Bi-Nuclear Tensor Schatten-p Norm Minimization for Multi-view Subspace Clustering

Shuqin Wang, Zhiping Lin, *Senior Member, IEEE,* Qi Cao, Yigang Cen and Yongyong Chen, *Member, IEEE*

***Abstract*—Multi-view subspace clustering aims to integrate the complementary information contained in different views to facilitate data representation. Currently, low-rank representation (LRR) serves as a benchmark method. However, we observe that these LRR-based methods would suffer from two issues: *limited clustering performance* and *high computational cost* since (1) they usually adopt the nuclear norm with biased estimation to explore the low-rank structures; (2) the singular value decomposition of large-scale matrices is inevitably involved. Moreover, LRR may not achieve low-rank properties in both intra-views and inter-views simultaneously. To address the above issues, this paper proposes the *Bi-nuclear tensor Schatten-p norm minimization for multi-view subspace clustering (BTMSC)*. Specifically, BTMSC constructs a third-order tensor from the view dimension to explore the high-order correlation and the subspace structures of multi-view features. The Bi-Nuclear Quasi-Norm (BiN) factorization form of the Schatten-p norm is utilized to factorize the third-order tensor as the product of two small-scale third-order tensors, which not only captures the low-rank property of the third-order tensor but also improves the computational efficiency. Finally, an efficient alternating optimization algorithm is designed to solve the BTMSC model. Extensive experiments with ten datasets of texts and images illustrate the performance superiority of the proposed BTMSC method over state-of-the-art methods.**

***Index Terms*—Multi-view subspace clustering, low-rank representation, tensor factorization, Schatten-p norm**

I. Introduction

DUE to the popularity of multi-view data and the difficulty of acquiring label information, multi-view clustering has been widely used in unsupervised knowledge discovery to automatically find underlying correlations [1], [2]. For example, one person may be identified by several different features, such as face poses, pedestrian walking postures and palmprint images. Various visual descriptors, such as local binary patterns, scale-invariant feature transform and histogram of oriented gradients, have been adopted to extract corresponding visual features for some specific tasks, such as recognition and detection. Researchers have long been working on different types of approaches for multi-view clustering. Among them, multi-view subspace clustering has attracted much attention.

The core of multi-view subspace clustering is how to build an intrinsic matrix to capture both the consistent and complementary principles of multiple views [3], [4]. One direct way is to adopt the cosine or heat kernel methods [5] to compute the pair-wise distances on the raw data without considering the noise disturbance. Furthermore, the kernel-based methods [6], [7] are limited to the predefined kernel and the corresponding parameters. Towards this issue, low-rank representation (LRR) [8] and sparse subspace clustering (SSC) [9] are proposed to overcome the noise interference with the basic assumption that data can be linearly represented by samples belonging to the same subspace. Many multi-view subspace clustering methods based on LRR have been proposed [10]–[13]. In terms of the construction form of self-representation coefficients, these methods can be divided into matrix representation [10], [11] and tensor representation methods [12], [13]. Matrix-based methods only explore the pair-wise correlation between views by processing each self-representation matrix or their fusion matrices separately, and ignore the correlation among views. Note that tensor-based methods commonly construct a third-order tensor to explore high-order correlation of multiple views.

Technically, the aforementioned multi-view subspace clustering methods utilize the nuclear norm to describe the low-rank property of a similarity matrix. However, there are two common limitations for the above methods: (1) the nuclear norm is the convex approximation of the rank function yet the biased estimation. To alleviate the bias of the nuclear norm, a large number of researchers [14]–[17] borrowed the nonconvex functions to further improve the approximation performance of the rank function, which is termed as nonconvex approximation. However, nonconvex approximation methods have nonconvex, nonsmooth and non-Lipschitz problems. Fortunately, the factorization form of the nonconvex Schatten-p norm is proved to be tractable [18]. (2) the high computational cost of each iteration has greatly limited their practical applicability since each iteration involves singular value decomposition (SVD) of large-scale matrices. To improve efficiency, Sun and Wang *et al.* [19], [20] applied anchor to represent the actual latent data distribution, which is essentially a matrix factorization-based method. Although the matrix factorization-based clustering methods [21]–[23] effectively reduce the computational complexity, they fail to closely combine the

This work was supported in part by the National Key Research and Development Program of China under Grant 2021YFE0110500, in part by the National Natural Science Foundation of China under Grant 62062021 and 62106063, in part by the Guangdong Natural Science Foundation under Grant 2022A1515010819, and in part by the Shenzhen Science and Technology Program under Grant RCBS20210609103708013. (Corresponding authors: Yigang Cen and Yongyong Chen.)

S. Wang and Y. Cen are with Institute of Information Science, Beijing Jiaotong University and Beijing Key Laboratory of Advanced Information Science and Network Technology, Beijing 100044, China (Email: ShuqinWang.cn@hotmail.com; ygcen@bjtu.edu.cn).

Z. Lin is with the School of Electrical and Electronic Engineering, Nanyang Technological University, 639798, Singapore (Email: ezplin@ntu.edu.sg).

Q. Cao is with the School of Computer Science, University of Glasgow, 567739, Singapore (Email: qi.cao@glasgow.ac.uk).

Y. Chen is with the Bio-Computing Research Center, Harbin Institute of Technology, Shenzhen 518055, China (Email: YongyongChen.cn@gmail.com).

multi-view matrix factorization with the essential structure of the view, resulting in unsatisfactory performance.

To overcome the above mentioned limitations, in this paper, we aim to design a multi-view subspace clustering method to achieve both clustering performance and computational efficiency. We propose the Bi-nuclear tensor Schatten-p norm minimization for multi-view subspace clustering (BTMSC). To be specific, to simultaneously explore the consistent principle and the complementary principle (*i.e.*, high-order correlation) of multi-view features, BTMSC constructs the third-order tensor from the view dimension. BTMSC utilizes the nonconvex Schatten-p norm and its Bi-Nuclear Quasi-Norm (BiN) factorization form instead of the nuclear norm to factorize the tensor into the product of two small-scale tensors, which not only reduces the computational complexity, but also better preserves the global low-rank structure information of multiple views. Consequently, the learned similarity matrix from the third-order tensor could accurately represent the latent data distribution. An efficient algorithm with seven subproblems is derived by the alternating direction method of multipliers. The main contributions of this paper are summarized as follows:

- To overcome the biased estimation and high computational cost of LRR-based methods, we propose the Bi-nuclear tensor Schatten-p norm minimization for multi-view subspace clustering (BTMSC). The proposed method combines the nonconvex low-rank constraint and tensor factorization to reduce the computational complexity and improve the clustering performance.
- Instead of the coefficient construction form of matrix representation, BTMSC strengthens the correlation among multiple views by constructing the self-representation matrices as a tensor along the view dimension. Simultaneously, BTMSC uses the factorized Bi-Nuclear Quasi-Norm (BiN) of Schatten-p norm as the rank relaxation to find a more robust low-dimensional subspace accurately and efficiently.
- We design an efficient iterative algorithm by the alternating direction method of multipliers to solve the resultant optimization problem and validate the clustering performance of the proposed method on ten real datasets. Comprehensive experiments demonstrate that the proposed BTMSC significantly outperforms state-of-the-art methods in clustering accuracy and computational efficiency.

The remainder of this paper is organized as follows. Some subspace clustering and matrix factorization-based clustering methods are reviewed in Section II. Section III shows some notations and preliminaries of Schatten-p norm and tensor. The proposed BTMSC method and its optimization process are described in Section IV. Section V shows the theoretical convergence of the proposed BTMSC method. We conduct the clustering experiments and model analysis on some datasets in Section VI. The conclusions of this paper are summarized in Section VII.

II. Related Work

In this section, we briefly review subspace clustering and matrix factorization-based clustering methods related to our proposed method. The nonnegative matrix factorization (NMF) methods are often used to find a low-rank approximation by the product of two nonnegative matrices. Given the feature matrix $X = [x_1, x_2, \cdots, x_N] \in R^{d\times N}$, each column of which is a d dimensional sample and N is the number of samples. Let $X \approx UV^T$, the objective function of the standard NMF model based on Frobenius norm $\|\cdot\|_F = \sqrt{\sum_{ij} x_{ij}^2}$ is defined as

$$\min_{U\geq 0, V\geq 0} \|X - UV^T\|_F^2, \tag{1}$$

where $U = [u_1, u_2, \cdots, u_k] \in R^{d\times k}$, $V = [v_1, v_2, \cdots, v_k] \in R^{N\times k}$. Currently, many studies apply NMF to clustering as a dimensionality reduction method. For example, Cai *et al.* [24] proposed the graph regularized nonnegative matrix factorization (GNMF) method, which incorporates manifold structure into NMF to explore the intrinsic geometric structure of the data. Peng *et al.* [22] integrated robust principal component analysis and GNMF into a unified framework, enhancing the ability to recover the nonlinear structure of data. Another matrix factorization strategy is to factorize the matrix nuclear norm into the product of two or three small-scale matrices [25]. Liu *et al.* [26] proposed an efficient matrix tri-factorization method for single-view clustering. For multi-view data with M views: $X^{(v)} = [x_1^{(v)}, x_2^{(v)}, \cdots, x_N^{(v)}] \in R^{d^{(v)}\times N}$, Guo *et al.* [27] proposed the rank consistency induced multi-view clustering by matrix tri-factorization. To preserve the local manifold structure, Wang *et al.* [28] develop a structural low-rank matrix factorization method. However, matrix factorization-based clustering methods often suffer from outliers or noise sensitivity [22].

Subspace clustering methods [29] divide the original feature space into low-dimensional subspaces and use different regularization terms to learn a similarity matrix with some specific prior structures. Typical examples are LRR [8], SSC [9] for single-view clustering and the works in [3], [12], [13] for multi-view clustering. Existing LRR-based clustering methods often use the nuclear norm to approximate the rank function to avoid the NP-hard problem. Taking multi-view data as an example, the feature matrix of each view is expressed under the self-representation technique as: $X^{(v)} = X^{(v)}Z^{(v)} + E^{(v)}$, where $Z^{(v)} \in \mathbb{R}^{N\times N}$ is the corresponding self-representation matrix; $E^{(v)}$ denotes outliers or noise. To reveal the common latent subspace shared by multiple views, the LRR-based multi-view subspace clustering methods are developed with their general model is shown in the following formulation:

$$\begin{aligned} &\min_{Z^{(v)}, E^{(v)}} \|Z^{(v)}\|_* + \lambda\|E^{(v)}\|_l \\ &s.t. \quad X^{(v)} = X^{(v)}Z^{(v)} + E^{(v)}, \quad v = 1, \cdots, M, \end{aligned} \tag{2}$$

where $\|\cdot\|_*$ denotes the nuclear norm (i.e., sum of singular values of a matrix); $\|\cdot\|_l$ means coding noise, which is for different types of noise and the sample-specific noise is mainly encoded by $l_{2,1}$ in the clustering task.

LRR-based multi-view subspace clustering methods are mainly divided into matrix representation [11], [30]–[32] and tensor representation [3], [12], [13], [33] methods. The former methods either fuse all self-representation matrices learned

from each view individually or discover one shared self-representation matrix. For example, the work in [11] learned a low-rank and sparse representation for each view while the work in [32] uncovered a shared low-rank representation. With the increase of the number and dimensions of views, it is much important to make full use of view information. The latter methods utilized the tensor representation to obtain high-order correlation among multiple views. For example, prior works [3], [12], [13], [33] constructed a tensor by storing all self-representation matrices with the low-rank constraint. All these above-mentioned methods are based on the matrix nuclear norm or tensor nuclear norm. However, the nuclear norm is a biased estimation of the rank function and computes the time-consuming SVD at each iteration inevitably [34]. Fortunately, the low-rank performance of the nonconvex approximation methods, such as weighted nuclear norm [14], truncated nuclear norm [15], Logdet rank [16], gamma-norm [17] are proven to be superior to the nuclear norm. In particular, the Schatten-p norm [35], [36] is equivalent to the nuclear norm if $p = 1$. When $0 < p < 1$, the Schatten-p norm is a nonconvex approximation of the rank function. However, the introduction of nonconvex functions cannot overcome the high computational complexity for large-scale data. Fortunately, the work in [18] has proved that the Schatten-p norm has an equivalent matrix factorization form when $p = 1/2$, which transforms a large-scale matrix calculation into smaller factor matrix calculations. Therefore, the BiN factorization form of the Schatten-$1/2$ norm could guarantee the performance and efficiency of the low-rank approximation. The factorability of BiN not only reduces the computational complexity, but also makes each subproblem a convex optimization problem, which is defined in Section III.

III. Notations and Preliminaries

In this section, we give the definition of the nonconvex Schatten-p norm and its BiN factorization form, as well as the relevant knowledge of the tensor nuclear norm. The nonconvex Schatten-p norm ($0 < p < 1$) of matrix $X \in R^{m\times n}$ is defined as

$$\|X\|_{S_p} \triangleq \Big(\sum_{i=1}^{n} \sigma_i^p(X)\Big)^{1/p}, \tag{3}$$

where $\sigma_i(X)$ represents the i-th singular value of X. The work in [37] has proved that when the rank number is larger, the nonconvex Schatten-p norm ($0 < p < 1$) is better than the nuclear norm. To further reduce the computational complexity, we introduce the BiN [18] of Schatten-p norm ($p = 1/2$) to explore the global low-rank structure information of multiple views.

Definition 1. *[18] For any matrix $X \in R^{m\times n}$ with $rank(X) = r \le k$, we can factorize it into two smaller matrices $U \in R^{m\times k}$ and $V \in R^{n\times k}$ such that $X = UV^T$. Then the BiN of X is defined as*

$$\|X\|_{BiN} \triangleq \min_{X=UV^T} \|U\|_*\|V\|_* \tag{4}$$

Property 1. *[18] For any matrix $X \in R^{m\times n}$ with $rank(X) = r \le k$, the following holds:*

$$\begin{aligned}\|X\|_{BiN} &= \min_{X=UV^T} \|U\|_*\|V\|_* = \min_{X=UV^T} \frac{\|U\|_*^2 + \|V\|_*^2}{2} \\ &= \min_{X=UV^T} \Big(\frac{\|U\|_* + \|V\|_*}{2}\Big)^2 = \|X\|_{S_{1/2}}.\end{aligned} \tag{5}$$

For a third-order tensor $\mathcal{X} \in \mathcal{R}^{n_1\times n_2\times n_3}$, the tensor nuclear norm based on tensor SVD (t-SVD) is defined as:

$$\|\mathcal{X}\|_{\circledast} = \sum_{i=1}^{min(n_1,n_2)} \sum_{k=1}^{n_3} \bar{\mathcal{S}}(i,i,k), \tag{6}$$

where $\bar{\mathcal{S}}$ is the Fourier transform of $\mathcal{S}$ along the third dimension, denoted as $\bar{\mathcal{S}} = \mathbf{fft}(\mathcal{S}, [], 3)$. $\mathcal{S}$ denotes an f-diagonal tensor obtained by the t-SVD.

Definition 2. *(f-Diagonal Tensor): A tensor is called f-diagonal if each of its frontal slices is diagonal matrix.*

Definition 3. *(t-SVD): Let $\mathcal{X} \in \mathcal{R}^{n_1\times n_2\times n_3}$, its t-SVD is defined as*

$$\mathcal{X} = \mathcal{A} * \mathcal{S} * \mathcal{B}^T, \tag{7}$$

where $\mathcal{A} \in \mathcal{R}^{n_1\times n_1\times n_3}$ and $\mathcal{B} \in \mathcal{R}^{n_2\times n_2\times n_3}$ are orthogonal tensors, $\mathcal{S} \in \mathcal{R}^{n_1\times n_2\times n_3}$ denotes an f-diagonal tensor.

IV. The Proposed BTMSC

Most current methods obtain the self-representation matrix $Z^{(v)}$ through the rank approximation learning based on nuclear norm in Eq. (2) and then use the similarity matrix $C = \frac{1}{M}\sum_{v=1}^{M}(|Z^{(v)}| + |Z^{(v)T}|)$ as the input of spectral clustering to obtain the clustering result. However, the definition and solution of nuclear norm suffer from the difficulties in biased estimation and high computational complexity. To address the above two limitations, we consider the nonconvex function *i.e.*, nonconvex Schatten-p norm and its BiN factorization form instead of the nuclear norm to approximate the rank function. The model using the nonconvex Schatten-p norm instead of the nuclear norm in Eq. (2) is as follows:

$$\begin{aligned}\min_{Z^{(v)},E^{(v)}} \ & \|Z^{(v)}\|_{S_p}^p + \lambda\|E^{(v)}\|_l \\ s.t. \ & X^{(v)} = X^{(v)}Z^{(v)} + E^{(v)}, \\ & Z^{(v)} = U^{(v)}V^{(v)T}, \ v = 1, \cdots, M.\end{aligned} \tag{8}$$

According to Property 1, $\|Z^{(v)}\|_{S_{1/2}}^{1/2}$ is equivalent to $\|Z^{(v)}\|_{BiN}^{1/2}$. To reduce the computational complexity, Eq. (8) can be transformed into the following BiN-based model:

$$\begin{aligned}\min_{Z^{(v)},E^{(v)},U^{(v)},V^{(v)}} \ & \frac{1}{2}\big(\|U^{(v)}\|_* + \|V^{(v)}\|_*\big) + \lambda\|E^{(v)}\|_l \\ s.t. \ & X^{(v)} = X^{(v)}Z^{(v)} + E^{(v)}, \\ & Z^{(v)} = U^{(v)}V^{(v)T}, \ v = 1, \cdots, M,\end{aligned} \tag{9}$$

where $\min_{Z^{(v)}=U^{(v)}V^{(v)T}} \frac{1}{2}\big(\|U^{(v)}\|_* + \|V^{(v)}\|_*\big) = \|Z^{(v)}\|_{BiN}^{1/2}$, $U^{(v)} \in R^{N\times k}$ and $V^{(v)} \in R^{N\times k}$, k is the upper bounding of the $rank(Z^{(v)})$.

Inspired by the great success of tensor representation methods [3], [12], [13], we consider tensorizing all $Z^{(v)}$ into tensor

$\mathcal{Z} \in R^{N\times N\times M}$. The BiN factorizes the self-representation tensor $\mathcal{Z}$ into $\mathcal{U} * \mathcal{V}^T$, which not only describes the low-rank property but also further reduces the computational complexity by optimizing the tensors with small dimensions. Tensors $\mathcal{U}$ and $\mathcal{V}$ are with dimensions as $N\times k\times M$, where k is the upper bounding of the $rank(\mathcal{Z})$. The proposed BTMSC model is shown as follows:

$$\begin{aligned}\min_{\mathcal{U},\mathcal{V},\mathcal{Z},E_1,\mathcal{E}_2} \ & \frac{1}{2}\big(\|\mathcal{U}\|_{\circledast}+\|\mathcal{V}\|_{\circledast}\big)+\lambda_1\|E_1\|_{2,1}+\lambda_2\|\mathcal{E}_2\|_F^2\\ s.t. \quad & X^{(v)}=X^{(v)}Z^{(v)}+E_1^{(v)}, v=1,\cdots,M,\\ & \mathcal{Z}=\mathcal{U}*\mathcal{V}^T+\mathcal{E}_2,\\ & E_1=\left[E_1^{(1)};E_1^{(2)};\cdots;E_1^{(M)}\right],\\ & \mathcal{Z}=\Phi(Z^{(1)},Z^{(2)},\cdots,Z^{(M)}),\end{aligned} \tag{10}$$

where $\|\cdot\|_{\circledast}$ denotes the tensor nuclear norm in Eq. (6). λ_1 and λ_2 are the trade-off parameters. E_1 represents noise embedding in multi-view features, which is yielded by vertically concatenating all noise matrix $E_1^{(v)}$ of views. Tensor $\mathcal{E}_2$ is the error term between $\mathcal{Z}$ and $\mathcal{U}*\mathcal{V}^T$. $Z^{(v)}$ represents the noise-free matrix and is constructed as $\mathcal{Z}$ by operator $\Phi(\cdot)$.

- BTMSC reduces the computational complexity by $\mathcal{Z}=\mathcal{U}*\mathcal{V}^T+\mathcal{E}_2$, and uses tensor nuclear norm based on t-SVD to describe the low-rank property of $\mathcal{U}$ and $\mathcal{V}$. The tensor rank is the sum of the tensor singular values in the Fourier domain. Unlike existing tensor-based multi-view subspace clustering methods [3], [13], [17] which imposed the low-rank constraint on tensor $\mathcal{Z}$ with size of $N\times N\times M$, our BTMSC imposes the tensor nuclear norm on $\mathcal{U}$ and $\mathcal{V}$ with size of $N\times k\times M$.
- Eq. (10) models the sample-specific corruptions noise term E_1 with $l_{2,1}$ norm ($\sum_i\sqrt{\sum_j x_{ij}^2}$) to force columns of E_1 to be closer to zero.
- Tensor $\mathcal{E}_2$ is constrained by Frobenius norm to satisfy the minimum loss condition according to Eq. (1).

After obtaining the minimizer $Z^{(v)}$, we use the similarity matrix C to obtain the clustering results of the original data $X^{(v)}$.

We introduce auxiliary variables $\mathcal{M}$ and $\mathcal{N}$ into Eq. (10) to make variables $\mathcal{U}$ and $\mathcal{V}$ separable, leading to the following optimization model:

$$\begin{aligned}\min_{\mathcal{M},\mathcal{N},\mathcal{U},\mathcal{V},\mathcal{Z},E_1,\mathcal{E}_2} \ & \frac{1}{2}\big(\|\mathcal{M}\|_{\circledast}+\|\mathcal{N}\|_{\circledast}\big)+\lambda_1\|E_1\|_{2,1}+\lambda_2\|\mathcal{E}_2\|_F^2\\ s.t. \quad & X^{(v)}=X^{(v)}Z^{(v)}+E_1^{(v)}, v=1,\cdots,M,\\ & \mathcal{Z}=\mathcal{U}*\mathcal{V}^T+\mathcal{E}_2,\\ & E_1=\left[E_1^{(1)};E_1^{(2)};\cdots;E_1^{(M)}\right],\\ & \mathcal{Z}=\Phi(Z^{(1)},Z^{(2)},\cdots,Z^{(M)}),\\ & \mathcal{U}=\mathcal{M},\ \mathcal{V}=\mathcal{N}.\end{aligned} \tag{11}$$

Meanwhile, we observe that it is not jointly convex when updating all variables simultaneously. Thus, we solve Eq. (11) by the alternating direction method of multipliers, and the corresponding Lagrangian function is shown as follows:

$$\begin{aligned}\mathcal{L}(\mathcal{M},\mathcal{N},\mathcal{U},\mathcal{V},\mathcal{Z},E_1,\mathcal{E}_2) = & \frac{1}{2}\big(\|\mathcal{M}\|_{\circledast}+\|\mathcal{N}\|_{\circledast}\big)+\lambda_1\|E_1\|_{2,1}\\ & +\lambda_2\|\mathcal{E}_2\|_F^2+\frac{\rho}{2}\|\mathcal{Z}-\mathcal{U}*\mathcal{V}^T-\mathcal{E}_2+\frac{\mathcal{Y}_2}{\rho}\|_F^2\\ & +\sum_{v=1}^{M}\big(\frac{\rho}{2}\|X^{(v)}-X^{(v)}Z^{(v)}-E_1^{(v)}+\frac{Y_1^{(v)}}{\rho}\|_F^2\big)\\ & +\frac{\rho}{2}\|\mathcal{U}-\mathcal{M}+\frac{\mathcal{Y}_3}{\rho}\|_F^2+\frac{\rho}{2}\|\mathcal{V}-\mathcal{N}+\frac{\mathcal{Y}_4}{\rho}\|_F^2,\end{aligned} \tag{12}$$

where the augmented Lagrangian function is constructed from the objective function and the penalty term with the l_2 norm constraint. Y_1, $\mathcal{Y}_2$, $\mathcal{Y}_3$ and $\mathcal{Y}_4$ represent the Lagrange multipliers; ρ denotes the penalty parameter.

We solve Eq. (12) according to an iterative mechanism. Specifically, we can update each variable while keeping others fixed.

$\mathcal{M}$ and $\mathcal{N}$-subproblems: We first fix variables $\mathcal{N},\mathcal{U},\mathcal{V},\mathcal{Z},E_1,\mathcal{E}_2$ and solve the optimization problem of $\mathcal{M}$ by Eq. (13).

$$\mathcal{M}_{t+1}=\arg\min_{\mathcal{M}}\frac{1}{2}\|\mathcal{M}\|_{\circledast}+\frac{\rho_t}{2}\|\mathcal{U}_t-\mathcal{M}+\frac{\mathcal{Y}_{3t}}{\rho_t}\|_F^2. \tag{13}$$

Similar to the $\mathcal{M}$-subproblem, the $\mathcal{N}$-subproblem is as follows:

$$\mathcal{N}_{t+1}=\arg\min_{\mathcal{N}}\frac{1}{2}\|\mathcal{N}\|_{\circledast}+\frac{\rho_t}{2}\|\mathcal{V}_t-\mathcal{N}+\frac{\mathcal{Y}_{4t}}{\rho_t}\|_F^2. \tag{14}$$

Due to $M<k$, we rotate $\mathcal{M}\in\mathcal{R}^{N\times k\times M}$ and $\mathcal{N}\in\mathcal{R}^{N\times k\times M}$ to $\tilde{\mathcal{M}}\in\mathcal{R}^{N\times M\times k}$ and $\tilde{\mathcal{N}}\in\mathcal{R}^{N\times M\times k}$ to reduce the computational complexity of SVD. Then, the optimization problems in Eq. (13) and Eq. (14) can be solved by the tensor tubal-shrinkage operator [3], as shown in Eq. (15) and Eq. (16).

$$\tilde{\mathcal{M}}_{t+1}=\mathcal{C}_{\frac{k}{2\rho_t}}(\mathcal{F}_{\mathcal{M}})=\mathcal{A}_{\mathcal{M}}*\mathcal{C}_{\frac{k}{2\rho_t}}(\mathcal{S}_{\mathcal{M}})*\mathcal{B}_{\mathcal{M}}^T, \tag{15}$$

$$\tilde{\mathcal{N}}_{t+1}=\mathcal{C}_{\frac{k}{2\rho_t}}(\mathcal{F}_{\mathcal{N}})=\mathcal{A}_{\mathcal{N}}*\mathcal{C}_{\frac{k}{2\rho_t}}(\mathcal{S}_{\mathcal{N}})*\mathcal{B}_{\mathcal{N}}^T, \tag{16}$$

where $\mathcal{F}_{\mathcal{M}}=\mathcal{U}_t+\frac{\mathcal{Y}_{3t}}{\rho_t}=\mathcal{A}_{\mathcal{M}}*\mathcal{S}_{\mathcal{M}}*\mathcal{B}_{\mathcal{M}}^T$ and $\mathcal{F}_{\mathcal{N}}=\mathcal{V}_t+\frac{\mathcal{Y}_{4t}}{\rho_t}=\mathcal{A}_{\mathcal{N}}*\mathcal{S}_{\mathcal{N}}*\mathcal{B}_{\mathcal{N}}^T$. $\mathcal{C}_{\frac{k}{2\rho_t}}(\mathcal{S}_{\mathcal{M}})=\mathcal{S}_{\mathcal{M}}*\mathbf{diag}\{\max(\bar{\mathcal{S}}_{\mathcal{M}}^{(j)}(i,i)-\frac{k}{2\rho_t},0)\}$ and $\mathcal{C}_{\frac{k}{2\rho_t}}(\mathcal{S}_{\mathcal{N}})=\mathcal{S}_{\mathcal{N}}*\mathbf{diag}\{\max(\bar{\mathcal{S}}_{\mathcal{N}}^{(j)}(i,i)-\frac{k}{2\rho_t},0)\}$. $\bar{\mathcal{S}}_{\mathcal{M}}=fft(\mathcal{S}_{\mathcal{M}},[],3)$, $\bar{\mathcal{S}}_{\mathcal{N}}=fft(\mathcal{S}_{\mathcal{N}},[],3)$ and $j=1,2,\cdots,k$.

$\mathcal{Z}$-subproblem: To solve subproblem $\mathcal{Z}$, variables $\mathcal{M},\mathcal{N},\mathcal{U},\mathcal{V},E_1,\mathcal{E}_2$ are fixed in the augmented Lagrangian function. We find that all terms are based on the Frobenius norm with respect to $\mathcal{Z}$, which makes all frontal slices of $\mathcal{Z}$ independent. Thus, we can solve $\mathcal{Z}$ slice-by-slice from the frontal side by minimizing Eq. (17):

$$\begin{aligned}Z_{t+1}^{(v)}=\arg\min_{Z^{(v)}}\ & \frac{\rho_t}{2}\|X^{(v)}-X^{(v)}Z^{(v)}-E_{1t}^{(v)}+\frac{Y_{1t}^{(v)}}{\rho_t}\|_F^2\\ & +\frac{\rho_t}{2}\|Z^{(v)}-U_t^{(v)}V_t^{(v)T}-E_{2t}^{(v)}+\frac{Y_{2t}^{(v)}}{\rho_t}\|_F^2.\end{aligned} \tag{17}$$

Different from the solutions of Eq. (13) and Eq. (14), by differentiating the Eq. (17) with respect to $Z^{(v)}$ and making it equal to zero, the closed-form solution for $Z_{t+1}^{(v)}$ can be obtained as follows:

$$Z_{t+1}^{(v)} = \Big(I + X^{(v)^T} X^{(v)}\Big)^{-1} \times \big(X^{(v)^T} X^{(v)} - X^{(v)^T} E_{1t}^{(v)} + X^{(v)^T} \frac{Y_{1t}^{(v)}}{\rho_t} + U_t^{(v)} V_t^{(v)^T} + E_{2t}^{(v)} + \frac{Y_{2t}^{(v)}}{\rho_t}\big), \tag{18}$$

where I denotes the identity matrix with proper size.

$\mathcal{U}$ and $\mathcal{V}$-subproblems: Similarly, the subproblems of updating $\mathcal{U}$ and $\mathcal{V}$ can also be obtained by fixing other variables. We write them as follows:

$$[\mathcal{U}_{t+1}, \mathcal{V}_{t+1}] = \arg\min_{\mathcal{U},\mathcal{V}} \frac{\rho_t}{2} \|\mathcal{Z}_{t+1} - \mathcal{U} * \mathcal{V}^T - \mathcal{E}_{2t} + \frac{\mathcal{Y}_{2t}}{\rho_t}\|_F^2 + \frac{\rho_t}{2} \|\mathcal{U} - \mathcal{M}_{t+1} + \frac{\mathcal{Y}_{3t}}{\rho_t}\|_F^2 + \frac{\rho_t}{2} \|\mathcal{V} - \mathcal{N}_{t+1} + \frac{\mathcal{Y}_{4t}}{\rho_t}\|_F^2. \tag{19}$$

Same to $\mathcal{Z}$-subproblem, all terms are based on the Frobenius norm with respect to $\mathcal{U}$ and $\mathcal{V}$, which makes all frontal slices of $\mathcal{U}$ and $\mathcal{V}$ independent. It is also beneficial to parallel compute $\mathcal{U}$ and $\mathcal{V}$ to improve the running efficiency. Subsequently, we can solve $\mathcal{U}$ and $\mathcal{V}$ slice-by-slice from the frontal side, whose v-th slice is updated by:

$$[U_{t+1}^{(v)}, V_{t+1}^{(v)}] = \arg\min_{U^{(v)},V^{(v)}} \frac{\rho_t}{2} \big(\|Z_{t+1}^{(v)} - U^{(v)} V^{(v)T} - E_{2t}^{(v)} + \frac{Y_{2t}^{(v)}}{\rho_t}\|_F^2 + \|U^{(v)} - M_{t+1}^{(v)} + \frac{Y_{3t}^{(v)}}{\rho_t}\|_F^2 + \|V^{(v)} - N_{t+1}^{(v)} + \frac{Y_{4t}^{(v)}}{\rho_t}\|_F^2\big) \tag{20}$$

By setting the derivative of Eq. (20) to zero, we obtain the closed-form solutions of $U^{(v)}$ and $V^{(v)}$ as follows:

$$U_{t+1}^{(v)} = \Big(\big(Z_{t+1}^{(v)} - E_{2t}^{(v)} + \frac{Y_{2t}^{(v)}}{\rho_t}\big) V_t^{(v)} + M_{t+1}^{(v)} - \frac{Y_{3t}^{(v)}}{\rho_t}\Big) \times \Big(V_t^{(v)T} V_t^{(v)} + I\Big)^{-1}, \tag{21}$$

$$V_{t+1}^{(v)} = \Big(\big(Z_{t+1}^{(v)} - E_{2t}^{(v)} + \frac{Y_{2t}^{(v)}}{\rho_t}\big)^T U_{t+1}^{(v)} + N_{t+1}^{(v)} - \frac{Y_{4t}^{(v)}}{\rho_t}\Big) \times \Big(U_{t+1}^{(v)T} U_{t+1}^{(v)} + I\Big)^{-1}. \tag{22}$$

E_1 and $\mathcal{E}_2$-subproblems: According to Eq. (12), with other variables fixed, we write the corresponding subproblems of E_1 and $\mathcal{E}_2$ as follows:

$$E_{1t+1} = \arg\min_{E_1} \lambda_1 \|E_1\|_{2,1} + \frac{\rho_t}{2} \|E_1 - D_t\|_F^2, \tag{23}$$

$$\mathcal{E}_{2t+1} = \arg\min_{\mathcal{E}_2} \lambda_2 \|\mathcal{E}_2\|_F^2 + \frac{\rho_t}{2} \|\mathcal{Z}_{t+1} - \mathcal{U}_{t+1}\mathcal{V}_{t+1}^T - \mathcal{E}_2 + \frac{\mathcal{Y}_{2t}}{\rho_t}\|_F^2), \tag{24}$$

where D_t is constructed by vertically concatenating $\{X^{(v)} - X^{(v)} Z_{t+1}^{(v)} + \frac{Y_{1t}^{(v)}}{\rho_t}\}$ along the column. The j-th column of optimal solution E_{1t+1} can be obtained by Eq. (25).

$$E_{1t+1}(:,j) = \begin{cases} \frac{\|D_t(:,j)\|_2 - \frac{\lambda_1}{\rho_t}}{\|D_t(:,j)\|_2} D_t(:,j), & if \quad \frac{\lambda_1}{\rho_t} < \|D_t(:,j)\|_2; \\ 0, & otherwise. \end{cases} \tag{25}$$

The closed-form solution of $\mathcal{E}_{2t+1}$ is obtained by setting the derivative with respect to $\mathcal{E}_2$ to zero as shown in Eq. (26).

$$\mathcal{E}_{2t+1} = \big(2\lambda_2 + \rho_t\big)^{-1} \times \rho_t (\mathcal{Z}_{t+1} - \mathcal{U}_{t+1} * \mathcal{V}_{t+1}^T + \frac{\mathcal{Y}_{2t}}{\rho_t}\big). \tag{26}$$

The Lagrange multipliers $Y_1, \mathcal{Y}_2, \mathcal{Y}_3, \mathcal{Y}_4$ and penalty parameter ρ are updated by Eq. (27)-Eq. (31).

$$Y_{1t+1}^{(v)} = Y_{1t}^{(v)} + \rho_t (X^{(v)} - X^{(v)} Z_{t+1}^{(v)} - E_{1t+1}^{(v)}); \tag{27}$$

$$\mathcal{Y}_{2t+1} = \mathcal{Y}_{2t} + \rho_t (\mathcal{Z}_{t+1} - \mathcal{U}_{t+1} * \mathcal{V}_{t+1}^T - \mathcal{E}_{2t+1}); \tag{28}$$

$$\mathcal{Y}_{3t+1} = \mathcal{Y}_{3t} + \rho_t (\mathcal{U}_{t+1} - \mathcal{M}_{t+1}); \tag{29}$$

$$\mathcal{Y}_{4t+1} = \mathcal{Y}_{4t} + \rho_t (\mathcal{V}_{t+1} - \mathcal{N}_{t+1}); \tag{30}$$

$$\rho_{t+1} = \min\{\beta * \rho_t, \rho_{max}\}, \tag{31}$$

where β is used to accelerate the convergence speed. Algorithm 1 shows the whole process of the proposed BTMSC.

Algorithm 1: BTMSC for multi-view subspace clustering

Input: multi-view features $\{X^{(v)}, v = 1, 2, \cdots, M\}$; parameter λ_1, λ_2, k;

Initialize: $\mathcal{M}$, $\mathcal{N}$, $\mathcal{E}_2$, E_1, $\mathcal{Y}_2$, $\mathcal{Y}_3$, $\mathcal{Y}_4$ initialized to $\mathbf{0}$; $Z^{(v)} = V^{(v)} = Y_1^{(v)} = \mathbf{0}$, $U^{(v)} = I_{N\times k}$, $v = 1, \cdots, M$; $\rho = 10^{-5}$, $\beta = 2$, $\rho_{max} = 10^{10}$, $tol = 10^{-7}$;

while not converged **do**
  Update $\mathcal{M}_{t+1}$ according to Eq. (15);
  Update $\mathcal{N}_{t+1}$ according to Eq. (16);
  for $v = 1$ to M **do**
    Update $Z_{t+1}^{(v)}$ according to Eq. (18);
    Update $U_{t+1}^{(v)}$ and $V_{t+1}^{(v)}$ according to Eq. (21) and Eq. (22);
  end for
  Update E_{1t+1} and $\mathcal{E}_{2t+1}$ according to Eq. (25) and Eq. (26);
  Update $Y_{1t+1}^{(v)}, \mathcal{Y}_{2t+1}, \mathcal{Y}_{3t+1}, \mathcal{Y}_{4t+1}$ and ρ_{t+1} according to Eqs. (27), (28), (29), (30) and (31);
  Check the convergence conditions:
  $$\max\left\{ \begin{array}{l} \|X^{(v)} - X^{(v)} Z^{(v)} - E_1^{(v)}\|_\infty, \\ \|\mathcal{Z} - \mathcal{U} * \mathcal{V}^{\mathcal{T}} - \mathcal{E}_2\|_\infty \end{array} \right\} \le tol$$
end while

Output: $Z_{t+1}^{(v)}$.

V. Convergence Analysis

We first introduce several lemmas and then show the theoretical convergence of the proposed BTMSC.

Lemma 1. *The sequences $\{\mathcal{M}_t \mathcal{N}_t, \mathcal{U}_t, \mathcal{V}_t, \mathcal{Z}_t, E_{1t}, \mathcal{E}_{2t}\}$ and $Y_t = \{Y_{1t}, \mathcal{Y}_{2t}, \mathcal{Y}_{3t}, \mathcal{Y}_{4t}\}$ are bounded if $\mathcal{Y}_{2t}$ is bounded and $\sum_{t=1}^{\infty} (\rho_t + \rho_{t-1})/(\rho_{t-1})^2 < \infty$.*

Proof. The optimal $\mathcal{M}_{t+1}, \mathcal{N}_{t+1}, E_{1t+1}$ satisfy the first-order optimality condition, *i.e.*,

TABLE I
COMPLEXITY OF THE COMPARED METHODS

Method	DiMSC	LT-MSC	ECMSC	tSVDMSC	ETLMSC	LMSC	BTMSC
Cost	$\mathcal{O}(TMN^3)$	$\mathcal{O}(TMN^3)$	$\mathcal{O}(T+M)N^3$	$\mathcal{O}(MN^3+TMN^2log(N))$	$\mathcal{O}(MN^3+TMN^2log(N))$	$\mathcal{O}(TMN^3)$	$\mathcal{O}((MN^3+2TMNklog(k))$

$$
\begin{aligned}
&0 \in \partial_{\mathcal{M}} \mathcal{L}_{\rho_t}(\mathcal{M}_{t+1}, \mathcal{N}_{t+1}, \mathcal{U}_{t+1}, \mathcal{V}_{t+1}, \mathcal{Z}_{t+1}, E_{1t+1}, \mathcal{E}_{2t+1}) \\
&= \frac{1}{2}\partial_{\mathcal{M}}\|\mathcal{M}\|_{\circledast} - \mathcal{Y}_{3t} - \rho_t(\mathcal{U}_{t+1} - \mathcal{M}_{t+1}) \\
&= \frac{1}{2}\partial_{\mathcal{M}}\|\mathcal{M}\|_{\circledast} - \mathcal{Y}_{3t+1}.
\end{aligned} \tag{32}
$$

$$
\begin{aligned}
&0 \in \partial_{\mathcal{N}} \mathcal{L}_{\rho_t}(\mathcal{M}_{t+1}, \mathcal{N}_{t+1}, \mathcal{U}_{t+1}, \mathcal{V}_{t+1}, \mathcal{Z}_{t+1}, E_{1t+1}, \mathcal{E}_{2t+1}) \\
&= \frac{1}{2}\partial_{\mathcal{N}}\|\mathcal{N}\|_{\circledast} - \mathcal{Y}_{4t} - \rho_t(\mathcal{V}_{t+1} - \mathcal{N}_{t+1}) \\
&= \frac{1}{2}\partial_{\mathcal{N}}\|\mathcal{N}\|_{\circledast} - \mathcal{Y}_{4t+1}.
\end{aligned} \tag{33}
$$

$$
\begin{aligned}
&0 \in \partial_{E_1} \mathcal{L}_{\rho_t}(\mathcal{M}_{t+1}, \mathcal{N}_{t+1}, \mathcal{U}_{t+1}, \mathcal{V}_{t+1}, \mathcal{Z}_{t+1}, E_{1t+1}, \mathcal{E}_{2t+1}) \\
&= \partial_{E_1}(\lambda_1\|E_1\|_{2,1}) - \sum_{v=1}^{M}\big(Y_{1t}^{(v)} + \rho_t(X^{(v)} - X^{(v)}Z_{t+1}^{(v)} - E_{1t+1}^{(v)}\big) \\
&= \partial_{E_1}(\lambda_1\|E_1\|_{2,1}) - \sum_{v=1}^{M} Y_{1t+1}^{(v)}
\end{aligned} \tag{34}
$$

According to Lemma 2, we can obtain $\|\mathcal{Y}_{3t+1}\|_2 \leq \frac{1}{2}$, $\|\mathcal{Y}_{4t+1}\|_2 \leq \frac{1}{2}$.

Lemma 2. *Let H be a real Hilbert space endowed with an inner product, and a nuclear norm, and $y \in \partial\|x\|$. Then the dual norm of the nuclear norm $\|y\|_2 = 1$ if $x \neq 0$, and $\|y\|_2 \leq 1$ if $x = 0$.*

The $\partial_{E_1}(\lambda_1\|E_1\|_{2,1})$ is bounded according to Eq. (35). Thus, $Y_{1t+1}^{(v)}$ is bounded.

$$
\partial_{E_1}(\|E_1\|_{2,1}) = \begin{cases} 0, & if\ \|E_{1t+1}(:,l)\|_2 = 0 \\ \frac{E_{1t+1}(:,l)}{\|E_{1t+1}(:,l)\|_2}, & otherwise. \end{cases} \tag{35}
$$

By the iterative mechanism of Lagrangian function, we have

$$
\begin{aligned}
&\mathcal{L}_{\rho_t}(\mathcal{M}_{t+1}, \mathcal{N}_{t+1}, \mathcal{U}_{t+1}, \mathcal{V}_{t+1}, \mathcal{Z}_{t+1}, E_{1t+1}, \mathcal{E}_{2t+1}) \\
&\leq \mathcal{L}_{\rho_t}(\mathcal{M}_t, \mathcal{N}_t, \mathcal{U}_t, \mathcal{V}_t, \mathcal{Z}_t, E_{1t}, \mathcal{E}_{2t}) \\
&= \mathcal{L}_{\rho_{t-1}}(\mathcal{M}_t, \mathcal{N}_t, \mathcal{U}_t, \mathcal{V}_t, \mathcal{Z}_t, E_{1t}, \mathcal{E}_{2t}) \\
&+ \frac{\rho_t + \rho_{t-1}}{2(\rho_{t-1})^2}\|Y_t - Y_{t-1}\|_F^2.
\end{aligned} \tag{36}
$$

Since $\sum_{t=1}^{\infty}(\rho_t + \rho_{t-1})/(\rho_{t-1})^2 < \infty$. Then the sequence $\{\mathcal{M}_t, \mathcal{N}_t, \mathcal{U}_t, \mathcal{V}_t, \mathcal{Z}_t, E_{1t}, \mathcal{E}_{2t}, Y_{1t}^{(v)}, \mathcal{Y}_{2t}, \mathcal{Y}_{3t}, \mathcal{Y}_{4t}\}$ is upper-bounded due to the boundedness of the sequences of all Lagrange multipliers $\{Y_{1t}^{(v)}, \mathcal{Y}_{2t}, \mathcal{Y}_{3t}, \mathcal{Y}_{4t}\}$. □

Theorem 1. *Suppose the sequence $\{\mathcal{M}_t, \mathcal{N}_t, \mathcal{U}_t, \mathcal{V}_t, \mathcal{Z}_t, E_{1t}, \mathcal{E}_{2t}, Y_{1t}^{(v)}, \mathcal{Y}_{2t}, \mathcal{Y}_{3t}, \mathcal{Y}_{4t}\}$ be generated by Algorithm 1 and $\{\mathcal{M}^*, \mathcal{N}^*, \mathcal{U}^*, \mathcal{V}^*, \mathcal{Z}^*, E_1^*, \mathcal{E}_2^*, Y_1^{(v)*}, \mathcal{Y}_2^*, \mathcal{Y}_3^*, \mathcal{Y}_4^*\}$ be the accumulation point. Then, $\{\mathcal{M}^*, \mathcal{N}^*, \mathcal{U}^*, \mathcal{V}^*, \mathcal{Z}^*, E_1^*, \mathcal{E}_2^*, Y_1^{(v)*}, \mathcal{Y}_2^*, \mathcal{Y}_3^*, \mathcal{Y}_4^*\}$ satisfies the first-order Karush-Kuhn-Tucker (KKT) conditions of Eq. (11).*

Proof. According to Lemma 1, the sequence $\{\mathcal{M}_t, \mathcal{N}_t, \mathcal{U}_t, \mathcal{V}_t, \mathcal{Z}_t, E_{1t}, \mathcal{E}_{2t}, Y_{1t}^{(v)}, \mathcal{Y}_{2t}, \mathcal{Y}_{3t}, \mathcal{Y}_{4t}\}$ is bounded. By Bolzano–Weierstrass theorem, the sequence has at least one accumulation point which is denoted as $\{\mathcal{M}^*, \mathcal{N}^*, \mathcal{U}^*, \mathcal{V}^*, \mathcal{Z}^*, E_1^*, \mathcal{E}_2^*, Y_1^{(v)*}, \mathcal{Y}_2^*, \mathcal{Y}_3^*, \mathcal{Y}_4^*\}$. Assume that $\{\mathcal{M}_t, \mathcal{N}_t, \mathcal{U}_t, \mathcal{V}_t, \mathcal{Z}_t, E_{1t}, \mathcal{E}_{2t}, Y_{1t}^{(v)}, \mathcal{Y}_{2t}, \mathcal{Y}_{3t}, \mathcal{Y}_{4t}\}$ converges to $\{\mathcal{M}^*, \mathcal{N}^*, \mathcal{U}^*, \mathcal{V}^*, \mathcal{Z}^*, E_1^*, \mathcal{E}_2^*, Y_1^{(v)*}, \mathcal{Y}_2^*, \mathcal{Y}_3^*, \mathcal{Y}_4^*\}$. According to the boundness of $\{Y_{1t}^{(v)}, \mathcal{Y}_{2t}, \mathcal{Y}_{3t}, \mathcal{Y}_{4t}\}$, we obtain

$$
\begin{aligned}
&\lim_{t\to\infty} X^{(v)} - X^{(v)}Z_{t+1}^{(v)} - E_{t+1}^{(v)} = \lim_{t\to\infty}(Y_{1t+1}^{(v)} - Y_{1t}^{(v)})/\rho_t = 0, \\
&\lim_{t\to\infty} \mathcal{Z}_{t+1} - \mathcal{U}_{t+1} * \mathcal{V}_{t+1}^T - \mathcal{E}_{2t+1} = \lim_{t\to\infty}(\mathcal{Y}_{2t+1} - \mathcal{Y}_{2t})/\rho_t = 0, \\
&\lim_{t\to\infty} \mathcal{U}_{t+1} - \mathcal{M}_{t+1} = \lim_{t\to\infty}(\mathcal{Y}_{3t+1} - \mathcal{Y}_{3t})/\rho_t = 0, \\
&\lim_{t\to\infty} \mathcal{V}_{t+1} - \mathcal{N}_{t+1} = \lim_{t\to\infty}(\mathcal{Y}_{4t+1} - \mathcal{Y}_{4t})/\rho_t = 0.
\end{aligned} \tag{37}
$$

Then we have that $X^{(v)} = X^{(v)}Z^{(v)*} - E^{(v)*}$, $\mathcal{Z}^* = \mathcal{U}^* * \mathcal{V}^{T*} - \mathcal{E}^*$, $\mathcal{U}^* = \mathcal{M}^*$, $\mathcal{V}^* = \mathcal{N}^*$.

According to Eqs. (32)-(37), we reach the conclusion that $\{\mathcal{M}^*, \mathcal{N}^*, \mathcal{U}^*, \mathcal{V}^*, \mathcal{Z}^*, E_1^*, \mathcal{E}_2^*, Y_1^{(v)*}, \mathcal{Y}_2^*, \mathcal{Y}_3^*, \mathcal{Y}_4^*\}$ satisfies the KKT conditions. Thus, $\{\mathcal{M}^*, \mathcal{N}^*, \mathcal{U}^*, \mathcal{V}^*, \mathcal{Z}^*, E_1,^* \mathcal{E}_2^*\}$ is a stationary point of the original problem (11). □

Complexity analysis The computational complexity of the proposed BTMSC method mainly comes from the solution of subproblems $U^{(v)}$, $V^{(v)}$, $\mathcal{M}$ and $\mathcal{N}$. The calculation of $U^{(v)}$ and $V^{(v)}$ involves multiplication for $N \times k$ matrix and the inverse for $k \times k$ matrix. Therefore, the computational complexity of $U^{(v)}$ and $V^{(v)}$ is $2(Nk^2 + kN^2 + k^3)$. The solution of $\mathcal{M}$ and $\mathcal{N}$ requires calculating the 3D FFT and 3D inverse FFT for $N \times M \times k$ tensor and k SVDs of $N \times M$ matrices in the Fourier domain. For each iteration, the computational complexity of FFT and inverse FFT is $\mathcal{O}(2MNklog(k))$ and the cost of k SVD for $N \times M$ matrices is $\mathcal{O}(kNM^2)$. According to [3], the computational complexity of the last step *i.e.*, spectral clustering for M matrices is usually $\mathcal{O}(MN^3)$. Since $log(k) > M$ and $N > k$, the total computational complexity of BTMSC is about $\mathcal{O}(2TMNklog(k)) + \mathcal{O}(MN^3)$, where T is the total number of iterations. The computational complexity of the compared methods is recorded in Table I.

VI. EXPERIMENTAL RESULTS

In this section, we experimentally compare the proposed BTMSC method with sixteen state-of-the-art clustering methods on ten real datasets based on the quantitative analysis

TABLE II
DETAILS OF TEN MULTI-VIEW DATASETS

Datasets	Types	View	Dimension	Samples	Clusters
BBCSport	text	2	3183/3203	544	5
UCI-3views	image	3	240/76/6	2000	10
COIL_20	image	3	1024/3304/6750	1440	20
Reuters	text	5	21526/24892/34121/15487/11539	600	6
MSRC-V1	image	5	24/576/512/256/254	210	7
Wikipedia	text	2	128/10	693	10
Scene-15	image	3	1800/1180/1240	4485	15
Caltech101	image	4	2048/4800/3540/1240	8677	101
Mnist	image	3	30/9/30	10000	10
SUNRGBD	image	2	4096/4096	10335	45

TABLE III
PARAMETERS SELECTION RANGE OF COMPARISON METHODS

Method	Parameter 1	Parameter 2	Parameter 3
SSC [9]	[0.01, 10]	-	-
LRR [8]	[0.01, 10]	-	-
RMSC [38]	[0.005, 0.1]	-	-
DiMSC [39]	[0.01, 0.03]	[20, 180]	-
LT-MSC [12]	[0.01, 100]	-	-
ECMSC [40]	[0.1, 0.5]	[0.2, 0.7]	1.2
GMC [41]	Initial value: 1	-	-
LMSC [30]	[0.001, 1000]	-	-
MLAN [42]	[1, 30]	-	-
tSVDMSC [3]	[0.1, 2]	-	-
ETLMSC [13]	[0.0008, 0.01]	-	-
LRTG [43]	[0.001, 100]	[5, 15]	-
GNLTA [17]	[0.001, 0.1]	[1.1, 2]	-
HLR-M^{2}VS [33]	[0.01, 0.2]	[0.1, 0.9]	-
OPMC [44]	-	-	-
SCGL [45]	[0.2, 1.4]	[10, 350]	-

results of evaluation metrics. The six clustering evaluation metrics are Accuracy (ACC), normalized mutual information (NMI), adjusted rand index (ARI), F-score, Precision and Recall. In addition to quantitative analysis, we also perform analysis on parameters selection, numerical convergence and running time. The experimental setting and results are as follows.

A. Datasets and Compared Methods

Datasets: Experiments are conducted on ten real multi-view datasets of text and images. The details of ten multi-view datasets are described in Table II. **BBCSport**[1] is a News stories dataset. It contains 544 documents from the BBC Sports website, which correspond to 5 clusters with 2 views; **UCI-3views**[2] includes 2000 instance with 10 clusters and three types of features including the $240d$ Fourier coefficients, the $76d$ pixel averages and the $6d$ morphological features are explored; **COIL_20**[3] **and Caltech101** are the Generic objects datasets. COIL_20 contains 1440 images with 32×32 pixels, which corresponds to 20 clusters with 3 views. Caltech101 contains 8677 images with 101 clusters and four types of features are explored. **Reuters** is an article written in 5 languages with 5 views. In this paper we follow [17] to randomly sample 100 documents from each class to obtain a dataset with 600 documents. **MSRC-V1** includes 210 images, which correspond to 7 clusters with 5 views. **Wikipedia** contains 693 documents, which correspond to 10 clusters associated with 2 views. **Scene-15**[4] contains 4485 outdoor and indoor scene images from 15 categories and each category has 210 to 410 images. **Mnist** includes 10000 images, which correspond to 10 clusters with 3 views. **SUNRGBD** contains 10335 images, which correspond to 45 clusters associated with 2 views.

Compared methods: We compare the proposed BTMSC with state-of-the-art clustering methods including two single-view methods: **SSC** [9] and **LRR** [8] and fourteen multi-view methods: **RMSC** [38], **DiMSC** [39], **LT-MSC** [12], **ECMSC** [40], **GMC** [41], **LMSC** [30], **MLAN** [42], **tSVDMSC** [3], **ETLMSC** [13], **LRTG** [43], **GNLTA** [17], **HLR-M^{2}VS** [33], **OPMC** [44] and **SCGL** [45]. Among these fourteen multi-view clustering methods, RMSC and LMSC utilize the convex matrix nuclear norm to realize the low-rankness while LT-MSC, tSVDMSC, ETLMSC and HLR-M^{2}VS adopt the convex tensor nuclear norm. LRTG is a one-step low-rank tensor graph learning method. GNLTA is the recently proposed non-convex multi-view subspace clustering method. For ECMSC, three parameters are tuned from interval $[0.1, 1]$, $[0.1, 1]$ and 1.2, respectively. Two parameters of HLR-M^{2}VS are selected from $[0.01, 0.2]$ and $[0.1, 0.9]$, respectively. Two parameters of SCGL are selected from $[0.2, 1.4]$ and $[10, 350]$, respectively. For other compared methods, we set the parameters as in [46].

B. Experimental Results

The clustering results compared with sixteen clustering methods on ten datasets are shown in Tables IV-VIII, in which each row represents the cluster metric value for the corresponding comparison method on a specific dataset. In these tables, the best results are highlighted in bold, and the second best results are underlined.

Observed in Tables IV-VIII, the multi-view clustering methods are overall better than the single-view clustering methods in most cases because the complementary information in different views can facilitate data representation. The top-ranked methods are tensor-based multi-view clustering methods. The reason is that exploring the tensor low-rank structure can preserve the high-order correlation of views. The superior clustering results of tSVDMSC and ETLMSC quantitatively confirm the above statements, since they are direct extensions of LRR and RMSC from matrix optimization to tensor optimization, respectively.

For the ACC metrics, the proposed BTMSC method achieves the most advantageous clustering performance. Specifically, the ACC values of BTMSC on all datasets are improved by 0.2%, 0.1%, 4.3%, 2.8%, 0.9%, 3.2%, 7.4%, 3.3%, 1.3%, 1.1% compared to the second-best method. The above results show the effectiveness of BTMSC. Compared to the second-best methods HLR-M^{2}VS, LRTG, GNLTA, tSVDMSC and ETLMSC, although the proposed BTMSC method is also based on the low-rank tensor approximation, it applies the BiN factorization form of the nonconvex Schatten-p norm as an approximation of the rank function, which avoids the

[1] http://mlg.ucd.ie/datasets/segment.html
[2] http://archive.ics.uci.edu/ml/datasets/Multiple+Features
[3] http://www.cs.columbia.edu/CAVE/software/softlib/
[4] http://www-cvr.ai.uiuc.edu/ponce_grp/data/

TABLE IV
CLUSTERING RESULTS (MEAN±STANDARD DEVIATION) ON *BBCSport* AND *UCI-3views* DATASETS.

Dataset	Method	ACC	NMI	AR	F-score	Precision	Recall
BBCSport	SSC [9]	0.627±0.003	0.534±0.008	0.364±0.007	0.565±0.005	0.427±0.004	0.834±0.004
	LRR [8]	0.836±0.001	0.698±0.002	0.705±0.001	0.776±0.001	0.768±0.001	0.784±0.001
	RMSC [38]	0.826±0.001	0.666±0.001	0.637±0.001	0.719±0.001	0.766±0.001	0.677±0.001
	DiMSC [39]	0.922±0.000	0.785±0.000	0.813±0.000	0.858±0.000	0.846±0.000	0.872±0.000
	LT-MSC [12]	0.460±0.046	0.222±0.028	0.167±0.043	0.428±0.014	0.328±0.028	0.629±0.053
	ECMSC [40]	0.285±0.014	0.027±0.013	0.009±0.011	0.267±0.020	0.244±0.007	0.297±0.045
	GMC [41]	0.807±0.000	0.760±0.000	0.722±0.000	0.794±0.000	0.727±0.000	0.875±0.000
	LMSC [30]	0.847±0.003	0.739±0.001	0.749±0.001	0.810±0.001	0.799±0.001	0.822±0.001
	MLAN [42]	0.721±0.000	0.779±0.000	0.591±0.000	0.714±0.000	0.567±0.000	0.962±0.000
	tSVDMSC [3]	0.879±0.000	0.765±0.000	0.784±0.000	0.834±0.000	0.863±0.000	0.807±0.000
	ETLMSC [13]	0.959±0.086	0.972±0.058	0.949±0.107	0.961±0.081	0.963±0.078	0.960±0.085
	LRTG [43]	0.943±0.005	0.869±0.009	0.840±0.012	0.879±0.010	0.866±0.006	0.892±0.014
	GNLTA [17]	0.980±0.064	0.986±0.043	0.973±0.086	0.979±0.065	0.979±0.067	0.980±0.064
	HLR-M^{2}VS [33]	0.998±0.000	0.993±0.000	0.994±0.000	0.996±0.000	0.996±0.000	0.995±0.064
	OPMC [44]	0.945±0.000	0.842±0.000	0.854±0.000	0.889±0.000	0.885±0.000	0.893±0.000
	SCGL [45]	0.906±0.000	0.778±0.000	0.784±0.000	0.836±0.000	0.849±0.000	0.823±0.000
	BTMSC	**1.000±0.000**	**1.000±0.000**	**1.000±0.000**	**1.000±0.000**	**1.000±0.000**	**1.000±0.000**
UCI-3views	SSC [9]	0.815±0.011	0.840±0.001	0.770±0.005	0.794±0.004	0.747±0.010	0.848±0.004
	LRR [8]	0.871±0.001	0.768±0.002	0.736±0.002	0.763±0.002	0.759±0.002	0.767±0.002
	RMSC [38]	0.915±0.024	0.822±0.008	0.789±0.014	0.811±0.012	0.797±0.017	0.826±0.006
	DiMSC [39]	0.703±0.010	0.772±0.006	0.652±0.006	0.695±0.006	0.673±0.005	0.718±0.007
	LT-MSC [12]	0.803±0.001	0.775±0.001	0.725±0.001	0.753±0.001	0.739±0.001	0.767±0.001
	ECMSC [40]	0.718±0.001	0.780±0.001	0.672±0.001	0.707±0.001	0.660±0.001	0.760±0.001
	GMC [41]	0.736±0.000	0.815±0.000	0.678±0.000	0.713±0.000	0.644±0.000	0.799±0.000
	LMSC [30]	0.893±0.000	0.815±0.000	0.783±0.000	0.805±0.000	0.798±0.000	0.812±0.000
	MLAN [42]	0.874±0.000	0.910±0.000	0.847±0.000	0.864±0.000	0.797±0.000	0.943±0.000
	tSVDMSC [3]	0.955±0.000	0.932±0.000	0.924±0.000	0.932±0.000	0.930±0.000	0.934±0.000
	ETLMSC [13]	0.958±0.078	0.977±0.028	0.953±0.069	0.958±0.062	0.940±0.088	0.980±0.029
	LRTG [43]	0.981±0.000	0.953±0.000	0.957±0.000	0.961±0.000	0.961±0.000	0.962±0.000
	GNLTA [17]	0.981±0.036	0.979±0.012	0.972±0.036	0.975±0.032	0.968±0.052	0.983±0.008
	HLR-M^{2}VS [33]	0.998±0.000	0.995±0.000	0.996±0.000	0.996±0.000	0.996±0.000	0.996±0.000
	OPMC [44]	0.898±0.000	0.821±0.000	0.793±0.000	0.814±0.000	0.812±0.000	0.816±0.000
	SCGL [45]	0.980±0.000	0.953±0.000	0.955±0.000	0.959±0.000	0.960±0.000	0.959±0.000
	BTMSC	**0.999±0.000**	**0.998±0.000**	**0.999±0.000**	**0.999±0.000**	**0.999±0.000**	**0.999±0.000**

biased estimation of the tensor nuclear norm. In addition, the factorization of the BiN improves the computational efficiency while further filtering out noise.

Taking the Caltech101 dataset as an example, ETLMSC and BTMSC respectively introduce the decomposition and factorization strategy to reduce the dimension of the data, and the clustering results are outstanding. However, for all datasets, the low-rank self-representation tensor of BTMSC outperforms the low-rank transition probability tensor of ETLMSC in clustering performance. In Fig. 1, we also plot the variation in the ACC of BTMSC over iterations. We can see that the ACC value of BTMSC increases rapidly in the first 5 iterations, and reaches stable convergence after 15 iterations. Therefore, the proposed BTMSC method can achieve fast convergence.

Overall, the proposed BTMSC method has two advantages. First, BTMSC exhibits the superior low-rank performance of nonconvex function over convex function. Second, BTMSC greatly contributes to the clustring efficiency and performance improvement on multiple datasets through the factorization strategy, which is confirmed by the results in Tables IV-VIII.

C. Parameter Selection

The proposed method involves balance parameters λ_1, λ_2 for low-rankess and sparsity and parameter k for tuning the dimension of the tensor factorization. During the experiment, we select their values from range: $\lambda_1 = [0.01 : 0.01 : 0.1]$, $\lambda_2 = [0.1 : 0.1 : 1.0]$, $k = [10 : 5 : 100]$. Fig. 2 shows the influence of the three parameters λ_1, λ_2 and k on the ACC value, where the first row is the value of ACC when k is fixed and λ_1, λ_2 traverses all parameters. The second row is the result of ACC when λ_1, λ_2 are fixed and k takes different values. The results show that UCI-3views and COIL_20 datasets are insensitive to parameters λ_1 and λ_2. For the BBCSport and Reuters datasets, when parameter λ_1 takes 0.03 to 0.1 and λ_2 takes 0.4 to 1, the ACC value is better. According to Definition 1, $k \geq rank(\mathcal{Z})$. For selecting an appropriate k, we use a rank estimation procedure to estimate rank following reference [18]. First, we compute the L largest singular values of the input data (usually $L = 100$), and then determine the number of clusters by the basic spectral gap technique. We finally obtain the estimated $rank(\mathcal{Z})$ of BBCSport, UCI-3views, COIL_20 and Reuters datasets to be 30, 20, 20, 10. From Fig. 2, it can be seen that the proposed BTMSC obtains lower ACC when k is smaller than 30, 20, 20, 10, respectively, which corresponds exactly to our estimated rank. When $k \geq 30$, the variation range of ACC are 0.002, 0.005, 0.05, 0.005 and the clustering results are relatively stable for BBCSport, UCI-3views, COIL_20 and Reuters datasets, respectively. In general, the selection range of the above parameters can obtain reasonable clustering results.

TABLE V
CLUSTERING RESULTS (MEAN±STANDARD DEVIATION) ON *COIL_20* AND *Reuters* DATASETS.

Dataset	Method	ACC	NMI	AR	F-score	Precision	Recall
COIL_20	SSC [9]	0.803±0.022	0.935±0.009	0.798±0.022	0.809±0.013	0.734±0.027	0.804±0.028
	LRR [8]	0.761±0.003	0.829±0.006	0.720±0.020	0.734±0.006	0.717±0.003	0.751±0.002
	RMSC [38]	0.685±0.045	0.800±0.017	0.637±0.044	0.656±0.042	0.620±0.057	0.698±0.026
	DiMSC [39]	0.778±0.022	0.846±0.002	0.732±0.005	0.745±0.005	0.739±0.007	0.751±0.003
	LT-MSC [12]	0.804±0.011	0.860±0.002	0.748±0.004	0.760±0.007	0.741±0.009	0.776±0.006
	ECMSC [40]	0.782±0.001	0.942±0.001	0.781±0.001	0.794±0.001	0.695±0.002	0.925±0.001
	GMC [41]	0.791±0.001	0.941±0.000	0.782±0.000	0.794±0.000	0.694±0.000	0.929±0.000
	LMSC [30]	0.749±0.018	0.866±0.006	0.699±0.025	0.715±0.023	0.655±0.041	0.790±0.017
	MLAN [42]	0.862±0.011	0.961±0.004	0.835±0.006	0.844±0.013	0.758±0.008	0.953±0.007
	tSVDMSC [3]	0.830±0.000	0.884±0.005	0.786±0.003	0.800±0.004	0.785±0.007	0.808±0.001
	ETLMSC [13]	0.877±0.065	0.947±0.024	0.862±0.057	0.869±0.054	0.830±0.065	0.914±0.045
	LRTG [43]	0.927±0.000	**0.976±0.000**	0.928±0.000	0.932±0.000	0.905±0.000	**0.961±0.000**
	GNLTA [17]	0.888±0.056	0.956±0.017	0.881±0.049	0.887±0.046	0.856±0.063	0.922±0.028
	HLR-M^{2}VS [33]	0.852±0.009	0.960±0.006	0.833±0.005	0.842±0.003	0.757±0.010	0.949±0.011
	OPMC [44]	0.619±0.000	0.769±0.000	0.595±0.000	0.617±0.000	0.574±0.000	0.667±0.000
	SCGL [45]	0.874±0.000	0.945±0.000	0.845±0.000	0.854±0.000	**0.971±0.000**	0.762±0.000
	BTMSC	**0.970±0.021**	0.968±0.000	**0.944±0.024**	**0.947±0.023**	0.940±0.030	0.951±0.015
Reuters	SSC [9]	0.433±0.002	0.278±0.002	0.134±0.002	0.324±0.001	0.244±0.001	0.481±0.002
	LRR [8]	0.512±0.001	0.361±0.001	0.246±0.001	0.393±0.001	0.332±0.001	0.482±0.001
	RMSC [38]	0.496±0.039	0.296±0.020	0.238±0.020	0.372±0.015	0.350±0.017	0.398±0.014
	DiMSC [39]	0.476±0.007	0.307±0.002	0.175±0.006	0.345±0.003	0.276±0.005	0.459±0.007
	LT-MSC [12]	0.435±0.001	0.273±0.003	0.171±0.005	0.337±0.004	0.279±0.011	0.431±0.035
	GMC [41]	0.277±0.000	0.235±0.000	0.027±0.000	0.287±0.000	0.177±0.000	0.748±0.000
	LMSC [30]	0.512±0.004	0.317±0.005	0.238±0.004	0.375±0.003	0.346±0.003	0.408±0.004
	MLAN [42]	0.490±0.003	0.355±0.001	0.250±0.003	0.389±0.002	0.346±0.002	0.444±0.002
	tSVDMSC [3]	0.907±0.001	0.841±0.001	0.801±0.001	0.835±0.001	0.823±0.001	0.846±0.001
	ETLMSC [13]	0.932±0.049	0.957±0.050	0.924±0.073	0.938±0.060	0.914±0.066	0.968±0.055
	LRTG [43]	0.525±0.000	0.413±0.000	0.210±0.000	0.382±0.000	0.290±0.000	0.557±0.000
	GNLTA [17]	0.967±0.077	0.967±0.042	0.955±0.077	0.963±0.063	0.956±0.086	0.972±0.035
	HLR-M^{2}VS [33]	0.708±0.000	0.686±0.000	0.567±0.000	0.645±0.000	0.589±0.000	0.713±0.000
	OPMC [44]	0.303±0.000	0.163±0.000	0.052±0.000	0.274±0.000	0.193±0.000	0.470±0.000
	SCGL [45]	0.573±0.000	0.396±0.000	0.314±0.000	0.440±0.000	0.493±0.000	0.397±0.000
	BTMSC	**0.995±0.000**	**0.986±0.000**	**0.988±0.000**	**0.990±0.000**	**0.990±0.000**	**0.990±0.000**

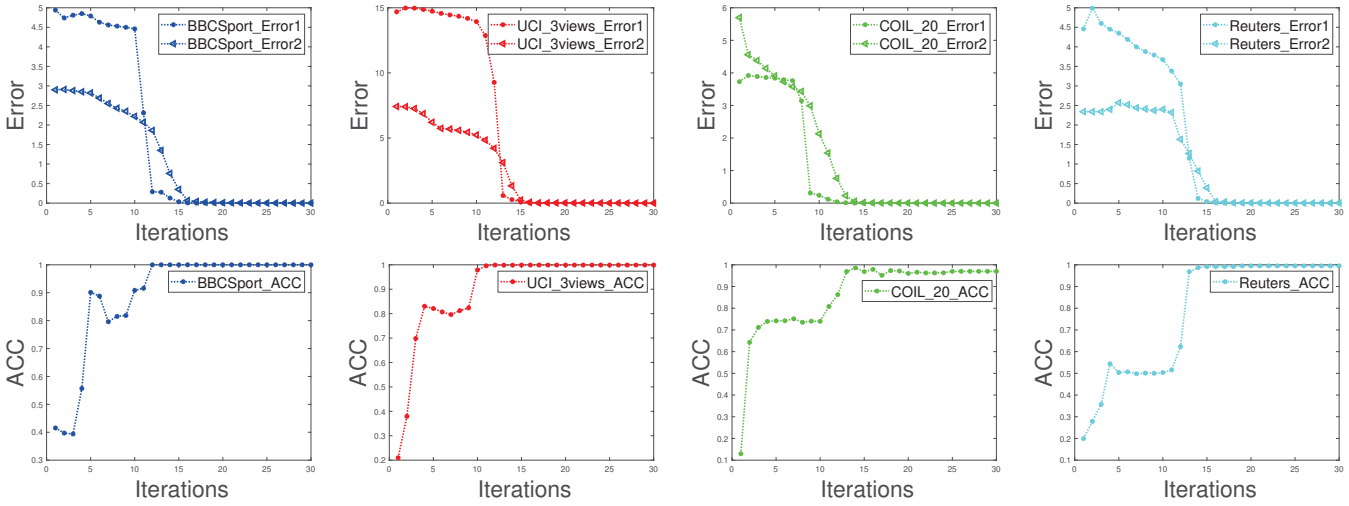

Fig. 1. Error and ACC versus iterations on BBCSport, UCI-3views, COIL_20 and Reuters datasets.

TABLE VI
CLUSTERING RESULTS (MEAN±STANDARD DEVIATION) ON *MSRC-V1* AND *Wikipedia* DATASETS.

Dataset	Method	ACC	NMI	AR	F-score	Precision	Recall
MSRC-V1	SSC [9]	0.791±0.007	0.750±0.005	0.651±0.006	0.701±0.005	0.670±0.008	0.736±0.003
	LRR [8]	0.695±0.000	0.590±0.000	0.491±0.000	0.562±0.000	0.560±0.000	0.564±0.000
	RMSC [38]	0.761±0.054	0.673±0.032	0.587±0.041	0.646±0.035	0.633±0.041	0.660±0.031
	DiMSC [39]	0.759±0.009	0.622±0.015	0.548±0.015	0.611±0.013	0.606±0.013	0.616±0.012
	LT-MSC [12]	0.831±0.003	0.743±0.004	0.665±0.004	0.712±0.004	0.699±0.004	0.725±0.003
	ECMSC [40]	0.795±0.002	0.750±0.002	0.681±0.001	0.727±0.001	0.705±0.001	0.750±0.001
	GMC [41]	0.748±0.000	0.771±0.000	0.640±0.000	0.697±0.000	0.612±0.000	0.809±0.000
	LMSC [30]	0.770±0.022	0.679±0.025	0.596±0.028	0.654±0.024	0.635±0.025	0.673±0.023
	MLAN [42]	0.859±0.003	0.751±0.003	0.709±0.004	0.750±0.003	0.727±0.004	0.776±0.002
	tSVDMSC [3]	0.991±0.000	0.982±0.000	0.978±0.000	0.981 ±0.000	0.980±0.000	0.982±0.000
	ETLMSC [13]	0.872±0.082	0.805±0.053	0.764±0.083	0.797±0.071	0.784±0.082	0.812±0.059
	LRTG [43]	0.895±0.000	0.829±0.000	0.775±0.000	0.807±0.000	0.794±0.000	0.821±0.000
	GNLTA [17]	0.894±0.089	0.880±0.053	0.829±0.092	0.854±0.078	0.836±0.095	0.873±0.060
	HLR-M^{2}VS [33]	**1.000±0.000**	**1.000±0.000**	**1.000±0.000**	**1.000±0.000**	**1.000±0.000**	**1.000±0.000**
	OPMC [44]	0.862±0.000	0.771±0.000	0.718±0.000	0.757±0.000	0.750±0.000	0.765±0.000
	SCGL [45]	0.900±0.000	0.808±0.000	0.787±0.000	0.817±0.000	0.819±0.000	0.814±0.000
	BTMSC	**1.000±0.000**	**1.000±0.000**	**1.000±0.000**	**1.000±0.000**	**1.000±0.000**	**1.000±0.000**
Wikipedia	SSC [9]	0.561±0.001	0.527±0.002	0.418±0.001	0.481±0.001	0.491±0.001	0.471±0.001
	LRR [8]	0.554±0.001	0.523±0.001	0.417±0.000	0.479±0.000	0.490±0.000	0.468±0.001
	RMSC [38]	0.558±0.015	0.513±0.015	0.430±0.020	0.491±0.019	0.501±0.021	0.483±0.017
	DiMSC [39]	0.547±0.007	0.500±0.003	0.397±0.002	0.461±0.002	0.478±0.002	0.445±0.002
	LT-MSC [12]	0.532±0.003	0.496±0.005	0.407±0.005	0.471±0.005	0.480±0.004	0.461±0.006
	ECMSC [40]	0.563±0.000	0.522±0.000	0.413±0.000	0.475±0.000	0.494±0.000	0.457±0.000
	GMC [41]	0.449±0.000	0.417±0.000	0.145±0.000	0.283±0.000	0.191±0.000	**0.550±0.000**
	LMSC [30]	0.186±0.000	0.070±0.000	0.021±0.000	0.123±0.000	0.128±0.000	0.123±0.000
	MLAN [42]	0.203±0.001	0.066±0.000	0.020±0.000	0.127±0.000	0.127±0.000	0.127±0.000
	t-SVD-MSC [3]	0.577±0.000	0.495±0.000	0.413±0.000	0.475±0.000	0.491±0.000	0.460±0.000
	ETLMSC [13]	0.561±0.000	0.499±0.000	0.398±0.000	0.462±0.000	0.472±0.000	0.453±0.000
	LRTG [43]	0.557±0.000	0.501±0.000	0.402 ±0.000	0.465±0.000	0.482±0.000	0.449 ±0.000
	GNLTA [17]	0.581±0.017	0.507±0.011	0.414±0.005	0.476±0.004	0.491±0.006	0.463±0.006
	HLR-M^{2}VS [33]	0.577±0.000	0.513±0.000	0.417±0.000	0.480±0.000	0.485±0.000	0.475±0.000
	OPMC [44]	0.189±0.000	0.066±0.000	0.021±0.000	0.128±0.000	0.128±0.000	0.127±0.000
	SCGL [45]	0.470±0.000	0.494±0.000	0.427±0.000	0.487±0.000	0.471±0.000	0.501±0.000
	BTMSC	**0.613±0.000**	**0.536±0.000**	**0.454±0.000**	**0.513±0.000**	**0.524±0.000**	0.502±0.000

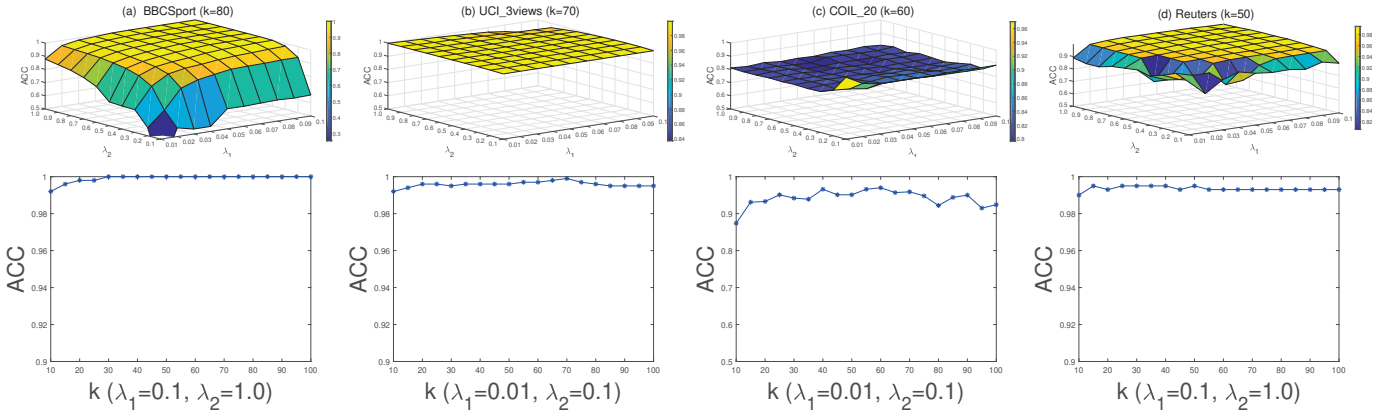

Fig. 2. Parameter tuning with respect to λ_1, λ_2 and k on (a) BBCSport, (b) UCI-3views, (c) COIL_20 and (d) Reuters datasets.

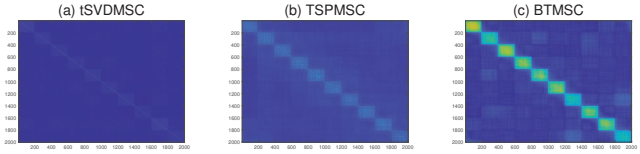

Fig. 3. Structural visualization comparison of the similarity matrices of (a) tSVDMSC, (b) TSPMSC and the proposed (c) BTMSC on UCI-3views dataset.

D. Ablation Experiment

We design the ablation experiment on Table IX to verify the effectiveness of BiN in the proposed BTMSC method. The baseline method is tSVDMSC based on tensor nuclear norm. The comparison methods are TSPMSC ($p = [0.1 : 0.1 : 0.9]$) based on tensor nonconvex Schatten-p norm and BTMSC based on BiN. Among them, we only record the p value of TSPMSC with the best clustering result. The clustering results in Table IX show that the ACC and NMI of BTMSC and TSPMSC are significantly improved compared to tSVDMSC, and BTMSC has the least running time. The difference between the clustering results of BTMSC and TSPMSC is very small, and the optimal result of TSPMSC is often obtained

TABLE VII
CLUSTERING RESULTS (MEAN±STANDARD DEVIATION) ON *Scene-15* AND *Caltech101* DATASETS.

Dataset	Method	ACC	NMI	AR	F-score	Precision	Recall
Scene-15	SSC [9]	0.444±0.003	0.470±0.002	0.279±0.001	0.337±0.002	0.292±0.001	0.397±0.001
	LRR [8]	0.445±0.013	0.426±0.018	0.272±0.015	0.324±0.010	0.316±0.011	0.333±0.015
	RMSC [38]	0.503±0.000	0.495±0.000	0.325±0.000	0.371±0.000	0.374±0.000	0.368±0.000
	DiMSC [39]	0.300±0.010	0.269±0.009	0.117±0.012	0.181±0.010	0.173±0.016	0.190±0.010
	LT-MSC [12]	0.574±0.009	0.571±0.011	0.424±0.010	0.465±0.007	0.452±0.003	0.479±0.008
	ECMSC [40]	0.457±0.001	0.463±0.002	0.303±0.001	0.357±0.001	0.318±0.001	0.408±0.001
	GMC [41]	0.381±0.000	0.519±0.000	0.191±0.000	0.281±0.000	0.174±0.000	0.732±0.000
	LMSC [30]	0.563±0.000	0.525±0.000	0.397±0.000	0.440 ±0.000	0.430±0.000	0.450±0.000
	MLAN [42]	0.331±0.000	0.475±0.000	0.151±0.000	0.248±0.000	0.150±0.000	0.731±0.000
	tSVDMSC [3]	0.812±0.007	0.858±0.007	0.771±0.003	0.788±0.001	0.743±0.006	0.839±0.003
	ETLMSC [13]	0.878±0.000	0.902±0.000	0.851±0.000	0.862±0.000	0.848±0.000	0.877±0.000
	LRTG [43]	0.615±0.016	0.657±0.005	0.486±0.016	0.525±0.014	0.485±0.023	0.572±0.005
	GNLTA [17]	0.881±0.043	0.895±0.013	0.850±0.032	0.861±0.030	0.846±0.041	0.876±0.020
	HLR-M^{2}VS [33]	0.878±0.003	0.895±0.005	0.850±0.003	0.861±0.005	0.850±0.008	0.871±0.010
	OPMC [44]	0.490±0.000	0.541±0.005	0.343±0.003	0.392±0.005	0.363±0.008	0.425±0.010
	SCGL [45]	0.599±0.000	0.587±0.000	0.439±0.000	0.478±0.000	0.486±0.000	0.471±0.000
	BTMSC	**0.955±0.003**	**0.935±0.000**	**0.910±0.005**	**0.916±0.004**	**0.922±0.004**	**0.911±0.005**
Caltech101	SSC [9]	0.420±0.015	0.723±0.005	0.303±0.011	0.317±0.012	0.441±0.025	0.248±0.010
	LRR [8]	0.510±0.009	0.728±0.014	0.304±0.017	0.339±0.008	0.627±0.012	0.231±0.010
	RMSC [38]	0.346±0.036	0.573±0.047	0.246±0.031	0.258±0.027	0.457±0.033	0.182±0.031
	DiMSC [39]	0.351±0.000	0.589±0.000	0.226±0.000	0.253±0.000	0.362±0.000	0.191±0.000
	LT-MSC [12]	0.559±0.012	0.788±0.005	0.393±0.007	0.403±0.003	0.670±0.009	0.288±0.012
	ECMSC [40]	0.359±0.004	0.606±0.001	0.273±0.003	0.286±0.006	0.433±0.020	0.214±0.030
	GMC [41]	0.331±0.000	0.544±0.000	0.031±0.000	0.081±0.000	0.044±0.000	0.470±0.000
	LMSC [30]	0.566±0.012	0.818±0.004	0.383±0.010	0.392±0.010	0.710±0.014	0.271±0.008
	MLAN [42]	0.579±0.024	0.748±0.020	0.222±0.015	0.265±0.015	0.173±0.009	0.560±0.016
	tSVDMSC [3]	0.607±0.005	0.858±0.003	0.430±0.005	0.440±0.010	0.742±0.007	0.323±0.009
	ETLMSC [13]	0.639±0.019	0.899±0.007	0.456±0.017	0.465±0.017	0.825±0.029	0.324±0.012
	LRTG [43]	0.490±0.000	0.750±0.000	0.340±0.000	0.350±0.000	0.547±0.000	0.260±0.000
	GNLTA [17]	0.604±0.016	0.875±0.005	0.444±0.017	0.453±0.016	0.776±0.018	0.320±0.015
	HLR-M^{2}VS [33]	0.650±0.000	0.872±0.000	0.463±0.000	0.472±0.000	0.760±0.000	0.343±0.000
	OPMC [44]	0.430±0.000	0.644±0.000	0.344±0.000	0.356±0.000	0.553±0.000	0.263±0.000
	SCGL [45]	0.617±0.000	0.813±0.000	0.422±0.000	0.432±0.000	0.313±0.000	**0.699±0.000**
	BTMSC	**0.683±0.014**	**0.903±0.000**	**0.504±0.024**	**0.512±0.023**	**0.843±0.015**	0.368±0.021

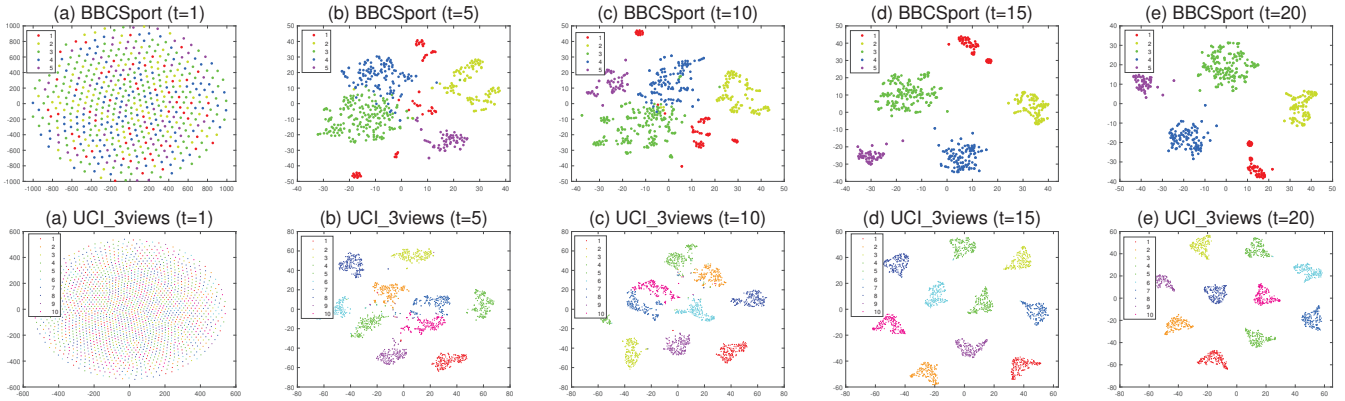

Fig. 4. t-SNE demonstration of similarity matrix of BTMSC during the iteration process on BBCSport and UCI-3views datasets.

when $0.3 \leq p \leq 0.6$. The above conclusions show that the nonconvex Schatten-p norm is effective, and BiN as the factorization form of Schatten-p norm ($p = 1/2$) plays an important role in clustering efficiency. The advantage of BTMSC over TSPMSC is that it introduces an additional noise term. Fig. 3 also show that the block-diagonal structure of the similarity matrix obtained by the proposed method is obvious. Overall, the proposed BTMSC method improves the clustering performance and efficiency over with the compared methods.

E. Numerical Convergence

We record the relative error and ACC versus iterations of the proposed BTMSC method on BBCSport, UCI-3views, COIL_20 and Reuters datasets in Fig. 1 to verify its convergence, where $Error1 = \frac{1}{M}\sum_{v=1}^{M}\|X^{(v)} - X^{(v)}Z^{(v)} - E_1^{(v)}\|_\infty$, $Error2 = \frac{1}{M}\sum_{v=1}^{M}\|Z^{(v)} - U^{(v)}V^{(v)T} - E_2^{(v)}\|_\infty$. It is easy to see that the error changes gently in the first 8 iterations and decreases rapidly after 10 iterations. With a few fluctuations, the error tends to 0 after 15 iterations. Corresponding to the change trend between ACC and iter-

TABLE VIII
CLUSTERING RESULTS (MEAN±STANDARD DEVIATION) ON *Mnist* AND *SUNRGBD* DATASETS.

Dataset	Method	ACC	NMI	AR	F-score	Precision	Recall
Mnist	SSC [9]	0.418±0.009	0.483±0.002	0.289±0.008	0.380±0.003	0.296±0.019	0.538±0.004
	LRR [8]	0.509±0.000	0.466±0.000	0.335±0.000	0.404±0.000	0.390±0.000	0.419±0.000
	RMSC [38]	0.545±0.021	0.502±0.012	0.382±0.011	0.445±0.009	0.434±0.012	0.457±0.007
	DiMSC [39]	0.548±0.000	0.385±0.000	0.275±0.000	0.352±0.000	0.332±0.000	0.374±0.000
	LT-MSC [12]	0.684±0.000	0.625±0.000	0.545±0.000	0.592±0.000	0.567±0.000	0.620±0.000
	GMC [41]	0.844±0.000	0.771±0.000	0.714±0.000	0.744±0.000	0.719±0.000	0.772±0.000
	LMSC [30]	0.445±0.001	0.379±0.004	0.254±0.002	0.330±0.002	0.323±0.003	0.337±0.001
	MLAN [42]	0.621±0.000	0.656±0.000	0.509±0.000	0.562±0.000	0.515±0.000	0.620±0.000
	tSVDMSC [3]	0.923±0.000	0.876±0.000	0.846±0.000	0.861±0.000	0.859±0.000	0.863±0.000
	ETLMSC [13]	0.888±0.000	0.891±0.000	0.830±0.000	0.848±0.000	0.827±0.000	0.871±0.000
	LRTG [43]	0.639±0.000	0.657±0.000	0.515±0.000	0.568±0.000	0.513±0.000	0.637±0.000
	GNLTA [17]	0.874±0.063	0.881±0.023	0.814±0.056	0.834±0.050	0.809±0.073	0.862±0.024
	HLR-M^{2}VS [33]	0.919±0.000	0.868±0.000	0.838±0.000	0.855±0.000	0.853±0.000	0.856±0.000
	OPMC [44]	0.552±0.000	0.517±0.000	0.398±0.000	0.461±0.000	0.438±0.000	0.487±0.000
	SCGL [45]	0.839±0.000	0.731±0.000	0.696±0.000	0.727±0.000	0.737±0.000	0.717±0.000
	BTMSC	**0.936±0.000**	**0.906±0.000**	**0.871±0.000**	**0.884±0.000**	**0.883±0.000**	**0.885±0.000**
SUNRGBD	SSC [9]	0.122±0.003	0.179±0.001	0.049±0.002	0.083±0.002	0.152±0.003	0.057±0.001
	LRR [8]	0.112±0.003	0.159±0.002	0.041±0.002	0.073±0.002	0.140±0.003	0.049±0.001
	RMSC [38]	0.200±0.007	0.265±0.007	0.101±0.005	0.131±0.005	0.252±0.008	0.089±0.003
	DiMSC [39]	0.127±0.004	0.156±0.003	0.040±0.001	0.075±0.001	0.134±0.004	0.052±0.001
	LT-MSC [12]	0.199±0.000	0.306±0.000	0.108±0.000	0.139±0.000	0.260±0.000	0.095±0.000
	GMC [41]	0.128±0.000	0.073±0.000	0.008±0.000	0.122±0.000	0.065±0.000	**0.944±0.000**
	LMSC [30]	0.087±0.003	0.115±0.004	0.024±0.002	0.057±0.002	0.107±0.003	0.039±0.001
	MLAN [42]	0.218±0.007	0.273±0.002	0.110±0.006	0.145±0.006	0.240±0.009	0.104±0.005
	tSVDMSC [3]	0.197±0.000	0.269±0.000	0.101±0.000	0.131±0.000	0.250±0.000	0.089±0.000
	ETLMSC [13]	0.195±0.000	0.264±0.000	0.098±0.000	0.128±0.000	0.246±0.000	0.087±0.000
	LRTG [43]	0.218±0.000	0.307±0.000	0.112±0.000	0.147±0.000	0.240±0.000	0.106±0.000
	GNLTA [17]	0.232±0.011	0.393±0.007	0.141±0.008	0.171±0.008	0.312±0.015	0.118±0.005
	HLR-M^{2}VS [33]	0.196±0.000	0.310±0.000	0.106±0.000	0.138±0.000	0.252±0.000	0.095±0.000
	OPMC [44]	0.201±0.000	0.259±0.000	0.105±0.000	0.136±0.000	0.253±0.000	0.093±0.000
	SCGL [45]	0.190±0.000	0.251±0.000	0.098±0.000	0.129±0.000	0.088±0.000	0.244±0.000
	BTMSC	**0.243±0.007**	**0.394±0.000**	**0.151±0.003**	**0.180±0.003**	**0.341±0.007**	0.123±0.002

TABLE IX
ABLATION EXPERIMENT: CLUSTERING RESULTS (MEAN±STANDARD DEVIATION) ON SIX DATASETS.

Method	P	*BBCSport*			P	*UCI-3views*		
		ACC	NMI	Time		ACC	NMI	Time
tSVDMSC	-	0.879±0.000	0.765±0.000	17.35	-	0.955±0.000	0.932±0.000	158.26
TSPMSC	0.6	**1.000±0.000**	**1.000±0.000**	20.66	1/2	0.996±0.001	0.989±0.002	185.60
BTMSC	1/2	**1.000±0.000**	**1.000±0.000**	**8.21**	1/2	**0.999±0.000**	**0.998±0.000**	**102.88**
Method	P	*COIL_20*			P	*Reuters*		
		ACC	NMI	Time		ACC	NMI	Time
tSVDMSC	-	0.830±0.000	0.884±0.005	169.10	-	0.907±0.001	0.841±0.001	101.01
TSPMSC	1/2	0.938±0.041	0.935±0.000	263.94	0.3	0.988±0.000	0.970±0.000	599.70
BTMSC	1/2	**0.970±0.021**	**0.968±0.000**	**99.10**	1/2	**0.995±0.000**	**0.986±0.000**	**99.27**
Method	P	*MSRC-V1*			P	*Wikipedia*		
		ACC	NMI	Time		ACC	NMI	Time
tSVDMSC	-	0.991±0.000	0.982±0.000	3.19	-	0.527±0.011	0.480±0.001	5.46
TSPMSC	0.4	**1.000±0.000**	**1.000±0.000**	3.38	0.6	0.574±0.001	0.498±0.002	9.97
BTMSC	1/2	**1.000±0.000**	**1.000±0.000**	**1.82**	1/2	**0.613±0.000**	**0.536±0.000**	**3.79**

ations, the ACC value increases rapidly after 5 iterations and reaches the optimal result after 15 iterations. The above phenomena indicate that the proposed BTMSC method is stable and convergent. In addition, we use t-SNE to visualize the clustering results of the similarity matrix C in Fig. 4. We present cluster visualization results for iterations 1, 5, 10, 15 and 20. As the iterations increasing, the similarity matrix C shows a clearer cluster structure after 15 iterations, which proves the convergence and superior clustering performance of the proposed method.

F. Running Time

For convenient comparison, we ignore the final k-means computation for all methods and only report the running time of the optimal similarity matrix. Fig. 5 shows the distribution of running time on seven datasets, where the y axis being log scale. We can intuitively see that BTMSC has the least running time. Among them, DiMSC and tSVDMSC have obvious running speed advantage on BBCSport, Reuters and MSRC-V1 datasets with relatively small samples, while on datasets with more than 2000 samples such as Scene-15,

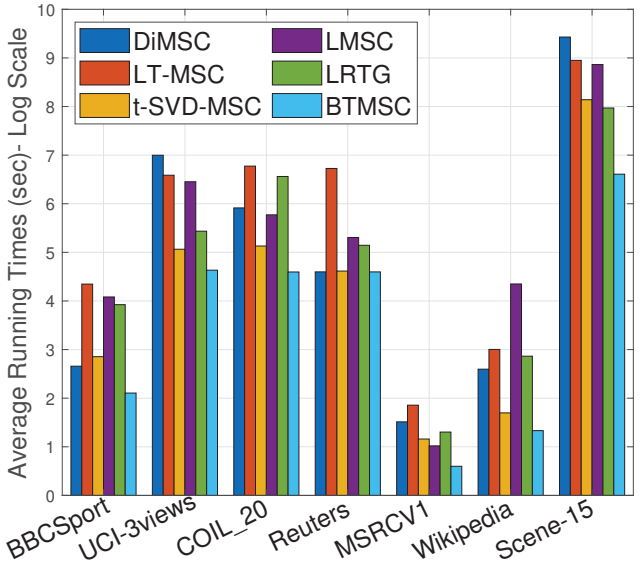

Fig. 5. The running time of compared methods and the proposed method on seven datasets.

the clustering efficiency of BTMSC is superior. The above conclusions demonstrate that the factorability of BTMSC is more easily applicable to large-scale datasets.

VII. Conclusion

In this paper, we proposed a Bi-nuclear tensor Schatten-p norm minimization method for multi-view subspace clustering. The proposed method has the following advantages. First, LRR-based multi-view subspace clustering methods often used the nuclear norm to explore the low-rank structure of views, which limits the clustering efficiency because the nuclear norm is a convex relaxation yet a biased approximation of the rank function, leading to suboptimal solution. The proposed method replaced the nuclear norm with the BiN as a nonconvex approximation of the rank function. Besides, the nuclear norm-based minimization problem may lead to high computational cost due to the SVD involving large-scale matrices. Considering the factorability of BiN of Schatten-p norm, we integrated nonconvex low-rank constraint and factorization to reduce the computational complexity. Furthermore, matrix-based LRR cannot simultaneously consider low-rank properties involving both intra-view and inter-view. Therefore, we applied the BiN to factorize the third-order tensor as the product of two small-scale tensors for low-rank constraint to explore higher-order correlation among views. The clustering results on ten datasets demonstrate that the proposed method can improve not only the clustering performance but also the computational efficiency. The limitation of the proposed method is that BTMSC is a tedious two-step clustering since the subspace and similarity matrix learning are independent. The future work will explore one-step multi-view clustering methods using the Wolfe dual algorithm to reduce the computational complexity.

References

[1] C. Zhang, S. Wang, J. Liu, S. Zhou, P. Zhang, X. Liu, E. Zhu, and C. Zhang, "Multi-view clustering via deep matrix factorization and partition alignment," in *Proc. Int. Conf. Multimedia*, 2021, pp. 4156–4164.

[2] Y. Zhang, X. Liu, S. Wang, J. Liu, S. Dai, and E. Zhu, "One-stage incomplete multi-view clustering via late fusion," in *Proc. Int. Conf. Multimedia*, 2021, pp. 2717–2725.

[3] Y. Xie, D. Tao, W. Zhang, Y. Liu, L. Zhang, and Y. Qu, "On unifying multi-view self-representations for clustering by tensor multi-rank minimization," *Int. J. Comput. Vis.*, vol. 126, no. 11, pp. 1157–1179, 2018.

[4] X. Wang, Z. Lei, X. Guo, C. Zhang, H. Shi, and S. Z. Li, "Multi-view subspace clustering with intactness-aware similarity," *Pattern Recognit.*, vol. 88, pp. 50–63, 2019.

[5] Z. Ren, Q. Sun, and D. Wei, "Multiple kernel clustering with kernel k-means coupled graph tensor learning," in *Proc. AAAI Conf. Artif. Intell.*, vol. 35, no. 11, 2021, pp. 9411–9418.

[6] Y. Tang, Y. Xie, X. Yang, J. Niu, and W. Zhang, "Tensor multi-elastic kernel self-paced learning for time series clustering," *IEEE Trans. Knowl. Data Eng.*, vol. 33, no. 3, pp. 1223–1237, 2019.

[7] Y. Xie, J. Liu, Y. Qu, D. Tao, W. Zhang, L. Dai, and L. Ma, "Robust kernelized multiview self-representation for subspace clustering," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 32, no. 2, pp. 868–881, 2020.

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

[9] E. Elhamifar and R. Vidal, "Sparse subspace clustering: Algorithm, theory, and applications," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 11, pp. 2765–2781, 2013.

[10] M. Meng, M. Lan, J. Yu, and J. Wu, "Multiview consensus structure discovery," *IEEE Trans. Cybern.*, vol. 52, no. 5, pp. 3469–3482, 2022.

[11] M. Brbic and I. Kopriva, "Multi-view low-rank sparse subspace clustering," *Pattern Recognit.*, vol. 73, pp. 247–258, 2018.

[12] C. Zhang, H. Fu, S. Liu, G. Liu, and X. Cao, "Low-rank tensor constrained multiview subspace clustering," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2015, pp. 1582–1590.

[13] J. Wu, Z. Lin, and H. Zha, "Essential tensor learning for multi-view spectral clustering," *IEEE Trans. Image Process.*, vol. 28, no. 12, pp. 5910–5922, 2019.

[14] S. Gu, L. Zhang, W. Zuo, and X. Feng, "Weighted nuclear norm minimization with application to image denoising," in *Proc. IEEE Conf. Comput. Vision Pattern Recognit.*, 2014, pp. 2862–2869.

[15] Y. Hu, D. Zhang, J. Ye, X. Li, and X. He, "Fast and accurate matrix completion via truncated nuclear norm regularization," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 9, pp. 2117–2130, 2012.

[16] Z. Kang, C. Peng, J. Cheng, and Q. Cheng, "Logdet rank minimization with application to subspace clustering," *Comput. Intell. Neuroscience*, vol. 2015, 2015.

[17] Y. Chen, X. Xiao, Z. Hua, and Y. Zhou, "Generalized nonconvex low-rank tensor approximation for multi-view subspace clustering," *IEEE Trans. Image Process.*, vol. 30, pp. 4022–4035, 2021.

[18] F. Shang, J. Cheng, Y. Liu, Z.-Q. Luo, and Z. Lin, "Bilinear factor matrix norm minimization for robust pca: Algorithms and applications," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 40, no. 9, pp. 2066–2080, 2017.

[19] M. Sun, P. Zhang, S. Wang, S. Zhou, W. Tu, X. Liu, E. Zhu, and C. Wang, "Scalable multi-view subspace clustering with unified anchors," in *Proc. Int. Conf. Multimedia*, 2021, pp. 3528–3536.

[20] S. Wang, X. Liu, X. Zhu, P. Zhang, Y. Zhang, F. Gao, and E. Zhu, "Fast parameter-free multi-view subspace clustering with consensus anchor guidance," *IEEE Trans. Image Process.*, vol. 31, pp. 556–568, 2021.

[21] J. Guo, S. Yin, Y. Sun, and Y. Hu, "Double manifolds regularized non-negative matrix factorization for data representation," in *Proc. IEEE Int. Conf. Pattern Recognit.* IEEE, 2021, pp. 901–906.

[22] C. Peng, Z. Kang, Y. Hu, J. Cheng, and Q. Cheng, "Robust graph regularized nonnegative matrix factorization for clustering," *ACM Trans. Knowl. Discov. Data*, vol. 11, no. 3, pp. 1–30, 2017.

[23] D. Kuang, "Nonnegative matrix factorization for clustering," Ph.D. dissertation, Georgia Institute of Technology, 2014.

[24] D. Cai, X. He, J. Han, and T. S. Huang, "Graph regularized nonnegative matrix factorization for data representation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 8, pp. 1548–1560, 2010.

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

[26] Y. Liu, L. Jiao, and F. Shang, "An efficient matrix factorization based low-rank representation for subspace clustering," *Pattern Recognit.*, vol. 46, no. 1, pp. 284–292, 2013.

[27] J. Guo, Y. Sun, J. Gao, Y. Hu, and B. Yin, "Rank consistency induced multiview subspace clustering via low-rank matrix factorization," *IEEE Trans. Neural Netw. Learn. Syst.*, 2021.

[28] Y. Wang, L. Wu, X. Lin, and J. Gao, "Multiview spectral clustering via structured low-rank matrix factorization," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 29, no. 10, pp. 4833–4843, 2018.

[29] Z. Ma, Z. Kang, G. Luo, L. Tian, and W. Chen, "Towards clustering-friendly representations: Subspace clustering via graph filtering," in *Proc. Int. Conf. Multimedia*, 2020, pp. 3081–3089.

[30] C. Zhang, H. Fu, Q. Hu, X. Cao, Y. Xie, D. Tao, and D. Xu, "Generalized latent multi-view subspace clustering," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 42, no. 1, pp. 86–99, 2020.

[31] K. Li, S. Li, Z. Ding, W. Zhang, and Y. Fu, "Latent discriminant subspace representations for multi-view outlier detection," in *Proc. AAAI Conf. Artif. Intell.*, 2018, pp. 3522–3529.

[32] Y. Chen, S. Wang, F. Zheng, and Y. Cen, "Graph-regularized least squares regression for multi-view subspace clustering," *Knowl.-Based Syst.*, vol. 194, p. 105482, 2020.

[33] Y. Xie, W. Zhang, and Y. Qu, "Hyper-laplacian regularized multilinear multiview self-representations for clustering and semisupervised learning," *IEEE Trans. Cybern.*, vol. 93, no. 2, pp. 572–586, 2020.

[34] S. Wang, Y. Chen, L. Zhang, Y. Cen, and V. Voronin, "Hyper-laplacian regularized nonconvex low-rank representation for multi-view subspace clustering," *IEEE Trans. Signal Inf. Process. over Networks*, vol. 8, pp. 376–388, 2022.

[35] M. Zhang, Z.-H. Huang, and Y. Zhang, "Restricted p-isometry properties of nonconvex matrix recovery," *IEEE Trans. Inf. Theory*, vol. 59, no. 7, pp. 4316–4323, 2013.

[36] Y. Xie, S. Gu, Y. Liu, W. Zuo, W. Zhang, and L. Zhang, "Weighted schatten p-norm minimization for image denoising and background subtraction," *IEEE Trans. Image Process.*, vol. 25, no. 10, pp. 4842–4857, 2016.

[37] F. Nie, H. Wang, X. Cai, H. Huang, and C. Ding, "Robust matrix completion via joint schatten p-norm and lp-norm minimization," in *Proc. IEEE 12th Int. Conf. Data Min.* IEEE, 2012, pp. 566–574.

[38] R. Xia, Y. Pan, L. Du, and J. Yin, "Robust multi-view spectral clustering via low-rank and sparse decomposition," in *Proc. AAAI Conf. Artif. Intell.*, vol. 28, no. 1, 2014.

[39] X. Cao, C. Zhang, H. Fu, S. Liu, and H. Zhang, "Diversity-induced multi-view subspace clustering," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2015, pp. 586–594.

[40] X. Wang, X. Guo, Z. Lei, C. Zhang, and S. Z. Li, "Exclusivity-consistency regularized multi-view subspace clustering," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2017, pp. 923–931.

[41] H. Wang, Y. Yang, and B. Liu, "GMC: Graph-based multi-view clustering," *IEEE Trans. Knowl. Data Eng.*, vol. 32, no. 6, pp. 1116–1129, 2020.

[42] F. Nie, G. Cai, J. Li, and X. Li, "Auto-weighted multi-view learning for image clustering and semi-supervised classification," *IEEE Trans. Image Process.*, vol. 27, no. 3, pp. 1501–1511, 2018.

[43] Y. Chen, X. Xiao, C. Peng, G. Lu, and Y. Zhou, "Low-rank tensor graph learning for multi-view subspace clustering," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 32, no. 1, pp. 92–104, 2022.

[44] J. Liu, X. Liu, Y. Yang, L. Liu, S. Wang, W. Liang, and J. Shi, "One-pass multi-view clustering for large-scale data," in *Proc. IEEE international Conf. Comput. Vision*, 2021, pp. 12 344–12 353.

[45] H. Yin, W. Hu, Z. Zhang, J. Lou, and M. Miao, "Incremental multi-view spectral clustering with sparse and connected graph learning," *Neural Netw.*, vol. 144, pp. 260–270, 2021.

[46] S. Wang, Y. Chen, Z. Lin, Y. Cen, and Q. Cao, "Robustness meets low-rankness: Unified entropy and tensor learning for multi-view subspace clustering," *IEEE Trans. Circuits Syst. Video Technol.*, 2023.