Chinese Journal of Electronics
Vol.32, No.4, July 2023

Adaptive Tensor Rank Approximation for Multi-View Subspace Clustering

SUN Xiaoli[1], HAI Yang[1], ZHANG Xiujun[2], and XU Chen[1,3]

(1. *College of Mathematics and Statistics, Shenzhen University, Shenzhen 518060, China*)

(2. *School of Electronic and Communication Engineering, Shenzhen Polytechnic, Shenzhen 518055, China*)

(3. *National Center for Applied Mathematics Shenzhen (NCAMS), Shenzhen 518055, China*)

Abstract — Multi-view subspace clustering under a tensor framework remains a challenging problem, which can be potentially applied to image classification, impainting, denoising, etc. There are some existing tensor-based multi-view subspace clustering models mainly making use of the consistency in different views through tensor nuclear norm (TNN). The diversity which means the intrinsic difference in individual view is always ignored. In this paper, a new tensorial multi-view subspace clustering model is proposed, which jointly exploits both the consistency and diversity in each view. The view representation is decomposed into view-consistent part (low-rank part) and view-specific part (diverse part). A tensor adaptive log-determinant regularization (TALR) is imposed on the low-rank part to better relax the tensor multi-rank, and a view-specific sparsity regularization is applied on the diverse part to ensure connectedness property. Although the TALR minimization is not convex, it has a closed-form analytical solution and its convergency is validated mathematically. Extensive evaluations on six widely used clustering datasets are executed and our model is demonstrated to have the superior performance.

Key words — Diversity, Multi-view subspace clustering, Tensor adaptive log-determinant, View-specific sparsity, Self-representation.

I. Introduction

To make better use of data, researchers prefer to depict the practical data through heterogeneous views, which are gathered from various domains and created through different feature extractors. Multi-view subspace clustering aims at clustering a given dataset into groups through fully using the consistent and complementary information provided by different views. There are various applications, such as image clustering, motion segmentation, data representation, etc.

It is known that different views depict specific perspectives of data. Obviously, multiple features integration and the concealed structures extraction are the keys for deeply understanding the data. In general, multi-view clustering normally relies on the following two principles [1]–[3]. The first one is the consistency principle, which refers to maximize the commonality among different features. The second one is the complementary principle, which states that each feature contains distinctive information. Thus, we should depict data in a more concise and comprehensive manner through multi-view representations.

To fully explore the multi-dimensional data and better maintain the correlation between independent data, several multi-view clustering models under the tensorial framework have been presented [4]–[10] in recent years. Specially, inspired by the typical single view low-rank representation subspace clustering model (LRRSC) [11], tensor low rank constraint is naturally extended to depict the low rank structure among multi-view data. More recently, several representative and specific regularization schemes were also proposed to combine consensus with complementary information from different views [3], [12]. However, they simply applied a tensor low rank constraint on the self-representation matrices. Each view is so specific that it is not proper to directly set a uniform constraint on the specialized view data. In addition, the study of tensor rank is still very hot, and some convex solutions [5], [8] and non-convex solutions [1], [2], [13] have been presented in recent years. Although tensor nuclear norm [14] may lead to relatively lower convex envelop of tensor multi-rank function, it has strongly punished the large singu-

Manuscript Received June 27, 2022; Accepted Nov. 7, 2022. This work was supported by the National Natural Science Foundation of China (61872429, 62272313, 12101415, 62202018) and the Project of Educational Commission of Guangdong Province (2022KTSCX106).
 DOI:10.23919/cje.2022.00.180

lar values in Fourier domain, which may lead to the inaccurate approximation of tensor multi-rank in practical applications.

Based on the above observations, in this paper, we have given several suggestions to address these issues. Specifically, we first extract a component from each view to excavate the consistency among multi-views, which is implemented through a low tensorial rank to exploit inherent consistency for different views. Furthermore, to better approximate the rank function, we develop a novel tensor adaptive log-determinant regularization to minimize the tensor multi-rank. And the diverse part corresponding to the specific difference in each view is separated to guarantee the view-specific structures. A view-specific sparsity regularization is applied on the diverse part to ensure that the self-representation matrix is usually dense. Our contributions are summarized as follows:

1) To efficiently utilize both consistent and specific knowledge existing in multiple views, the view-consistent and view-specific components are both considered in our model. A tensor adaptive log-determinant regularizer is defined to make use of the consistency among different views. A view-specific sparsity regularization is proposed to enhance the diverse information among different feature spaces and keep the similar representations for the features in the same view.

2) We have deduced a closed-form solution for our non-convex model. Furthermore, the convergence of our algorithm has been proved mathematically, which means the convergence of this non-convex minimization problem is guaranteed.

3) Since the appropriate approximation function is adaptively obtained according to the intrinsic characteristics of each dataset, our clustering accuracy is greatly improved. We conduct a lot of experiments on six challenging clustering datasets. Compared with eleven latest convex and non-convex approaches, our proposed method has reached superior performance.

We introduce the following sections. We give a review of the related work in Section II, and present our model detailedly in Section III. All experimental results are presented in Section IV to demonstrate our superior performance. Finally, we reach the conclusion in Section V. In Appendix A and Appendix B, we give the proof of Theorem 1 and Theorem 2, respectively.

II. Related Work

Generally, the current multi-views subspace clustering (MSC) methods are briefly classified into three categories: 1) co-training approaches, 2) ensemble clustering, 3) subspace learning algorithms.

Co-training approaches try to find some shared representation by combining different views, that is, they tend to maximize the consistency between two different views [15]–[17]. Assuming that view data is a hybrid model, Bickel *et al.* [15] discovered that expectation-maximization method optimized the consistency between views. Kumar *et al.* [16] first developed a co-training algorithm to tackle MSC problem by utilizing the spectral of a view to control the adjacent matrices in the rest views. In [17], Yu *et al.* also performed the co-training subspace clustering using a Bayesian undirected graphical model.

The second type is called ensemble clustering [18], [19]. The core idea is to combine kernels in different views for obtaining the final grouping output in a linearly or nonlinearly way. Greene *et al.* [18] used a late integration strategy to combine information from related views. Tzortzis and Likas [19] expressed views through kernel matrices and proposed a kernel-based method for MSC problem. The weights of kernels indicated the quality of corresponding views.

In subspace learning algorithms, each view can be formulated and generated based on a latent subspace, which captures the common parts among different views. In [20], the authors achieved compatible clustering performance via a non-negative matrix factorization (NMF). In [21], Jiang *et al.* solved the low-rank matrix completion problem via nuclear norm minimization and Frobenius norm minimization function. Based on the two typical spectral-type single-view subspace clustering models, sparse subspace clustering (SSC) [22] and low-rank representation model (LRR) [11], some multi-view subspace clustering models have been proposed [3], [10], [23]–[25]. Xia *et al.* [23] proposed a Markov chain based multi-view spectral clustering (RMSC) through a sparse low-rank decomposition. Wang *et al.* [24] proposed the low-rank and sparse subspace clustering method. Cao *et al.* [3] maximized the diverse information among multi-views and explored the complementarity among multi-view data using the proposed method called Hilbert Schmidt independence criterion (HSIC). Wang *et al.* [25] adopted the exclusivity and consistency regularizers to capture the consistent and exclusive information. Wang *et al.* [10] proposed to use the intact space learning method for MSC problem. Zhao *et al.* [26] proposed the shape clustering method aiming at quickly clustering of arbitrary shapes and greatly reducing the storage space.

The above methods are all matrix-based methods for multi-view subspace clustering problem. Actually, tensor is more suitable to depict the high dimensional data of the real world. Motivated by this observation, some researchers attempted to find the inherent relationship among multi-views through high order tensor,

such as [4]–[8].

In [4], a tensorial framework was firstly used to ming the heterogeneous multi-view features. In [5], the feature tensor was obtained by stacking different self-representation matrices and constrained by a tensor nuclear norm to exploit the common parts among different features. In [6], Piao *et al.* further applied the multi-view subspace clustering which was based on sparse and low-rank to the clustering problem with multi-way data submodules. In [7], a t-product based tensor representation method was proposed to combine the multiple views. In [27], a new semi-supervised clustering method was proposed, which was based on Gaussian mixture models (GMM). In [28], a novel method was proposed to identify the clusters by topological graph partition (TGP). In [8], a new multi-view subspace clustering method used t-SVD [29] to depict the common information across multi-views, which was defined as follows.

Definition 1 (**t-SVD** [29]). Given $\mathcal{X} \in \mathbb{R}^{k_1 \times k_2 \times k_3}$, the t-SVD of $\mathcal{X} \in \mathbb{R}^{k_1 \times k_2 \times k_3}$ is described as

$$\mathcal{X} = \mathcal{U} * \mathcal{D} * \mathcal{V}^{\mathrm{T}} \tag{1}$$

where $\mathcal{U} \in \mathbb{R}^{k_1 \times k_1 \times k_3}$ and $\mathcal{V} \in \mathbb{R}^{k_2 \times k_2 \times k_3}$ are two orthogonal tensors. $\mathcal{D} \in \mathbb{R}^{k_1 \times k_2 \times k_3}$ is f-diagonal. "*" denotes the t-product [29].

In [12], hyper-Laplacian regularizer was further used to enhance the regional geometrical structure in different views. The above tensor-based approaches usually adopt tensor nuclear norm [14] (t-TNN, in short) to relax the tensor rank function, which is defined in Definition 2. As shown, in the t-TNN based models, the rank function is approximated by the L_1 norm of the singular values in Fourier domain. But due to its over-punishment to large singular values, the clustering accuracy is not satisfactory.

Definition 2 (**t-TNN** [14]). Given $\mathcal{X} \in \mathbb{R}^{k_1 \times k_2 \times k_3}$, the t-TNN of $\mathcal{X} \in \mathbb{R}^{k_1 \times k_2 \times k_3}$ is defined as

$$\begin{aligned} \|\mathcal{X}\|_* &= \frac{1}{k_3} \sum_{j=1}^{k_3} \mathrm{trace} \sqrt{\hat{X}^{(j)\mathrm{T}} \hat{X}^{(j)}} \\ &= \frac{1}{k_3} \sum_{i=1}^{\min(k_1,k_2)} \sum_{j=1}^{k_3} \left| \hat{\mathcal{D}}(i,i,j) \right| \end{aligned} \tag{2}$$

where $\hat{\mathcal{D}} = \mathrm{fft}(\mathcal{D}, [], 3)$, which is achieved through t-SVD operator of $\mathcal{X}$ in Fourier domain.

Recently, a few non-convex approximation methods of tensor rank function were proposed [2], [13], [30]. However, they always applied a uniform consensus regularized term to make use of the consistency among multi-views and ignored specialized subspace structures in each view. At the same time, since they were non-convex models, the existence of solutions and the convergence of algorithms have not been guaranteed in mathematics.

In this paper, a non-convex regularization is developed to approximate the tensor rank, and a new consensus and diversity multi-view subspace clustering approach equipped on a novel tensor framework is proposed to address above issues. A closed-form solution is mathematically deduced to tackle the non-convex optimization problem. Finally, we confirm the convergence of the algorithm in mathematics.

III. Our Model

1. Notations and preliminaries

Here, related definitions and preliminary notations are firstly provided. For simplicity, we summarize the related notations in Table 1.

Table 1. Notations summary

Notation	Definition	Notation	Definition
x	A vector	$x(i)$	The i-th entry of x
X	A matrix	$X(i,j)$	The (i,j)-th entry of X
$\mathcal{X}$	A tensor	$\mathcal{X}(i,j,k)$	The (i,j,k)-th entry of $\mathcal{X}$
$\mathcal{X}(k,:,:)$	The k-th horizontal slice of $\mathcal{X}$	$\hat{\mathcal{X}} \leftrightarrow \mathcal{X}$	$\hat{\mathcal{X}} = \mathrm{fft}(\mathcal{X}, [], 3)$
$\mathcal{X}(:,k,:)$	The k-th lateral slice of $\mathcal{X}$		$\mathcal{X} = \mathrm{ifft}(\hat{\mathcal{X}}, [], 3)$
$\mathcal{X}(:,:,k)$	The k-th frontal slice of $\mathcal{X}$	$\|\mathcal{X}\|_{2,1}$	$\|\mathcal{X}\|_{2,1} = \sum_{i,j} \|\mathcal{X}(i,j,:)\|_2$
$\mathcal{X}(:,i,j)$	mode-1 fibers of $\mathcal{X}$	$\|\mathcal{X}\|_{\mathrm{F}}$	$\|\mathcal{X}\|_{\mathrm{F}} = \sqrt{\sum_{i,j,k} \|\mathcal{X}(i,j,k)\|^2}$
$\mathcal{X}(i,:,j)$	mode-2 fibers of $\mathcal{X}$	$\|\mathcal{X}\|_{\infty}$	$\|\mathcal{X}\|_{\infty} = \max_{i,j,k} \|\mathcal{X}(i,j,k)\|$
$\mathcal{X}(i,j,:)$	mode-3 fibers of $\mathcal{X}$	$\|\mathcal{X}\|_*$	Tensor nuclear norm
$\mathcal{X}^{(k)}$	$\mathcal{X}^{(k)} = \mathcal{X}(:,:,k)$	$\|\mathcal{X}\|_{\mathrm{TALR}}$	Tensor adaptive log-determinant regularizer

2. Tensor adaptive log-determinant regularization on the consensus part

In the traditional tensor based methods, the same approximation of rank function was used during the whole iteration without considering the magnitude of singular values. Thus, different singular values were

punished equally. In contrast, in this paper, we propose a tensor adaptive log-determinant regularizer (TALR, in short), as the following Definition 3. The TALR is imposed on the low rank part, and it can automatically select the appropriate approximation function according to the magnitude of singular values, which leads to the higher performance.

Definition 3 (Tensor adaptive log-determinant regularization (TALR)). Given $\mathcal{X}\in\mathbb{R}^{k_1\times k_2\times k_3}$, the tensor adaptive log-determinant regularization (TALR) is defined as

$$\begin{aligned}\|\mathcal{X}\|_{\mathrm{TALR}} &= \frac{1}{k_3}\sum_{j=1}^{k_3}\log\det(\gamma\sqrt{\hat{X}^{(j)\mathrm{T}}\hat{X}^{(j)}}+I)\\ &= \frac{1}{k_3}\sum_{i=1}^{\min(k_1,k_2)}\sum_{j=1}^{k_3}\log(\gamma\hat{\mathcal{D}}(i,i,j)+1)\end{aligned}\tag{3}$$

where $\hat{\mathcal{D}}$ can be obtained through t-SVD of $\mathcal{X}$ in Fourier domain. $\gamma>0$ is the adaptive shrinkage factor, ensuring regularization stability.

As shown in Fig.1(a), blue line represents rank function, and red line represents the t-TNN based surrogate function. Yellow line represents our proposed adaptive approximation function adjusted through γ. We can clearly notice that a tighter surrogate function will approximate rank function better, so γ could be determined by minimizing the shaded area $S(\gamma)$, which is calculated as follows:

$$\begin{aligned}&S(\gamma)\\ &=\int_0^{\frac{e-1}{\gamma}}[1-\log(\gamma|x|+1)]\mathrm{d}x+\int_{\frac{e-1}{\gamma}}^{M}[\log(\gamma|x|+1)-1]\mathrm{d}x\\ &=M\log(\gamma M+1)-M+\int_0^{\frac{e-1}{\gamma}}\left(1-\frac{1}{\gamma x+1}\right)\mathrm{d}x\\ &\quad-\int_{\frac{e-1}{\gamma}}^{M}\left(1-\frac{1}{\gamma x+1}\right)\mathrm{d}x\end{aligned}\tag{4}$$

where $M=\sigma_{f,k-1}^{(n3)}(i,i)$ and $\sigma_{f,k-1}^{(n3)}(i,i)$ is obtained by t-SVD of $\mathcal{H}_f$ in the $k-1$ step. Let $S'(\gamma)=0$, then we could get the proper value of γ, which is related to the value of M. Different values of γ represent different approximation function curves, just as shown in Fig.1(b). According to singular values of the $k-1$ iteration, we obtain the best approximation functions in the k iteration. Because the surrogate functions are adaptively adjusted, it greatly improves the clustering accuracy.

3. View-specific sparsity regularization on the diversity part

Ideally, different views depict distinctive information of data. Thus, the diversity information existing in each view is completely different. That means the relationships among different views are sparse, encoding

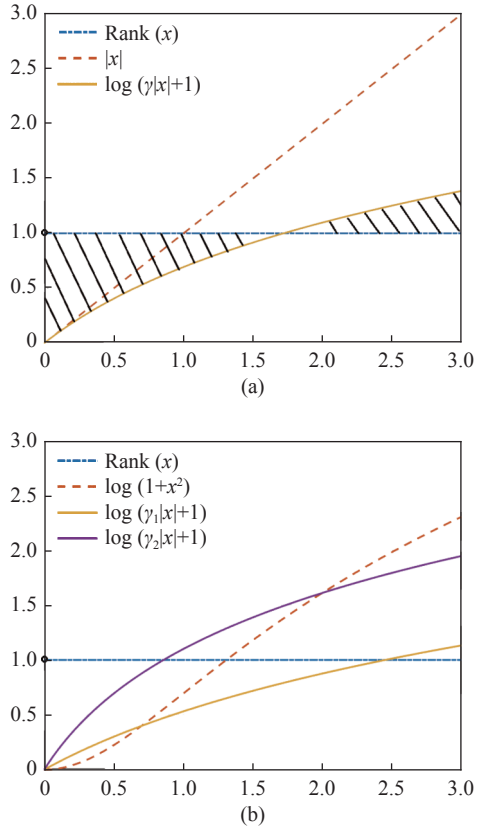


Fig. 1. The general framework and detailed flowchart based on our proposed model.

with l_1 norm. And in the same view, the features should share the similar representations, encoding with $\|\cdot\|_{\mathrm{F}}$ norm. Then we propose the view-specific sparsity regularization as the following definition.

Definition 4 (View-specific sparsity regularization (VSSR)). $D^{(v)}$ represents the diverse matrix of v-th view, and the view-specific sparsity regularization (VSSR) is defined as

$$\|\mathcal{D}\|_{\mathrm{VSSR}}=\sum_{v=1}^{V}\left\|D^{(v)}\right\|_{\mathrm{F}}^2,(v=1,2,\ldots,V)\tag{5}$$

As shown in Definition 4, the view-specific sparsity regularization has enhanced the diversity among different views as much as possible. Meanwhile, it keeps the similar representation for the features in the same view and ensures the dense connectedness property.

4. Our model

Let $X^{(1)},X^{(2)},\ldots,X^{(V)}$ denote V views of clustering data. $Z^{(1)},Z^{(2)},\ldots,Z^{(V)}$ denote the corresponding view-common matrices and $D^{(1)},D^{(2)},\ldots,D^{(V)}$ are the corresponding view-diverse matrices. For making use of the common structures, the view-common matrices are firstly stacked into a 3-order tensor. Then, an $N\times V\times N$ tensor is obtained by rotation, which is to reduce the calculation complexity [8]. In Fourier domain, instead of

the tensor nuclear norm [14], we adopt the tensor adaptive log-determinant regularization defined in Definition 3 on the rotated coefficient tensor $\mathcal{Z}$. At the same time, the view-specific sparsity regularization is applied on the diverse part to ensure connectedness property of the representation matrices. The Lagrangian multipliers are augmented and applied to iteratively optimize the related equations. Here, we can obtain the optimal tensor $\mathcal{Z}^*$ and the matrices $\{D^{(v)*}\}_{v=1}^{V}$ when iterations stop. The final $\{Z^{(v)*}\}_{v=1}^{V}$ can be developed through unfolding $\mathcal{Z}^*$. The similarity matrix generated from $\{Z^{(v)*}\}_{v=1}^{V}$ and $\{D^{(v)*}\}_{v=1}^{V}$ can be used to obtain the final clustering result. The whole construction of our approach is illustrated in Fig.2.

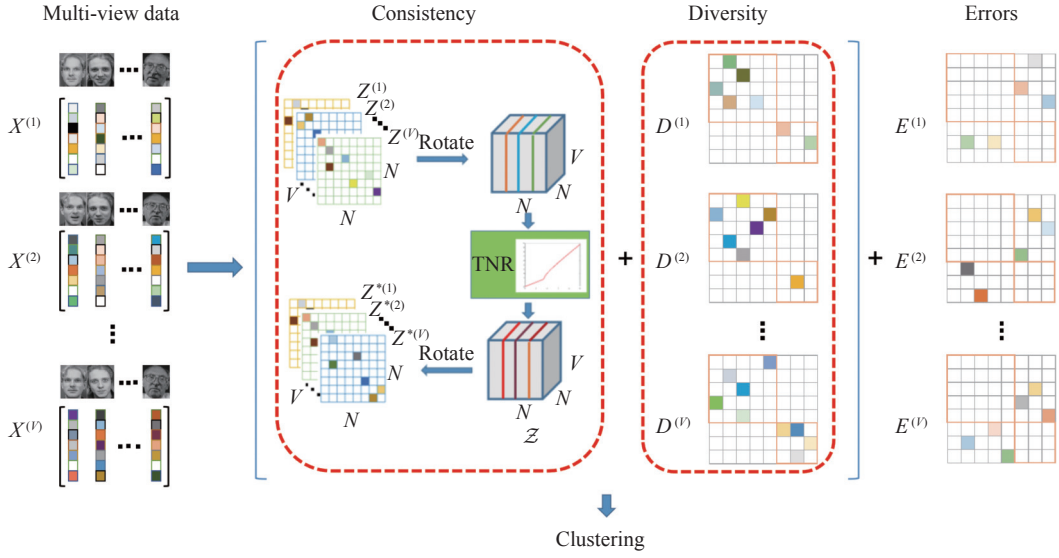


Fig. 2. Different approximation functions of the rank function. (a) The proper method of selecting parameter γ; (b) Comparisons among different log-determinant curves and $\gamma_1 = 0.7$, $\gamma_2 = 2$.

After getting the low rank tensor $\mathcal{Z}$ and the connected diverse part $\{D^{(v)*}\}_{v=1}^{V}$, we formulate our model as follows:

$$\begin{aligned}&\min_{Z^{(v)},D^{(v)},E^{(v)}} \underbrace{\|\mathcal{Z}\|_{\mathrm{TALR}}}_{\text{Consensus Part}} + \underbrace{\lambda\|E\|_{2,1}}_{\text{Error Part}} + \underbrace{\beta\sum_{v=1}^{V}\left\|D^{(v)}\right\|_{\mathrm{F}}^{2}}_{\text{Diversity Part}}\\ &\text{s.t.}\quad X^{(v)} = X^{(v)}(Z^{(v)}+D^{(v)})+E^{(v)}, v=1,2,\ldots,V\\ &\qquad E=\left[E^{(1)};E^{(2)};\ldots;E^{(V)}\right]\\ &\qquad \mathcal{Z}=\Phi(Z^{(1)},Z^{(2)},\ldots,Z^{(V)})\end{aligned} \tag{6}$$

where $\Phi(\cdot)$ represents the merging and rotating operation defined in [8], and $\{E^{(v)}\}_{v=1}^{V}$ denote error matrices. Formula (6) aims at finding the optimal view-consistent part $Z^{(1)}, Z^{(2)}, \ldots, Z^{(V)}$ and view-diverse part $D^{(1)}, D^{(2)}, \ldots, D^{(V)}$. Consequently, the similarity matrix is obtained by

$$S=\sum_{v=1}^{V}\left(\frac{\left|Z^{(v)}\right|+\left|Z^{(v)^{\mathrm{T}}}\right|}{2}+\frac{\left|D^{(v)}\right|+\left|D^{(v)^{\mathrm{T}}}\right|}{2}\right) \tag{7}$$

5. Optimization approach

To efficiently separate variables, we introduce a variable $\mathcal{H}$ and let $\mathcal{H}=\mathcal{Z}$. Thus, the augmented Lagrange function of (6) is

$$\begin{aligned}&\mathcal{L}\left(Z^{(1)},\ldots,Z^{(V)};E^{(1)},\ldots,E^{(V)};D^{(1)},\ldots,D^{(V)};\mathcal{H}\right)\\ &=\|\mathcal{H}\|_{\mathrm{TALR}}+\lambda\|E\|_{2,1}+\beta\sum_{v=1}^{V}\left\|D^{(v)}\right\|_{\mathrm{F}}^{2}\\ &\quad+\sum_{v=1}^{V}\Big(\langle Y^{(v)},X^{(v)}-X^{(v)}(Z^{(v)}+D^{(v)})-E^{(v)}\rangle\\ &\quad+\langle\mathcal{W},\mathcal{Z}-\mathcal{H}\rangle+\frac{\mu}{2}\left\|X^{(v)}-X^{(v)}(Z^{(v)}+D^{(v)})-E^{(v)}\right\|_{\mathrm{F}}^{2}\Big)\\ &\quad+\frac{\rho}{2}\|\mathcal{Z}-\mathcal{H}\|_{\mathrm{F}}^{2}\end{aligned} \tag{8}$$

where the tensor $\mathcal{W}$ and the matrix $Y^{(v)}$ are both Lagrangian multipliers. ρ and μ are penalty parameters. We can solve the optimal problem (5) through alternately updating the variables while keeping others fixed. We will depict the detailed procedure in the following.

Subproblem of $Z^{(v)}$ Fixing E, D and $\mathcal{H}$, and let $\Phi_{(v)}^{-1}(\mathcal{W})=W^{(v)}$, $\Phi_{(v)}^{-1}(\mathcal{H})=H^{(v)}$, the subproblem of $Z^{(v)}$ is

$$\begin{aligned}&\min_{Z^{(v)}}\langle Y^{(v)},X^{(v)}-X^{(v)}(Z^{(v)}+D^{(v)})-E^{(v)}\rangle\\ &\quad+\frac{\mu}{2}\|X^{(v)}-X^{(v)}(Z^{(v)}+D^{(v)})\\ &\quad-E^{(v)}\Big\|_{\mathrm{F}}^{2}+\langle W^{(v)},Z^{(v)}-H^{(v)}\rangle+\frac{\rho}{2}\Big\|Z^{(v)}-H^{(v)}\|_{\mathrm{F}}^{2}\end{aligned} \tag{9}$$

The first-order optimal condition of (9) is

$$-X^{(v)\mathrm{T}}Y^{(v)}-\mu X^{(v)\mathrm{T}}\left(X^{(v)}-X^{(v)}(Z^{(v)}+D^{(v)})-E^{(v)}\right)+W^{(v)}+\rho\left(Z^{(v)}-H^{(v)}\right)=0 \tag{10}$$

which leads to

$$Z^{(v)*}=\left(I+\frac{\mu}{\rho}X^{(v)^{\mathrm{T}}}X^{(v)}\right)^{-1}\left(\left(X^{(v)^{\mathrm{T}}}Y^{(v)}+\mu X^{(v)^{\mathrm{T}}}X^{(v)}-\mu X^{(v)^{\mathrm{T}}}X^{(v)}D^{(v)}-\mu X^{(v)^{\mathrm{T}}}E^{(v)}-W^{(v)}\right)/\rho+H^{(v)}\right) \tag{11}$$

Subproblem of $E^{(v)}$: Fixing Z, D and $\mathcal{H}$, subproblem of error matrix E is

$$\begin{aligned}E^*&=\arg\min_E \lambda\|E\|_{2,1}+\sum_{v=1}^{V}\left(\langle Y^{(v)},-E^{(v)}\rangle+\frac{\mu}{2}\left\|X^{(v)}-X^{(v)}(Z^{(v)}+D^{(v)})-E^{(v)}\right\|_{\mathrm{F}}^2\right)\\&=\arg\min_E \frac{\lambda}{\mu}\|E\|_{2,1}+\frac{1}{2}\|E-F\|_{\mathrm{F}}^2\end{aligned} \tag{12}$$

Here, F is obtained through vertically stacking the matrix $X^{(v)}-X^{(v)}(Z^{(v)}+D^{(v)})+\frac{1}{\mu}Y^{(v)}$ along the column [11].

Subproblem of $D^{(v)}$: Fixing E, Z and $\mathcal{H}$, the subproblem of diverse matrix $D^{(v)}$ is

$$\min_{D^{(v)}}\beta\left\|D^{(v)}\right\|_{\mathrm{F}}^2+\langle Y^{(v)},X^{(v)}-X^{(v)}(Z^{(v)}+D^{(v)})-E^{(v)}\rangle+\frac{\mu}{2}\|X^{(v)}-X^{(v)}(Z^{(v)}+D^{(v)})-E^{(v)}\|_{\mathrm{F}}^2 \tag{13}$$

The first-order optimal condition of (12) is

$$2\beta D^{(v)}-X^{(v)\mathrm{T}}Y^{(v)}-\mu X^{(v)\mathrm{T}}\left(X^{(v)}-X^{(v)}(Z^{(v)}+D^{(v)})-E^{(v)}\right)=0 \tag{14}$$

leading to

$$D^{(v)*}=(2\beta I+\mu X^{(v)^{\mathrm{T}}}X^{(v)})^{-1}(X^{(v)^{\mathrm{T}}}Y^{(v)}+\mu X^{(v)^{\mathrm{T}}}X^{(v)}-\mu X^{(v)^{\mathrm{T}}}X^{(v)}Z^{(v)}-\mu X^{(v)^{\mathrm{T}}}E^{(v)}) \tag{15}$$

Subproblem of $\mathcal{H}$: Fixing Z, D and E, the subproblem of $\mathcal{H}$ is

$$\begin{aligned}\mathcal{H}^*&=\operatorname*{argmin}_{\mathbf{H}}\|\mathcal{H}\|_{\mathrm{TALR}}+\langle\mathcal{W},\mathcal{Z}-\mathcal{H}\rangle+\frac{\rho}{2}\|\mathcal{Z}-\mathcal{H}\|_{\mathrm{F}}^2\\&=\operatorname*{argmin}_{\mathbf{H}}\|\mathcal{H}\|_{\mathrm{TALR}}+\frac{1}{2}\|\mathcal{H}-\mathcal{M}\|_{\mathrm{F}}^2\end{aligned} \tag{16}$$

Here $\tau=\frac{1}{\rho}$, and $\mathcal{M}=\mathcal{Z}+\frac{1}{\rho}\mathcal{W}$. Then equation (15) can be solved through Theorem 1. We prove Theorem 1 in Appendix A. Algorithm 1 summarizes the updating process of $\mathcal{H}$.

Theorem 1 Suppose $\tau>0$, $\mathcal{X},\mathcal{Y}\in\mathbb{R}^{k_1\times k_2\times k_3}$, and $\mathcal{Y}$ has the t-SVD: $\mathcal{Y}=\mathcal{U}*\mathcal{D}_y*\mathcal{V}^{\mathrm{T}}$ as defined in Definition 2 in Section II. If $\mathcal{X}^*$ is the optimal solution of

$$\operatorname*{argmin}_{\mathcal{X}}\tau\|\mathcal{X}\|_{\mathrm{TALR}}+\frac{1}{2}\|\mathcal{X}-\mathcal{Y}\|_{\mathrm{F}}^2 \tag{17}$$

then $\mathcal{X}^*=\mathcal{U}*\mathcal{D}_x*\mathcal{V}^{\mathrm{T}}$, where $\mathcal{D}_x\in\mathbb{R}^{k_1\times k_2\times k_3}$ is f-diagonal. For $1\le i\le\min(k_1,k_2),1\le j\le k_3$, let $\hat{\mathcal{D}}_y=\mathrm{fft}(\mathcal{D}_y,[],3)$ and $\hat{\mathcal{D}}_x=\mathrm{fft}(\mathcal{D}_x,[],3)$. Then, the diagonal elements of $\hat{\mathcal{D}}_x$ is able to be achieved through solving the equation: $(\hat{\mathcal{D}}_x(i,i,j))^2+(\frac{1}{\gamma}-\hat{\mathcal{D}}_y(i,i,j))\hat{\mathcal{D}}_x(i,i,j)+\frac{1}{\gamma}(\tau-\hat{\mathcal{D}}_y(i,i,j))=0$. That is

$$\hat{\mathcal{D}}_x(i,i,j)=\begin{cases}\frac{\left(\hat{\mathcal{D}}_y(i,i,j)-\frac{1}{\gamma}\right)+\sqrt{\Delta}}{2}, & \Delta\ge 0\\ 0, & \Delta<0\end{cases} \tag{18}$$

where $\Delta=\left(\frac{1}{\gamma}-\hat{\mathcal{D}}_y(i,i,j)\right)^2-\frac{4}{\gamma}\left(\tau-\hat{\mathcal{D}}_y(i,i,j)\right)$.

Algorithm 1 Update $\mathcal{H}$ via Theorem 1

Input: $\mathcal{Z}$, $\mathcal{W}$, ρ and γ.

1: $\tau=\frac{1}{\rho},\mathcal{M}=\mathcal{Z}+\frac{1}{\rho}\mathcal{W}$;

2: $\hat{\mathcal{M}}=\mathrm{fft}(\mathcal{M},[],3)$;

3: **for** $j=1:k_3$ **do**

$\left[\hat{\mathcal{U}}^{(j)},\hat{\mathcal{D}_m}^{(j)},\hat{\mathcal{V}}^{(j)}\right]=\mathrm{SVD}\left(\hat{\mathcal{M}}^{(j)}\right)$;

obtain $\hat{\mathcal{D}}_h(i,i,j)$ through Theorem 1;

$\hat{\mathcal{H}}^{(j)}=\hat{\mathcal{U}}^{(j)}\,\hat{\mathcal{D}_h}^{(j)}\,\hat{\mathcal{V}}^{(j)^{\mathrm{T}}}$;

4: **end for**

5: $\mathcal{H}=\mathrm{ifft}(\hat{\mathcal{H}},[],3)$;

Output: Tensor $\mathcal{H}$.

Update $Y^{(v)}$:

$$Y^{(v)*}=Y^{(v)}+\mu\left(X^{(v)}-X^{(v)}(Z^{(v)}+D^{(v)})-E^{(v)}\right) \tag{19}$$

Update $\mathcal{W}$:

$$\mathcal{W}^*=\mathcal{W}+\rho(\mathcal{Z}-\mathcal{H}) \tag{20}$$

So far, we can summarize our algorithm in Algorithm 2. Besides, we also demonstrate its convergence as in Theorem 2, which has been proven in Appendix B.

Theorem 2 Let $\{Z^{(v)^t},E^{(v)^t},D^{(v)^t},\mathcal{H}^t\}$ be the sequence generated by Algorithm 2, and the superscript t denotes the iterations. Suppose that the sequence $\{Y^{(v)^t},\mathcal{W}^t\}$ is bounded, μ^t and ρ^t are non-de-

creasing, and $\sum_{t=0}^{\infty}\frac{\mu^{t+1}}{(\mu^t)^2}<\infty, \sum_{t=0}^{\infty}\frac{\rho^{t+1}}{(\rho^t)^2}<\infty$, then, 1) The sequence $\{Z^{(v)^t}, E^{(v)^t}, D^{(v)^t}, \mathcal{H}^t\}$ is bounded. 2) The sequence $\{Z^{(v)^t}, E^{(v)^t}, D^{(v)^t}, \mathcal{H}^t\}$ has at least one accumulation point, and any accumulation point is a stationary KKT point for the optimization issue of (7).

Algorithm 2 Diversity induced MSC method via TALR

Input: V views of clustering data: $X^{(1)}, X^{(2)}, \ldots, X^{(V)}$, λ, β, and number of clusters: K.

Initialized $Z^{(v)}, D^{(v)}, E^{(v)}, Y^{(v)}, \mathcal{H}, \mathcal{W}$ to zero tensors; $v=1,2,\ldots,V; \mu=10^{-4}, \rho=10^{-4}, \eta=1.2, \mu_{\max}=\rho_{\max}=10^{12}, \varepsilon=10^{-5}$;

Do

  for $v=1:V$

    Calculate $Z^{(v)}$ by solving (10);

  end for

  Calculate E by using (11);

  for $v=1:V$

    Calculate $D^{(v)}$ by solving (14);

  end for

  for $v=1:V$

    Calculate $Y^{(v)}$ by solving (18);

  end for

  Calculate $\mathcal{Z}=\Phi(Z^{(1)}, Z^{(2)}, \ldots, Z^{(V)})$;

  Calculate $\mathcal{H}$ by Algorithm 1;

  Calculate $\mathcal{W}$ by solving (20);

  Calculate ρ, μ: $\rho=\min(\eta\rho, \rho_{\max})$, $\mu=\min(\eta\mu, \mu_{\max})$;

  $(H^{(1)}, H^{(2)}, \ldots, H^{(V)})=\Phi^{-1}(\mathcal{H})$;

  Check if the convergence terms are satisfied:

$$\|X^{(v)}-X^{(v)}(Z^{(v)}+D^{(v)})-E^{(v)}\|_{\infty}<\varepsilon \text{ and } \|Z^{(v)}-H^{(v)}\|_{\infty}<\varepsilon;$$

Until converge

Obtain the similarity matrix using (6);

Employ spectral clustering through calculatd affinity matrix S;

Output: The clustering results.

IV. Experiments

1. Experimental conditions

In this section, we performed a lot of experiments upon 6 widely used datasets and compared our approach with 11 representative clustering models. All of our experiments were executed using Matlab R2020 on a computer equipped with i7-1070 CPU and 64 GB memory. For all experiments, we have run 10 times and recorded the average performances. Our code could be download from *https://github.com/mathxlsun/TALR*.

1) Datasets introduction

We adopted 6 commonly used image datasets to evaluate our proposed method. We also applied our method to 3 typical clustering scenes, including scene clustering, face classification, and generic objects recognition. The 6 datasets have been detailedly described in Table 2. As shown, for face clustering problem, two datasets were selected, including: Yale dataset and Extended YaleB dataset. For scene classification, two datasets were conducted, including: Scene-15 dataset and MITIndoor-67 dataset [31]. And for generic clustering problem, two datasets were conducted, including COIL-20 dataset and Caltech-101 dataset [32].

2) View descriptions

Similar to references [6] and [12], for two face clustering datasets and COIL-20 dataset, three features were used as "Views", including LBP feature [33], image intensity, and Gabor feature [34].

On the other hand, we also extracted other three features for MITIndoor-67, Caltech-101, and Scene-15 datasets, including CENTRIST [35], PRI-CoLBP [36] and PHOW [37], which represented the census transform histogram, pairwise rotation invariant co-occurrence local binary pattern feature, and pyramid histograms of visual words, respectively.

Table 2. Statistics of datasets and multiview features in the experiments.

Datasets	Objective	Clusters	Images	Views/Features
Yale	Face	15	165	(Intensity, LBP, Gabor)
ExtendedYaleB	Face	10	650	(Intensity, LBP, Gabor)
Scene-15	Scene	15	4485	(PHOW, PRI-CoLBP, CENTRIST)
MITIndoor-67	Scene	67	5360	(PHOW, PRI-CoLBP, CENTRIST, VGG19)
Coil-20	Generic object	20	1440	(Intensity, LBP, Gabor)
Caltech-101	Generic object	101	8677	(PHOW, PRI-CoLBP, CENTRIST, InceptionV3)

To complement the handcrafted features and improve the performance, two other deep features (e.g., VGG19 [38] for MITIndoor-67 dataset and InceptionV3 [39] for Caltech-101 dataset) were further imported as new "Views".

All features utilized in the experiments are listed in Table 2. More detailed operations for features extraction can be found in [8] and [12].

3) Evaluation metrics and compared methods

In our experiments, following [40], we also used 6 metrics to measure the performance, including accuracy (ACC), F-score, normalized mutual information (NMI), adjusted rank index (AR), recall and precision. Obviously, methods with better clustering performance

usually have the higher value in all above six metrics.

In this subsection, we compared our approach with 11 latest clustering methods, including SPC_{best} [41], LRR_{best} [11], RMSC [23], DiMSC [3], LTMSC [5], ECMSC [25], t-SVD-MSC [8], HLR-M^2VS [12], S_p-MSC [13], GLMSC [2], and TLRMSC [30]. Among these eleven methods, the first eight methods are convex, and the last three methods are non-convex.

2. Comparison with other methods

In this subsection, we have shown some compared results on six popular clustering datasets in Table 3, Table 4, Table 5, Table 6, Table 7, and Table 8. In the tables, if the results of one method were labeled as "–" on one dataset, it means that the method didn't publish their results on this dataset. To ensure a fair comparison, the best results for all competitors were selected or reported in [3], [5], [8], [12].

Table 3. Experimental results on Yale. Here, $\lambda = 0.092$ and $\beta = 3.3$ for our model

Methods	ACC	F-score	NMI	AR	Recall	Precision
SPC_{best}	0.619	0.456	0.653	0.441	0.501	0.456
LRR_{best}	0.697	0.547	0.709	0.512	0.567	0.529
RMSC	0.642	0.517	0.684	0.485	0.535	0.500
DiMSC	0.709	0.564	0.727	0.535	0.586	0.543
LTMSC	0.741	0.598	0.765	0.570	0.629	0.569
ECMSC	–	–	–	–	–	–
t-SVD-MSC	0.963	0.915	0.953	0.910	0.927	0.904
HLR-M2VS	–	–	–	–	–	–
Ours	**0.983**	**0.975**	**0.987**	**0.973**	**0.980**	**0.970**

For face clustering problem, results for Yale dataset and Extended YaleB dataset are provided in Table 3 and Table 4, respectively. As shown, on Yale dataset and Extended YaleB dataset, our methods both achieve the best performance. Especially, on Extended YaleB dataset, our approach improves the superior performance (reported in [12]) from NMI 0.703 and ACC 0.670 to NMI 0.960 and ACC 0.975.

For scene clustering problem, experimental results for Scene-15 dataset and MITIndoor-67 dataset are illustrated as in Table 5 and Table 6, respectively. On Scene-15 dataset, our method has gained significant improvements of 7.9%, 10.8%, 12.3%, 11.4%, 12.4%, and 10.5%, in terms of six metrics, respectively. On the MITIndoor-67 dataset, our method also significantly outperforms other competitors with a relatively clear margin, about 12.3%, 15.3%, 22.8%, 22.5%, 22.1% and 22.2% in terms of six metrics, respectively.

Table 4. Experimental results on Extended YaleB. Here, $\lambda = 0.0022$ and $\beta = 0.38$ for our model

Methods	ACC	F-score	NMI	AR	Recall	Precision
SPC_{best}	0.367	0.308	0.361	0.226	0.311	0.297
LRR_{best}	0.615	0.508	0.627	0.451	0.539	0.481
RMSC	0.210	0.155	0.157	0.060	0.159	0.150
DiMSC	0.615	0.504	0.636	0.453	0.534	0.481
LTMSC	0.626	0.521	0.637	0.459	0.539	0.485
ECMSC	0.783	0.597	0.759	0.544	0.718	0.513
HLR-M2VS	0.670	0.577	0.703	0.529	0.595	0.560
t-SVD-MSC	0.652	0.550	0.667	0.500	0.590	0.514
Ours	**0.975**	**0.952**	**0.960**	**0.947**	**0.954**	**0.950**

Table 5. Experimental results on Scene-15. Here, $\lambda = 0.008$ and $\beta = 0.005$ for our model

Methods	ACC	F-score	NMI	AR	Recall	Precision
SPC_{best}	0.437	0.322	0.422	0.271	0.328	0.315
LRR_{best}	0.445	0.324	0.426	0.272	0.333	0.316
RMSC	0.507	0.437	0.564	0.394	0.450	0.425
DiMSC	0.300	0.181	0.269	0.117	0.190	0.173
LTMSC	0.574	0.465	0.571	0.424	0.479	0.452
ECMSC	0.457	0.357	0.463	0.303	0.408	0.318
HLR-M2VS	0.878	0.861	0.895	0.850	0.871	0.850
t-SVD-MSC	0.812	0.788	0.858	0.771	0.839	0.743
Ours	**0.986**	**0.975**	**0.974**	**0.973**	**0.976**	**0.974**

Table 6. Experimental results on MITIndoor-67. Here, $\lambda = 0.002$ and $\beta = 0.003$ for our model

Methods	ACC	F-score	NMI	AR	Recall	Precision
SPC_{best}	0.445	0.314	0.558	0.306	0.341	0.295
LRR_{best}	0.120	0.045	0.226	0.031	0.047	0.044
RMSC	0.232	0.123	0.342	0.110	0.125	0.121
DiMSC	0.246	0.141	0.383	0.128	0.144	0.138
LTMSC	0.431	0.290	0.546	0.280	0.306	0.279
ECMSC	0.353	0.228	0.489	0.216	0.247	0.213
HLR-M2VS	0.802	0.734	0.866	0.730	0.757	0.713
t-SVD-MSC	0.684	0.562	0.750	0.555	0.582	0.543
Ours	**0.955**	**0.959**	**0.989**	**0.958**	**0.979**	**0.934**

For generic objects clustering problem, the results are presented in Tables 7 and 8 for COIL-20 dataset and Caltech-101 dataset, respectively. As shown, our method yields the best performance on both datasets, which confirms the effectiveness of our method. Our method is non-convex, so we have also compared our method with some non-convex multi-view subspace clustering methods including S_p-MSC [13], GLMSC [2] and TLRMSC [30], which has been published during the last two years. The comparison results are shown in Table 9.

For fair comparison, we maintain the original results in the referred papers. Although they are all non-convex approximations of tensor multi-rank function, our TALR approximation achieves the superior performance. There are two reasons. Firstly, a closed-form analytical solution is deduced for our optimization problem. Secondly, our surrogate functions are adaptively adjusted, which greatly improves the clustering accuracy. Besides, each view contains some knowledge that other views do not have, so a diverse regularized term is

Table 7. Experimental results on COIL-20. Here, $\lambda = 0.002$ and $\beta = 0.05$ for our model

Methods	ACC	F-score	NMI	AR	Recall	Precision
SPC_{best}	0.673	0.641	0.805	0.618	0.693	0.595
LRR_{best}	0.761	0.734	0.829	0.720	0.751	0.717
RMSC	0.685	0.656	0.800	0.637	0.698	0.620
DiMSC	0.778	0.745	0.846	0.732	0.751	0.739
LTMSC	0.804	0.760	0.860	0.748	0.776	0.741
ECMSC	0.782	0.794	0.942	0.781	0.925	0.695
HLR-M2VS	0.852	0.842	0.960	0.833	0.949	0.757
t-SVD-MSC	0.830	0.800	0.884	0.786	0.808	0.785
Ours	**0.992**	**0.986**	**0.993**	**0.985**	**0.987**	**0.984**

Table 8. Experimental results on Caltech-101. Here, $\lambda = 0.001$, $\beta = 0.002$

Methods	ACC	F-score	NMI	AR	Recall	Precision
SPC_{best}	0.485	0.341	0.724	0.318	0.234	0.596
LRR_{best}	0.510	0.339	0.728	0.304	0.231	0.627
RMSC	0.346	0.258	0.573	0.246	0.182	0.457
DiMSC	0.351	0.253	0.589	0.226	0.191	0.362
LTMSC	0.559	0.403	0.788	0.393	0.288	0.670
ECMSC	0.359	0.286	0.606	0.273	0.214	0.433
HLR-M2VS	0.650	0.442	0.872	0.463	0.343	0.760
t-SVD-MSC	0.607	0.440	0.858	0.430	0.323	0.742
Ours	**0.655**	**0.491**	**0.898**	**0.472**	**0.353**	**0.808**

introduced to further enhance our performance. Therefore, our model efficiently combines consistency and diversity associated with multi-view data. Accordingly, it can usually represent state-of-the-art performance.

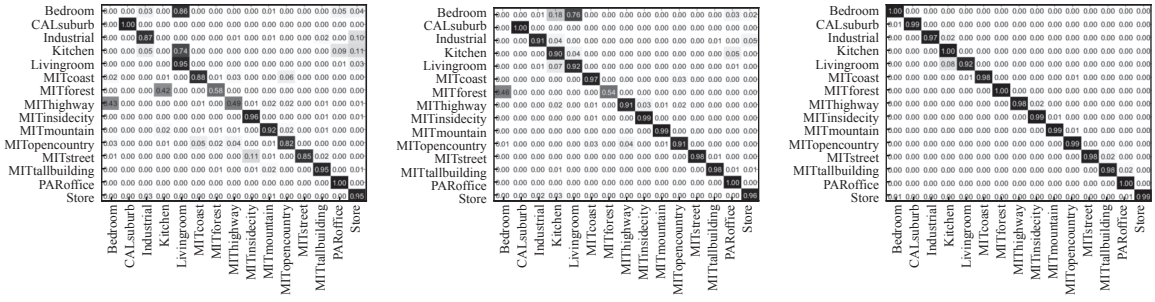
3. Comparisons of confusion matrices

Compared with the existing clustering methods, although t-SVD-MSC and HLR-M^{2}VS have achieved the great performance, the performance can be further improved by integrating consistency and diversity across multi-views, just like our method. This can be further illustrated by the confusion matrices calculated based on three different approaches on the Scene-15 dataset in the Fig.3.

As can be seen, the most confusing scenes are "bedroom", "living room" and "MIT forest". For HLR-M^{2}VS, the false alarm rate of clustering "living room" as "bedroom" is 0.76, and the false alarm rate of clustering "bedroom" as "MIT forest" is 0.46. It is notable that our model has reached very high performance in all categories, and the lowest accuracy is 0.92 for the scene of "living room". Especially, compared to the previous best performance reported in HLR-M^{2}VS [12], we can clearly observe the accuracy improvement in "bedroom" and "MIT forest" from 0.46 to 1.00 and from 0.54 to 1.00, respectively. For Scene-15 dataset, its images are all taken indoors with rich backgrounds, and the images in

Table 9. Comparisons to the other non-convex methods. The results of compared methods are from references [11], [12], and [25]

Datasets	Methods	ACC	F-score	NMI	AR	Recall	Precision
Yale	S_p-MSC	–	–	–	–	–	–
	GLMSC	0.979	0.958	0.975	**0.979**	0.955	0.959
	TLRMSC	0.976	0.966	0.983	0.964	0.973	0.960
	Ours	**0.983**	**0.975**	**0.987**	0.973	**0.980**	**0.970**
Extended YaleB	S_p-MSC	0.825	0.790	0.857	0.766	0.815	0.766
	GLMSC	–	–	–	–	–	–
	TLRMSC	0.966	0.934	0.947	0.927	0.936	0.932
	Ours	**0.975**	**0.952**	**0.960**	**0.947**	**0.954**	**0.950**
Scene-15	S_p-MSC	0.892	0.879	0.905	0.870	0.891	0.867
	GLMSC	0.904	0.901	0.932	0.933	0.893	0.916
	TLRMSC	0.970	0.942	0.942	0.938	0.940	0.945
	Ours	**0.986**	**0.975**	**0.974**	**0.973**	**0.976**	**0.974**



(a) t-SVD-MSC (b) HLR-M^{2}VS (c) Ours

Fig. 3. Confusion matrices of different methods on Scene-15 dataset. The vertical coordinate is the predicted labels, and the horizontal coordinate is the true labels.

the same category have both theme clarity and background diversity. Therefore, benefiting from combining consistency with diversity for the multi-view subspace clustering, our method has obtained the superior performance to the latest methods.

4. Ablation studies

In order to validate the roles of the different components in our model, we have removed the diversity part from our model and tuned the parameters to get the best performance. The HLR-M^{2}VS [12] is used as the "Baseline" method.

In Fig.4, we have shown the results of ablation study on the Extended YaleB dataset. The proposed method only with low-rank part has achieved the better performance than the baseline. Benefiting from the diversity part, the performance has continued to improve in term of all metrics.

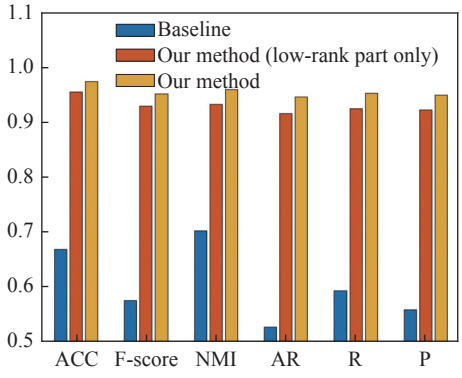


Fig. 4. Ablation study of the proposed method on the Extended YaleB dataset.

5. Analysis of parameters sensitivity

In our method, λ and β are two tuning parameters, in which λ denotes the noise level and β is a balance coefficient between the diversity regularizer and the consistency constraint on the representation tensor. Specifically, the optimal values of λ and β are determined by a exhausting grid searching in our experiments. Through carefully parameters tuning, we empirically find the best parameters for each dataset, just as shown in Tables 3–8.

For easily understanding, we have shown the evaluation process on Extended YaleB dataset in Fig.5. In this figure, the parameters λ and β are changed from 0.0021 to 0.004 and from 0.02 to 0.4 with steps 0.0001 and 0.02, respectively. The corresponding values of ACC and NMI are recorded. It is noteworthy to mention that the choice of λ may have a greater impact for the final scores. When the value of λ changed from 0.0021 to 0.004, the final performances have obvious fluctuations. In contrast, when the value of λ is fixed while the values of β are changed from 0.02 to 0.4, the final performances fluctuate in a small range.

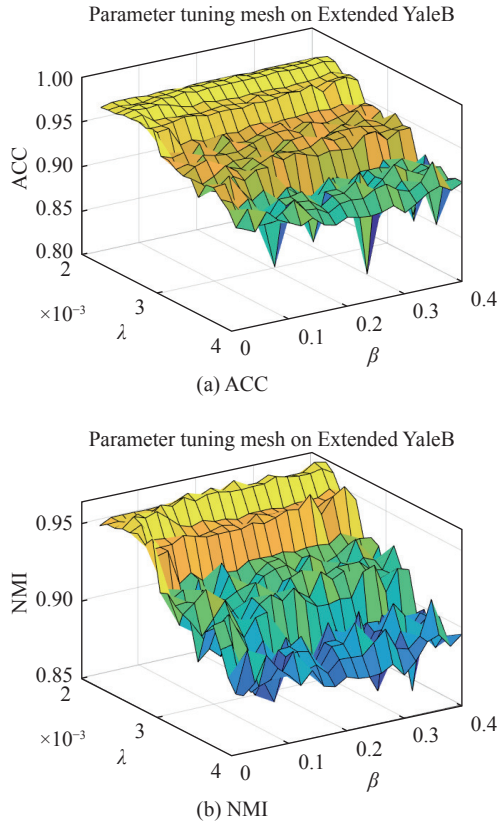


Fig. 5. Impact of different parameters (λ and β) for accuracy metrics (ACC and NMI) on Extended YaleB dataset.

6. Analysis of computational complexity

Since the inverse matrices can be calculated in advance, the computational cost of Algorithm 2 mainly depends on the updating of E and $\mathcal{H}$. For each iteration in updating E, the total cost is $\mathcal{O}(dNV)$ for V views. Here, we use d to denote the maximal dimension of data towards all views. To update $\mathcal{G}$, we need calculate in advance several preparatory values, including both 3D FFT and IFFT values for a $N \times V \times N$ tensor, and N number of SVD values for the $N \times V$ matrices $(N \geq V)$ under the Fourier domain. The computation cost is $\mathcal{O}\left(2N^2V\log(N) + N^2V^2\right)$ in each iteration.

Finally, we can obtain the total computational cost in Algorithm 2 by comprehensively integrating both the spectral clustering cost (e.g., $\mathcal{O}\left(N^3\right)$) and the iteration number (t) together, which can be written as $\mathcal{O}\left(N^3\right) + \mathcal{O}\left(t(2N^2V\log(N) + N^2V^2\right)$. Here, N and V denote the sample amount and view number respectively.

7. Analysis of algorithm convergence

Theorem 2 in Section IV ensures the convergence of our algorithm. Moreover, the reconstruction error (RE) as well as match error (ME) are defined as

$$\mathrm{RE} = \frac{1}{V}\sum_{v=1}^{V}\left\|X^{(v)} - X^{(v)}(Z^{(v)} + D^{(v)}) - E^{(v)}\right\|_{\infty} \quad (21)$$

$$\mathrm{ME}=\frac{1}{V}\sum_{v=1}^{V}\left\|Z^{(v)}-H^{(v)}\right\|_{\infty} \tag{22}$$

In Fig.6, we plot the convergence speed in terms of both RE and ME according to different iteration steps. The stopping criterion of our method has been given in Algorithm 2. If it is satisfied, the iterations are stopped. To demonstrate convergence speed in a more concise manner, we select the most representative convergence curves in Extended YaleB dataset. As shown, our algorithm generally converges quickly with a number of iterations from the Fig.6.

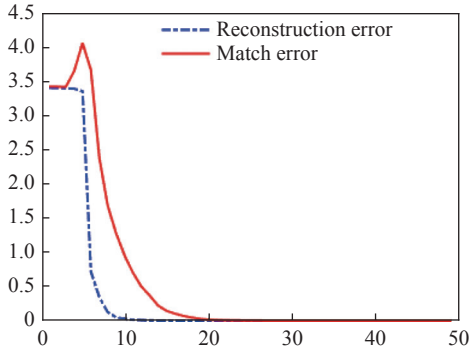


Fig. 6. Convergence curves of our method with the RE and ME vs. iterations on Extended YaleB dataset.

V. Conclusions

In the paper, we propose a new method for MSC problem. Based on consensus and complementary principles, a self-representation of multi-view data is decomposed into consensus part and diverse part. For consensus part, we propose to use the tensor adaptive log-determinant regularization term (TALR) as the surrogate of the tensorial multi-rank. It can reduce the over-penalization for large singular values in Fourier domain and improves the clustering accuracy by a large margin. For diverse part, we utilize a view-specific sparsity regularization to explore the inherent diversity in each view. We have deduced an optimal closed-form solution for the tensor non-convex minimization problem. Mathematically, we have proven the convergence of the whole algorithm. Extensive evaluations on six representative datasets show that our performance is superior to the eleven latest convex or non-convex methods.

Appendix A. Proof of Theorem 1

Let us consider the optimization problem as follows:

$$\underset{\mathcal{X}}{\operatorname{argmin}}\,\tau\|\mathcal{X}\|_{\mathrm{TALR}}+\frac{1}{2}\|\mathcal{X}-\mathcal{Y}\|_{\mathrm{F}}^{2} \tag{A-1}$$

We denote $f(\mathcal{X})=\tau\|\mathcal{X}\|_{\mathrm{TALR}}+\frac{1}{2}\|\mathcal{X}-\mathcal{Y}\|_{\mathrm{F}}^{2}$. According to the definition of t-product, in the Fourier domain, $f(\mathcal{X})$ can be re-written as following:

$$\begin{aligned}\hat{f}(\hat{\mathcal{X}})&=\tau\|\operatorname{bdiag}(\hat{\mathcal{X}})\|_{\mathrm{TALR}}+\frac{1}{2}\|\operatorname{bdiag}(\hat{\mathcal{X}}-\hat{\mathcal{Y}})\|_{\mathrm{F}}^{2}\\&=\sum_{j=1}^{k_3}\left(\tau\log\det(I+\gamma\hat{\mathcal{X}}^{(j)})+\frac{1}{2}\|\hat{\mathcal{X}}^{(j)}-\hat{\mathcal{Y}}^{(j)}\|_{\mathrm{F}}^{2}\right)\end{aligned} \tag{A-2}$$

If $\hat{\mathcal{X}}^{*}$ minimize $\hat{f}(\hat{\mathcal{X}})$, the optimal solution $\mathcal{X}^{*}$ of (A-1) can be obtained through inverse FFT: $\mathcal{X}^{*}=\mathrm{ifft}(\hat{\mathcal{X}}^{*},[],3)$. Since the minimization problem is separable, we can solve it slice by slice. Without loss of generality, for each frontal slice $\hat{\mathcal{X}}^{(j)}(1\le j\le k_3)$ of $\hat{\mathcal{X}}$, we just need to consider the following problem:

$$\underset{\hat{\mathcal{X}}^{(j)}}{\operatorname{argmin}}\,\tau\log\det(I+\gamma\hat{\mathcal{X}}^{(j)})+\frac{1}{2}\|\hat{\mathcal{X}}^{(j)}-\hat{\mathcal{Y}}^{(j)}\|_{\mathrm{F}}^{2} \tag{A-3}$$

Suppose $\mathcal{Y}$ has the t-SVD: $\mathcal{Y}=\mathcal{U}*\mathcal{D}_y*\mathcal{V}^{\mathrm{T}}$. We obtain k_3 blocks matrix SVD: $\hat{\mathcal{Y}}^{(j)}=\hat{\mathcal{U}}^{(j)}\hat{\mathcal{D}}_y^{(j)}(\hat{\mathcal{V}}^{(j)})^{\mathrm{T}}(1\le j\le k_3)$ in the Fourier domain. Equation (A-3) equals to

$$\underset{\hat{\mathcal{D}}_x(i,i,j)}{\operatorname{argmin}}\;\tau\log(\gamma\hat{\mathcal{D}}_x(i,i,j)+1)+\frac{1}{2}(\hat{\mathcal{D}}_x(i,i,j)-\hat{\mathcal{D}}_y(i,i,j))^{2} \tag{A-4}$$

Let $\hat{\mathcal{D}}_x(i,i,j)=\sigma$, we consider the following smooth function:

$$g(\sigma)=\tau\log(\gamma\sigma+1)+\frac{1}{2}(\sigma-\hat{\mathcal{D}}_y(i,i,j))^{2} \tag{A-5}$$

The first-order necessary optimal condition is

$$g'(\sigma)=\frac{\tau}{\gamma\sigma+1}+\sigma-\hat{\mathcal{D}}_y(i,i,j)=0 \tag{A-6}$$

which is equivalent to

$$\sigma^{2}+\left(\frac{1}{\gamma}-\hat{\mathcal{D}}_y(i,i,j)\right)\sigma+\frac{1}{\gamma}(\tau-\hat{\mathcal{D}}_y(i,i,j))=0 \tag{A-7}$$

Let $\Delta=\left(\frac{1}{\gamma}-\hat{\mathcal{D}}_y(i,i,j)\right)^{2}-\frac{4}{\gamma}\left(\tau-\hat{\mathcal{D}}_y(i,i,j)\right)$, the optimal solution of $\hat{\mathcal{D}}_x(i,i,j)$ can be denoted as

$$\hat{\mathcal{D}}_x(i,i,j)=\begin{cases}\dfrac{\left(\hat{\mathcal{D}}_y(i,i,j)-\frac{1}{\gamma}\right)+\sqrt{\Delta}}{2}, & \Delta\ge 0\\ 0, & \Delta<0\end{cases} \tag{A-8}$$

Consequently, the minimization problem of (A-1) is obtained through $\mathcal{X}^{*}=\mathcal{U}*\mathcal{D}_x*\mathcal{V}^{\mathrm{T}}$. The proof is completed.

Appendix B. Proof of Theorem 2

Let $\mathcal{P}^{t}=\{Z^{(v)^{t}},E^{(v)^{t}},D^{(v)^{t}},\mathcal{H}^{t}\},\mathcal{Y}^{t}=\{Y^{(v)^{t}},\mathcal{W}^{t}\},1\le t<\infty$. Through the iterative scheme of Algorithm 2, we have

$$\min_{\mathcal{P}}\mathcal{L}\left(\mathcal{P}^{t},\mathcal{Y}^{t},\mu^{t},\rho^{t}\right)=\mathcal{L}\left(\mathcal{P}^{t+1},\mathcal{Y}^{t},\mu^{t},\rho^{t}\right) \tag{B-1}$$

It yields

$$\mathcal{L}\left(\mathcal{P}^{t+1},\mathcal{Y}^{t},\mu^{t},\rho^{t}\right)\le\mathcal{L}\left(\mathcal{P}^{t},\mathcal{Y}^{t},\mu^{t},\rho^{t}\right) \tag{B-2}$$

For convenience, suppose $\Delta^{t}=X^{(v)}-X^{(v)}(Z^{(v)^{t}}+D^{(v)^{t}})-E^{(v)^{t}}$. From (8), we know

$$
\begin{aligned}
&\mathcal{L}\left(\mathcal{P}^t,\mathcal{Y}^t,\mu^t,\rho^t\right)\\
&=\mathcal{L}\left(\mathcal{P}^t,\mathcal{Y}^{t-1},\mu^{t-1},\rho^{t-1}\right)+\sum_{v=1}^{V}\left(\langle Y^{(v)^t},\Delta^t\rangle+\frac{\mu^t}{2}\left\|\Delta^t\right\|_{\mathrm{F}}^2\right)\\
&\quad-\sum_{v=1}^{V}\left(\langle Y^{(v)^{t-1}},\Delta^t\rangle+\frac{\mu^{t-1}}{2}\left\|\Delta^t\right\|_{\mathrm{F}}^2\right)+\langle\mathcal{W}^t,\mathcal{Z}^t-\mathcal{H}^t\rangle\\
&\quad+\frac{\rho^t}{2}\left\|\mathcal{Z}^t-\mathcal{H}^t\right\|_{\mathrm{F}}^2-\langle\mathcal{W}^{t-1},\mathcal{Z}^t-\mathcal{H}^t\rangle-\frac{\rho^{t-1}}{2}\left\|\mathcal{Z}^t-\mathcal{H}^t\right\|_{\mathrm{F}}^2\\
&=\mathcal{L}\left(\mathcal{P}^t,\mathcal{Y}^{t-1},\mu^{t-1},\rho^{t-1}\right)\\
&\quad+\sum_{v=1}^{V}\left(\langle Y^{(v)^t}-Y^{(v)^{t-1}},\Delta^t\rangle+\frac{\mu^t-\mu^{t-1}}{2}\left\|\Delta^t\right\|_{\mathrm{F}}^2\right)\\
&\quad+\langle\mathcal{W}^t-\mathcal{W}^{t-1},\mathcal{Z}^t-\mathcal{H}^t\rangle+\frac{\rho^t-\rho^{t-1}}{2}\left\|\mathcal{Z}^t-\mathcal{H}^t\right\|_{\mathrm{F}}^2
\end{aligned}
\tag{B-3}
$$

In addition, from equations (19) and (20), we can notice that

$$
\Delta^t=\frac{Y^{(v)^t}-Y^{(v)^{t-1}}}{\mu^{t-1}},\quad \mathcal{Z}^t-\mathcal{H}^t=\frac{\mathcal{W}^t-\mathcal{W}^{t-1}}{\rho^{t-1}}
\tag{B-4}
$$

Then

$$
\begin{aligned}
&\mathcal{L}\left(\mathcal{P}^t,\mathcal{Y}^t,\mu^t,\rho^t\right)\\
&=\mathcal{L}\left(\mathcal{P}^t,\mathcal{Y}^{t-1},\mu^{t-1},\rho^{t-1}\right)\\
&\quad+\sum_{v=1}^{V}\left(\frac{1}{\mu^{t-1}}\left\|Y^{(v)^t}-Y^{(v)^{t-1}}\right\|_{\mathrm{F}}^2+\frac{\mu^t-\mu^{t-1}}{2(\mu^{t-1})^2}\left\|Y^{(v)^t}-Y^{(v)^{t-1}}\right\|_{\mathrm{F}}^2\right)\\
&\quad+\frac{1}{\rho^{t-1}}\left\|\mathcal{W}^t-\mathcal{W}^{t-1}\right\|_{\mathrm{F}}^2+\frac{\rho^t-\rho^{t-1}}{2(\rho^{t-1})^2}\left\|\mathcal{W}^t-\mathcal{W}^{t-1}\right\|_{\mathrm{F}}^2\\
&=\mathcal{L}\left(\mathcal{P}^t,\mathcal{Y}^{t-1},\mu^{t-1},\rho^{t-1}\right)+\frac{\mu^t+\mu^{t-1}}{2(\mu^{t-1})^2}\sum_{v=1}^{V}\left\|Y^{(v)^t}-Y^{(v)^{t-1}}\right\|_{\mathrm{F}}^2\\
&\quad+\frac{\rho^t+\rho^{t-1}}{2(\rho^{t-1})^2}\left\|\mathcal{W}^t-\mathcal{W}^{t-1}\right\|_{\mathrm{F}}^2
\end{aligned}
\tag{B-5}
$$

Combined with (B-2), we get

$$
\begin{aligned}
&\mathcal{L}\left(\mathcal{P}^{t+1},\mathcal{Y}^t,\mu^t,\rho^t\right)\\
&\le\mathcal{L}\left(\mathcal{P}^t,\mathcal{Y}^{t-1},\mu^{t-1},\rho^{t-1}\right)\\
&\quad+\frac{\mu^t+\mu^{t-1}}{2(\mu^{t-1})^2}\sum_{v=1}^{V}\left\|Y^{(v)^t}-Y^{(v)^{t-1}}\right\|_{\mathrm{F}}^2+\frac{\rho^t+\rho^{t-1}}{2(\rho^{t-1})^2}\left\|\mathcal{W}^t-\mathcal{W}^{t-1}\right\|_{\mathrm{F}}^2
\end{aligned}
\tag{B-6}
$$

Iterating the inequality in (B-6) for t times, we can arrive at

$$
\begin{aligned}
&\mathcal{L}\left(\mathcal{P}^{t+1},\mathcal{Y}^t,\mu^t,\rho^t\right)\le\mathcal{L}\left(\mathcal{P}^1,\mathcal{Y}^0,\mu^0,\rho^0\right)\\
&\quad+\sum_{j=1}^{t}\left(\frac{\mu^j+\mu^{j-1}}{2(\mu^{j-1})^2}\sum_{v=1}^{V}\left\|Y^{(v)^j}-Y^{(v)^{j-1}}\right\|_{\mathrm{F}}^2\right.\\
&\quad\left.+\frac{\rho^j+\rho^{j-1}}{2(\rho^{j-1})^2}\left\|\mathcal{W}^j-\mathcal{W}^{j-1}\right\|_{\mathrm{F}}^2\right)
\end{aligned}
\tag{B-7}
$$

Here, $\{Y^{(v)^t},\mathcal{W}^t\}$ is bounded, μ^t,ρ^t is non-decreasing, and $\sum_{t=0}^{\infty}\frac{\mu^{t+1}}{(\mu^t)^2}<\infty$, $\sum_{t=0}^{\infty}\frac{\rho^{t+1}}{(\rho^t)^2}<\infty$, so $\mathcal{L}\left(\mathcal{P}^{t+1},\mathcal{Y}^t,\mu^t,\rho^t\right)$ is bounded.

Adding $\frac{1}{2(\mu^{t-1})}\sum_{v=1}^{V}\|Y^{(v)^{t-1}}\|_{\mathrm{F}}^2+\frac{1}{2(\rho^{t-1})}\|\mathcal{W}^{t-1}\|_{\mathrm{F}}^2$ to the augmented Lagrangian function (8), and considering (B-4), we can rewrite (8) as

$$
\begin{aligned}
&\|\mathcal{H}^t\|_{\mathrm{TALR}}+\lambda\|E^t\|_{2,1}+\beta\sum_{v=1}^{V}\left\|D^{(v)^t}\right\|_{\mathrm{F}}^2\\
&=\mathcal{L}\left(Z^{(v)^t},E^{(v)^t},D^{(v)^t},\mathcal{H}^t,Y^{(v)^{t-1}},\mathcal{W}^{t-1},\mu^{t-1},\rho^{t-1}\right)\\
&\quad-\sum_{v=1}^{V}\left(\langle Y^{(v)^{t-1}},\Delta^t\rangle+\frac{\mu^{t-1}}{2}\left\|\Delta^t\right\|_{\mathrm{F}}^2\right)\\
&\quad-\langle\mathcal{W}^{t-1},\mathcal{Z}^t-\mathcal{H}^t\rangle-\frac{\rho^{t-1}}{2}\left\|\mathcal{Z}^t-\mathcal{H}^t\right\|_{\mathrm{F}}^2\\
&=\mathcal{L}\left(Z^{(v)^t},E^{(v)^t},D^{(v)^t},\mathcal{H}^t,Y^{(v)^{t-1}},\mathcal{W}^{t-1},\mu^{t-1},\rho^{t-1}\right)\\
&\quad-\frac{1}{2(\mu^{t-1})}\sum_{v=1}^{V}\left\|Y^{(v)^t}\right\|_{\mathrm{F}}^2-\frac{1}{2(\rho^{t-1})}\left\|\mathcal{W}^{(v)^t}\right\|_{\mathrm{F}}^2
\end{aligned}
\tag{B-8}
$$

Because the right side of (B-8) is bounded and each term on the left side is nonnegative, each term on the left side must be bounded. It indicates that $\{E^{(v)^t},D^{(v)^t},\mathcal{H}^t\}$ are all bounded. Simultaneously, since $\mathcal{Z}^t=\mathcal{H}^t+\frac{\mathcal{W}^t-\mathcal{W}^{t-1}}{\rho^{t-1}}$, we can deduce that $\{Z^{(v)^t}\}$ is bounded. The proof of Theorem 2.1) is completed.

As aforementioned, $\{Z^{(v)^t},E^{(v)^t},D^{(v)^t},\mathcal{H}^t,Y^{(v)^t},\mathcal{W}^t\}$ is a bounded sequence. According to the Bolzano-Weierstrass Theorem, this sequence must have at least one cluster point. We assume that $\{Z^{(v)^t},E^{(v)^t},D^{(v)^t},\mathcal{H}^t,Y^{(v)^t},\mathcal{W}^t\}$ converges to $\{Z^{(v)^*},E^{(v)^*},D^{(v)^*},\mathcal{H}^*,Y^{(v)^*},\mathcal{W}^*\}$ without loss of generality, and $\mathcal{Z}^*=\Phi(Z^{(1)^*},Z^{(2)^*},\ldots,Z^{(v)^*})$. According to (B-8), we have $X^{(v)}=X^{(v)}(Z^{(v)^*}+D^{(v)^*})+E^{(v)^*},\mathcal{H}^*=\mathcal{Z}^*$.

In the subproblem for $\mathcal{H}$, the first-order optimal condition of (16) is equivalent to

$$
\nabla_{\mathcal{H}}\left(\|\mathcal{H}\|_{\mathrm{TALR}}\right)|_{\mathcal{H}^{t+1}}-\mathcal{W}^{t+1}+\rho^t\left(\mathcal{H}^{t+1}-\mathcal{Z}^{t+1}\right)=0
\tag{B-9}
$$

Then we have

$$
\nabla_{\mathcal{H}}\left(\|\mathcal{H}\|_{\mathrm{TALR}}\right)|_{\mathcal{H}^*}=\mathcal{W}^*
\tag{B-10}
$$

Similarly, we can easily verify that

$$
\begin{aligned}
&\nabla_{Z^{(v)}}\mathcal{L}\left(Z^{(v)},E^{(v)},D^{(v)},\mathcal{H},Y^{(v)},\mathcal{W}\right)\Big|_{Z^{(v)^*}}\\
&=-X^{(v)\mathrm{T}}Y^{(v)^*}+W^{(v)^*}=0
\end{aligned}
\tag{B-11}
$$

$$
\begin{aligned}
&\nabla_{E^{(v)}}\mathcal{L}\left(Z^{(v)},E^{(v)},D^{(v)},\mathcal{H},Y^{(v)},\mathcal{W}\right)\Big|_{E^{(v)^*}}\\
&=\lambda\,\nabla_{E^{(v)}}\left(\|E\|_{2,1}\right)\big|_{E^{(v)^*}}-Y^{(v)^*}=0
\end{aligned}
\tag{B-12}
$$

and

$$
\begin{aligned}
&\nabla_{D^{(v)}}\mathcal{L}\left(Z^{(v)},E^{(v)},D^{(v)},\mathcal{H},Y^{(v)},\mathcal{W}\right)\Big|_{D^{(v)^*}}\\
&=2\beta D^{(v)^*}-X^{(v)\mathrm{T}}Y^{(v)^*}=0
\end{aligned}
\tag{B-13}
$$

Now we observe that $\{Z^{(v)^*},E^{(v)^*},D^{(v)^*},\mathcal{H}^*,Y^{(v)^*},\mathcal{W}^*\}$ satisfies the KKT conditions of $\{Z^{(v)^t},E^{(v)^t},D^{(v)^t},\mathcal{H}^t,Y^{(v)^t},\mathcal{W}^t\}$.

Therefore, $\{Z^{(v)^*},E^{(v)^*},D^{(v)^*},\mathcal{H}^*\}$ is a stationary KKT point of the optimization problem (8). The proof is completed.

References

[1] Y. Y. Chen, S. Q. Wang, C. Peng, *et al.*, "Generalized nonconvex low-rank tensor approximation for multi-view subspace clustering," *IEEE Transactions on Image Processing*,

vol.30, pp.4022–4035, 2021.

[2] Q. X. Guo, W. Xia, Z. Z. Wan, *et al.*, "Tensor-SVD based graph learning for multi-view subspace clustering," in *Proceedings of the Thirty-Fourth AAAI Conference on Artificial Intelligence*, New York, NY, USA, pp.3930–3937, 2020.

[3] X. C. Cao, C. Q. Zhang, H. Z. Fu, *et al.*, "Diversity-induced multi-view subspace clustering," in *Proceedings of 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Boston, MA, USA, pp.586–594, 2015.

[4] X. H. Liu, S. W. Ji, W. Glänzel, *et al.*, "Multiview partitioning via tensor methods," *IEEE Transactions on Knowledge and Data Engineering*, vol.25, no.5, pp.1056–1069, 2013.

[5] C. Q. Zhang, H. Z. Fu, S. Liu, *et al.*, "Low-rank tensor constrained multiview subspace clustering," in *Proceedings of 2015 IEEE International Conference on Computer Vision (ICCV)*, Santiago, Chile, pp.1582–1590, 2015.

[6] X. L. Piao, Y. L. Hu, J. B. Gao, *et al.*, "Tensor sparse and low-rank based submodule clustering method for multi-way data," *arXiv preprint*, arXiv: 1601.00149, 2016.

[7] M. Yin, J. B. Gao, S. L. Xie, *et al.*, "Multiview subspace clustering via tensorial t-product representation," *IEEE Transactions on Neural Networks and Learning Systems*, vol.30, no.3, pp.851–864, 2019.

[8] Y. Xie, D. C. Tao, W. S. Zhang, *et al.*, "On unifying multi-view self-representations for clustering by tensor multi-rank minimization," *International Journal of Computer Vision*, vol.126, no.11, pp.1157–1179, 2018.

[9] Y. Y. Chen, X. L. Xiao, and Y. Y. Zhou, "Multi-view subspace clustering via simultaneously learning the representation tensor and affinity matrix," *Pattern Recognition*, vol.106, article no.107441, 2020.

[10] X. B. Wang, Z. Lei, X. J. Guo, *et al.*, "Multi-view subspace clustering with intactness-aware similarity," *Pattern Recognition*, vol.88, pp.50–63, 2019.

[11] G. C. Liu, Z. C. Lin, S. C. Yan, *et al.*, "Robust recovery of subspace structures by low-rank representation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol.35, no.11, pp.171–184, 2013.

[12] Y. Xie, W. S. Zhang, Y. Y. Qu, *et al.*, "Hyper-Laplacian regularized multilinear multiview self-representations for clustering and semisupervised learning," *IEEE Transactions on Cybernetics*, vol.50, no.2, pp.572–586, 2020.

[13] Y. L. Liu, X. Q. Zhang, G. Y. Tang, *et al.*, "Multi-view subspace clustering based on tensor Schatten-p norm," in *Proceedings of 2019 IEEE International Conference on Big Data (Big Data)*, Los Angeles, CA, USA, 2019.

[14] C. Y. Lu, J. S. Feng, Y. D. Chen, *et al.*, "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.

[15] S. Bickel and T. Scheffer, "Multi-view clustering," in *Proceedings of Fourth IEEE International Conference on Data Mining*, Brighton, UK, pp.19–26, 2004.

[16] A. Kumar and H. Daume III, "A co-training approach for multi-view spectral clustering," in *Proceedings of the 28th International Conference on International Conference on Machine Learning*, Bellevue, WA, USA, 2011, pp.393–400.

[17] S. P. Yu, B. Krishnapuram, R. Rosales, *et al.*, "Bayesian co-training," *The Journal of Machine Learning Research*, vol.12, pp.2649–2680, 2011.

[18] D. Greene and P. Cunningham, "A matrix factorization approach for integrating multiple data views," in *Proceedings of the European Conference on Machine Learning and Knowledge Discovery in Databases*, Bled, Slovenia, pp.423–438, 2009.

[19] G. Tzortzis and A. Likas, "Kernel-based weighted multi-view clustering," in *Proceedings of the IEEE 12th International Conference on Data Mining*, Brussels, Belgium, pp.675–684, 2012.

[20] J. L. Liu, C. Wang, J. Gao, *et al.*, "Multi-view clustering via joint nonnegative matrix factorization," in *Proceedings of the 2013 SIAM International Conference on Data Mining*, Austin, TX, USA, pp.252–260, 2013.

[21] X. Jiang, B. Y. Zheng, L. Wang, *et al.*, "Clustering for topological interference management," *Chinese Journal of Electronics*, vol.31, no.5, pp.844–850, 2022.

[22] E. Elhamifar and R. Vidal, "Sparse subspace clustering: Algorithm, theory, and applications," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol.35, no.11, pp.2765–2781, 2013.

[23] R. K. Xia, Y. Pan, L. Du, *et al.*, "Robust multi-view spectral clustering via low-rank and sparse decomposition," in *Proceedings of the Twenty-Eighth AAAI Conference on Artificial Intelligence*, Quebec City, Canada, pp.2149–2155, 2014.

[24] Y. X. Wang, H. Xu, and C. L. Leng, "Provable subspace clustering: When LRR meets SSC," *IEEE Transactions on Information Theory*, vol.65, no.9, pp.5406–5432, 2019.

[25] X. B. Wang, X. J. Guo, Z. Lei, *et al.*, "Exclusivity-consistency regularized multi-view subspace clustering," in *Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Honolulu, HI, USA, pp.1–9, 2017.

[26] B. X. Zhao, S. L. Wang, and C. L. Liu, "State: A clustering algorithm focusing on edges instead of centers," *Chinese Journal of Electronics*, vol.30, no.5, pp.902–908, 2021.

[27] W. L. Huang, X. D. Wang, G. H. Li, *et al.*, "Semi-supervised artificial immune mixture models clustering," *Chinese Journal of Electronics*, vol.25, no.2, pp.249–255, 2016.

[28] S. L. Wang, Q. Li, H. N. Yuan, *et al.*, "Robust clustering with topological graph partition," *Chinese Journal of Electronics*, vol.28, no.1, pp.76–84, 2019.

[29] M. E. Kilmer and C. D. Martin, "Factorization strategies for third-order tensors," *Linear Algebra and Its Applications*, vol.435, no.3, pp.641–658, 2011.

[30] X. L. Sun, Y. J. Wang, and X. J. Zhang, "Multi-view subspace clustering via non-convex tensor rank minimization," in *Proceedings of 2020 IEEE International Conference on Multimedia and Expo (ICME)*, London, UK, pp.1–6, 2020.

[31] A. Quattoni and A. Torralba, "Recognizing indoor scenes," in *Proceedings of the 2009 IEEE Conference on Computer Vision and Pattern Recognition*, Miami, FL, USA, pp.413–420, 2009.

[32] F. F. Li, R. Fergus, and P. Perona, "Learning generative visual models from few training examples: An incremental Bayesian approach tested on 101 object categories," *Computer Vision and Image Understanding*, vol.106, no.1, pp.59–70, 2007.

[33] T. Ojala, M. Pietikainen, and T. Maenpaa, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol.24, no.7, pp.971–987, 2002.

[34] M. Lades, J. C. Vorbruggen, J. Buhmann, *et al.*, "Distortion invariant object recognition in the dynamic link architecture," *IEEE Transactions on Computers*, vol.42, no.3, pp.300–311, 1993.

[35] J. X. Wu and J. M. Rehg, "CENTRIST: A visual descriptor for scene categorization," *IEEE Transactions on Pattern*

Analysis and Machine Intelligence, vol.33, no.8, pp.1489–1501, 2011.

[36] X. B. Qi, R. Xiao, C. G. Li, *et al.*, "Pair wise rotation invariant co-occurrence local binary pattern," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol.36, no.11, pp.2199–2213, 2014.

[37] A. Bosch, A. Zisserman, and X. Munoz, "Image classification using random forests and ferns," in *Proceedings of the 2007 IEEE 11th International Conference on Computer Vision*, Rio De Janeiro, Brazil, pp.1–8, 2007.

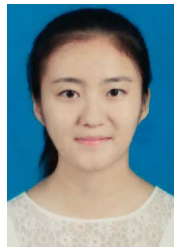
[38] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," in *Proceedings of the 3rd International Conference on Learning Representations*, San Diego, CA, USA, 2015.

[39] C. Szegedy, V. Vanhoucke, S. Ioffe, J. *et al.*, "Rethinking the inception architecture for computer vision," in *Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, NV, USA, pp.2818–2826, 2016.

[40] C. D. Manning, P. Raghavan, and H. Schütze, *Introduction to Information Retrieval*, Cambridge University Press, Cambridge, UK, 2008.

[41] A. Y. Ng, M. I. Jordan, and Y. Weiss, "On spectral clustering: Analysis and an algorithm," in *Proceedings of the 14th International Conference on Neural Information Processing System*, Vancouver, British, pp.849–856, 2001.

SUN Xiaoli received the Ph.D. degree in applied mathematics from Xidian University in 2007. She is currently an Associate Professor at the College of Mathematics and Statistics, Shenzhen University, Shenzhen. Her research interests include subspace clustering, saliency detection, and computer vision. (Email: xlsun@szu.edu.cn)



HAI Yang received the B.S. degree in mathematics and applied mathematics from Jiangxi Normol University in 2019. She is currently a master student at the College of Mathematics and Statistics, Shenzhen University, Shenzhen, China. Her research interests are in the fields of pattern recognition and intelligent computing.
(Email: 2060201016@email.szu.edu.cn)

ZHANG Xiujun (corresponding author) received the B.S. degree and M.S. degree in electrical engineering from Xidian University, Xi'an, China, in 2002 and 2005, respectively, and received the Ph.D. degree from Shenzhen University, Shenzhen, China in 2016. He is now working at the School of Electronic and Communication Engineering, Shenzhen Polytechnic, Shenzhen. His current research interests include saliency detection, subspace clustering, sparse and low-rank representation in image processing.
(Email: zhangxiujun@szpt.edu.cn)

XU Chen received the B.S. and M.S. degrees in Mathematics from Xidian University in 1986 and 1989, the Ph.D. degree in Mathematics from Xi'an Jiaotong University in 1992 respectively. He is currently a Professor of mathematics and Ph.D. Supervisor at Shenzhen University, Shenzhen. His research fields are information and computational science, analysis and application of wavelet.
(Email: xuchen_szu@szu.edu.com)