γ

Fig. 3. The sensitivity of our proposed MKCTM method to fixed parameters β, fixed parameters λ and γ on GLI-85 and Leukemia datasets. (a) GLI-85 with fixed β; (b) GLI-85 with fixed λ and γ; (c) Leukemia with fixed β; (d) Leukemia with fixed λ and γ.

tensor representation, our proposed model achieves consensus information learning between kernel spaces from a high-order perspective to eliminate noise and redundant information. In contrast to traditional tensor fusion methods, we further propose to perform tensor fusion on the Grassmann manifold to generate an enhanced kernel matrix for clustering. In our experiments, we demonstrate the effectiveness of the proposed method on six high-dimensional gene expression datasets. Compared to existing mainstream methods, our approach consistently achieves superior results. Furthermore, through visualization, we provide evidence that the learned kernels accurately capture the relationships between high-dimensional data. In future work, we aim to enhance the computational efficiency and scalability of our method. For example, we plan to explore the construction of bipartite graphs to enable efficient clustering on large-scale high-dimensional datasets.

REFERENCES

[1] Y. Lu, Y.-M. Cheung, and Y. Y. Tang, "Self-adaptive multiprototype-based competitive learning approach: A k-means-type algorithm for imbalanced data clustering," *IEEE Transactions on Cybernetics*, vol. 51, no. 3, pp. 1598–1612, 2021.

[2] "Database resources of the national genomics data center, china national center for bioinformation in 2023," *Nucleic Acids Research*, vol. 51, no. D1, pp. D18–D28, 2023.

[3] T. Kanungo, D. M. Mount, N. S. Netanyahu, C. D. Piatko, R. Silverman, and A. Y. Wu, "An efficient k-means clustering algorithm: Analysis and implementation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 7, pp. 881–892, 2002.

[4] U. Von Luxburg, "A tutorial on spectral clustering," *Statistics and Computing*, vol. 17, no. 4, pp. 395–416, 2007.

[5] F. Liu, D. Choi, L. Xie, and K. Roeder, "Global spectral clustering in dynamic networks," *Proceedings of the National Academy of Sciences*, vol. 115, no. 5, pp. 927–932, 2018.

[6] Z. Chi, Z. Wang, B. Wang, Z. Fang, Z. Zhu, D. Li, and W. Du, "Multiple kernel subspace learning for clustering and classification," *IEEE Transactions on Knowledge and Data Engineering*, pp. 1–14, 2022.

[7] Y. Peng, W. Huang, W. Kong, F. Nie, and B.-L. Lu, "Jgsed: An end-to-end spectral clustering model for joint graph construction, spectral embedding and discretization," *IEEE Transactions on Emerging Topics in Computational Intelligence*, pp. 1–15, 2023.

[8] F. Nie, X. Wang, and H. Huang, "Clustering and projected clustering with adaptive neighbors," in *Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2014, pp. 977–986.

[9] C. Lu, J. Feng, Z. Lin, T. Mei, and S. Yan, "Subspace clustering by block diagonal representation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 41, no. 2, pp. 487–501, 2018.

[10] I. Dhillon, Y. Guan, and B. Kulis, "Kernel k-means, spectral clustering and normalized cuts," *KDD-2004 - Proceedings of the Tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, vol. 551-556, 07 2004.

[11] J. Liu, X. Liu, Y. Yang, X. Guo, M. Kloft, and L. He, "Multiview subspace clustering via co-training robust data representation," *IEEE Transactions on Neural Networks and Learning Systems*, 2021.

[12] J. Liu, X. Liu, J. Xiong, Q. Liao, S. Zhou, S. Wang, and Y. Yang, "Optimal neighborhood multiple kernel clustering with adaptive local kernels," *IEEE Transactions on Knowledge and Data Engineering*, vol. 34, no. 6, pp. 2872–2885, 2022.

[13] J. Lou and Y.-M. Cheung, "Robust low-rank tensor minimization via a new tensor spectral k -support norm," *IEEE Transactions on Image Processing*, vol. 29, pp. 2314–2327, 2020.

[14] L. Luo, J. Xu, C. Deng, and H. Huang, "Robust metric learning on grassmann manifolds with generalization guarantees," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 33, no. 01, 2019, pp. 4480–4487.

[15] J. Hamm and D. D. Lee, "Grassmann discriminant analysis: a unifying view on subspace-based learning," in *Proceedings of the 25th International Conference on Machine Learning*, 2008, pp. 376–383.

[16] X. Dong, P. Frossard, P. Vandergheynst, and N. Nefedov, "Clustering on multi-layer graphs via subspace analysis on grassmann manifolds," *IEEE Transactions on Signal Processing*, vol. 62, no. 4, pp. 905–918, 2014.

[17] C. Lu, J. Feng, Y. Chen, W. Liu, Z. Lin, and S. Yan, "Tensor robust principal component analysis with a new tensor nuclear norm," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 42, no. 4, pp. 925–938, 2020.

[18] V. Q. Vu, J. Cho, J. Lei, and K. Rohe, "Fantope projection and selection: A near-optimal convex relaxation of sparse pca," in *NIPS*, 2013.

[19] W. Hu, D. Tao, W. Zhang, Y. Xie, and Y. Yang, "The twist tensor nuclear norm for video completion," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 28, no. 12, pp. 2961–2973, 2016.

[20] 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, 2012.

[21] J. Malick, "A dual approach to semidefinite least-squares problems," *SIAM Journal on Matrix Analysis and Applications*, vol. 26, no. 1, pp. 272–284, 2004.

[22] X. Zhu, S. Zhang, Y. Li, J. Zhang, L. Yang, and Y. Fang, "Low-rank sparse subspace for spectral clustering," *IEEE Transactions on Knowledge and Data Engineering*, vol. 31, no. 8, pp. 1532–1543, 2019.

[23] M. Chen and X. Li, "Robust matrix factorization with spectral embedding," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 32, no. 12, pp. 5698–5707, 2020.

[24] H. Peng, Y. Hu, J. Chen, H. Wang, Y. Li, and H. Cai, "Integrating tensor similarity to enhance clustering performance," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 44, no. 5, pp. 2582–2593, 2022.

[25] M. Sun, S. Wang, P. Zhang, X. Liu, X. Guo, S. Zhou, and E. Zhu, "Projective multiple kernel subspace clustering," *IEEE Transactions on Multimedia*, vol. 24, pp. 2567–2579, 2021.

[26] X. Liu, "Simplemkkm: Simple multiple kernel k-means," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 45, no. 4, pp. 5174–5186, 2023.