Multi-view Clustering by Spectral Structure Fusion and Novel Low-rank Approximation

Yin Long[1,2*], Xiaobo Liu[1], and Simon Murphy[2]
[1] School of Computer and Science Technology, Southwest University of Science and Technology
Mianyang 621000 P.R. China
[e-mail: yinnuoloong@163.com]
[2] School of Computer Science, University of Alberta
Edmonton, Alberta T6G 2R3 Canada
[*]Corresponding author: Yin Long

Received November 17, 2021; revised January 18, 2022; accepted March 8, 2022; published March 31, 2022

Abstract

In multi-view subspace clustering, how to integrate the complementary information between perspectives to construct a unified representation is a critical problem. In the existing works, the unified representation is usually constructed in the original data space. However, when the data representation in each view is very diverse, the unified representation derived directly in the original data domain may lead to a huge information loss. To address this issue, different to the existing works, inspired by the latest revelation that the data across all perspectives have a very similar or close spectral block structure, we try to construct the unified representation in the spectral embedding domain. In this way, the complementary information across all perspectives can be fused into a unified representation with little information loss, since the spectral block structure from all views shares high consistency. In addition, to capture the global structure of data on each view with high accuracy and robustness both, we propose a novel low-rank approximation via the tight lower bound on the rank function. Finally, experimental results prove that, the proposed method has the effectiveness and robustness at the same time, compared with the state-of-art approaches.

Keywords: Multi-view subspace clustering, rank-norm approximation, multi-view fusion, spectral structure, ADMM.

This research is supported by the National Natural Science Foundation of China (No. 62101467) and Natural Science Foundation of SWUST (No.19zx7141).

http://doi.org/10.3837/tiis.2022.03.004 **ISSN : 1976-7277**

1. Introduction

In some real-world problems, only using single view information does not satisfy the real-world demand very well. Thus, solving problems by the representation with respect to multiple views, i.e. multi-view learning [1], becomes an important approach. Even further, when the multi-view learning is applied into the field of clustering, the multi-view clustering is proposed [2], which needs to integrate these multiple views together to perform clustering. Recently, much progress has been made in developing multi-view clustering field, especially multi-view subspace clustering [3]. Since multi-view subspace clustering can make it easier to deal with high-dimensional data by learning a unified representation for data on all views, it attracts much attention. For example, in [4], authors separate noise from each view and learn a shared low-rank transition probability matrix for multi-view spectral clustering. In [5], the low-rank sparse subspace clustering is extended to multi-view clustering, and the consistent clustering results are obtained. In addition, to make full use of the complementary information across views, work [6] uses the Hilbert-Schmidt independent criterion to explore the complementary information between each view and reduces redundancy in the self-representation of each view data. In [7], a consistency item is developed on the basis of [6], and the authors jointly optimize the diversity representation and spectral clustering. All these approaches aim to maximize the complementary information between views but ignore the differences between views. Therefore, reasonably assigning different weights to the views can make the clustering results more accurate. In [8], by considering the difference between the data on each view, authors propose a self-adaptive weighting method to assign different weight to each view. In [9], a multi-graph fusion algorithm of multi-view spectral clustering is proposed, where the self-adaptive weighting method is also adopted.

As shown in the current works mentioned above, the complementary information between perspectives is captured straightly in the original data space. However, when the original data in each view is very diverse, the complementary information captured directly in the original data space may have a huge loss. Therefore, different to the previous works, inspired by the latest revelation that the data over all perspectives have a very similar or close spectral block structure, we are able to construct the complementary information in the spectral embedding domain. Besides, to better capture the global structure of the given data in subspace clustering, we propose a novel rank-norm approximation method. Experimental results demonstrate that the proposed method has better tradeoff between clustering effect and robustness.

2. Preliminary

2.1 Low-rank Sparse Representation

The aim of low-rank representation (LRR) [10] is to find a low-rank representation matrix $\mathbf{C} \in \mathbb{R}^{N\times N}$ for input data $\mathbf{X} = \{\mathbf{x}_i \in \mathbb{R}^D\}_{i=1}^N$ where D and N represent the dimension of each data point and the number of data points, respectively. Mathematically, LRR problem is described as

$$\min_{\mathbf{C}} rank(\mathbf{C}), \ \ s.t. \ \ \mathbf{X} = \mathbf{XC}, \tag{1}$$

where $rank(\mathbf{C})$ is the rank function. Since LRR is able to reconstruct the data space under the low-rank constraint, it is widely leveraged to capture the global structure of the given data.

Besides the low-rank constraint, to represent each data point by a small number of data points from the same subspace, the sparse constraint is also imposed on the matrix **C**. Thus, we have

$$\min_{\mathbf{C}} \left\|\mathbf{X}-\mathbf{X}\mathbf{C}\right\|_F^2 + \beta_1 rank\left(\mathbf{C}\right) + \beta_2 \left\|\mathbf{C}\right\|_1, \ s.t.\ diag\left(\mathbf{C}\right)=\mathbf{0}, \tag{2}$$

where $\left\|\cdot\right\|_1$ is the tightest convex relaxation of the sparse norm ℓ_0, the parameters β_1 and β_2 are the balance parameters for the low-rank constraint and the sparse constraint, respectively. Here, $diag\left(\mathbf{C}\right)=\mathbf{0}$ represents that the diagonal vector of matrix **C** is a null vector, which is used for avoiding trivial solution.

2.2 Multi-view Subspace Clustering

Given a *M*-view dataset $\mathbf{X}=\{\mathbf{X}^{(1)},\mathbf{X}^{(2)},...,\mathbf{X}^{(M)}\}$, where *M* is the number of views and $\mathbf{X}^{(m)}$ is the data on the *m*-th view, optimization problem (2) can be extended as

$$\begin{aligned} &\min_{\mathbf{C}^{(m)}} \left\|\mathbf{X}^{(m)}-\mathbf{X}^{(m)}\mathbf{C}^{(m)}\right\|_F^2 + \beta_1 rank\left(\mathbf{C}^{(m)}\right) + \beta_2 \left\|\mathbf{C}^{(m)}\right\|_1, \\ &s.t.\ diag\left(\mathbf{C}^{(m)}\right)=\mathbf{0},\ m=1,2...,M, \end{aligned} \tag{3}$$

where $\mathbf{C}^{(m)} \in \mathbb{R}^{N\times N}$ is the self-representation matrix of the *m*-th view. Since there exists common information shared among all views [11], we need to find a unified representation matrix **S** across all views. Consequently, the optimization problem about the unified representation matrix $\mathbf{S} \in \mathbb{R}^{N\times N}$ can be written as

$$\min_{\mathbf{S}} \sum_{m=1}^{M} p_m \left\|\mathbf{U}^{(m)}-\mathbf{S}\right\|_F^2, \ s.t.\ rank\left(\mathbf{L}_s\right)=N-k, \tag{4}$$

where $\mathbf{U}^{(m)}$ is the affinity matrix on the *m*-th view, expressed as $\mathbf{U}^{(m)}=(\mathbf{C}^{(m)}+(\mathbf{C}^{(m)})^T)/2$, $\mathbf{L}_s=\mathbf{I}_N-\mathbf{D}^{(-1/2)}\mathbf{S}\mathbf{D}^{(-1/2)}$ is the Laplace matrix of matrix **S** with **D** being its degree matrix whose diagonal element $\mathbf{D}_{ii}=\sum_{j\neq i}\mathbf{S}_{ij}$, the constraint $rank\left(\mathbf{L}_s\right)=N-k$ is used to guarantee that **S** contains *k* connected components [12], p_m is the weight characterizing the importance of the *m*-th view. With regard to the weight of the *m*-th view, it should be inversely proportional to $\left\|\mathbf{U}^{(m)}-\mathbf{S}\right\|_F^2$. Thus, we leverage the weighting scheme in [8], where p_m is set as $1/(2\left\|\mathbf{U}^{(m)}-\mathbf{S}\right\|_F^2)$.

Moreover, the optimization problem of multi-view subspace clustering under low-rank sparse constraint can be formulated as

$$\begin{aligned} \min_{\mathbf{C}^{(m)},\mathbf{S},\mathbf{F}} \sum_{m=1}^{M} \Big\{ &\left\|\mathbf{X}^{(m)}-\mathbf{X}^{(m)}\mathbf{C}^{(m)}\right\|_F^2 + \beta_1 rank\left(\mathbf{C}^{(m)}\right) + \beta_2 \left\|\mathbf{C}^{(m)}\right\|_1 \\ &+ p_m \left\|\mathbf{U}^{(m)}-\mathbf{S}\right\|_F^2 \Big\} + \eta Tr\left(\mathbf{F}^T\mathbf{L}_s\mathbf{F}\right), \\ s.t.\ &\mathbf{C}_{ij}^{(m)}>0,\ diag\left(\mathbf{C}^{(m)}\right)=\mathbf{0},\ \mathbf{F}^T\mathbf{F}=\mathbf{I}_k, \end{aligned} \tag{5}$$

where $\min_{\mathbf{F}} Tr(\mathbf{F}^T\mathbf{L}_s\mathbf{F})$ is equivalent to the constraint $rank(\mathbf{L}_s)=N-k$, η represents its

corresponding balance parameter. When problem (5) is solved, the clustering results can be derived straightly without performing other clustering.

3. The Proposed Approach

In this section, we first propose a novel rank-norm approximation which provides better tradeoff between the accuracy and robustness for LRR. Secondly, we propose an alternative way to integrate the complementary information across different views.

3.1 Novel Rank-norm Approximation

As we know, LRR problem inevitably involves solving a rank minimization problem, which is known to be non-convex, thus it is a NP-hard issue. An alternative for it is to relax the rank function by the nuclear norm. Such that the relaxed problem is convex and can be readily solved by the soft-thresholding operation on singular values. However, the nuclear norm and its enhanced methods are the sum of singular values essentially, leading to a biased estimation for the rank function. Towards this issue, a series of non-convex relaxations are proposed, such as the Schatten-*p* norm [13] and the Gamma norm [14] proposed recently. The Gamma norm is an extension of the min-max concave plus function, which can be nearly unbiased to approximate the rank function [15]. To illustrate it with more details, we first assume that the singular value decomposition (SVD) of the matrix $\mathbf{C}$ can be written as $\mathbf{C}=\mathbf{U}\mathbf{\Sigma}\mathbf{V}^T$, where $\mathbf{U}=\left[\mathbf{u}_1,\mathbf{u}_2,...,\mathbf{u}_N\right]$ and $\mathbf{V}=\left[\mathbf{v}_1,\mathbf{v}_2,...,\mathbf{v}_N\right]$ denote two unitary matrices, respectively, and $\mathbf{\Sigma}=diag(\sigma_1,\sigma_2,...,\sigma_N)$ is a diagonal matrix with $\sigma_1\geq\sigma_2\geq...\geq\sigma_N\geq 0$. Then, the Gamma norm of matrix $\mathbf{C}$ can be defined as

$$\left\|\mathbf{C}\right\|_\gamma=\sum_{i=1}^{N}J_\gamma\left(\sigma_i\right), \tag{6}$$

where $J_\gamma(\sigma_i)$ is a piecewise function written as

$$J_\gamma\left(\sigma_i\right)=\int_0^{\sigma_i}\left[1-\frac{x}{\gamma}\right]_+dx=\begin{cases}\dfrac{\gamma}{2}, & \text{if} \quad \sigma_i\geq\gamma\\ \sigma_i-\dfrac{\sigma_i^2}{2\gamma}, & \text{if} \quad \sigma_i<\gamma\end{cases}, \tag{7}$$

where $\left[x\right]_+=\max\left(x,0\right)$. As shown in (7), the Gamma norm is an extremely large or small concave function, so the approximation error would be very small. Besides, one may observe that, the parameter γ decides on which piecewise of function would be exploited for $J_\gamma(\sigma_i)$. However, the piecewise function in (7) is not continuous, and $J_\gamma(\sigma_i)$ is heavily dependent on the parameter γ [16]. Hence, the Gamma norm based approximation of the rank function is very sensitive to the parameter γ, which leads to a weak robustness to γ.

To obtain a better tradeoff between the approximation error and robustness of the rank-norm approximation, we propose a novel rank-norm approximation method. In this method, by resorting to skills of the matrix transformation, small approximation error with good robustness can be achieved. Specifically, we first assume *r* and σ_i represent the rank and the *i*-th singular value of matrix $\mathbf{C}$, respectively. Then, for the rank function of matrix $\mathbf{C}$, we have

$$rank(\mathbf{C}) = r = \sum_{i=1}^{r} \frac{\sigma_i^2}{\sigma_i^2} \geq \sum_{i=1}^{r} \frac{\sigma_i^2}{\sigma_i^2 + \varepsilon} \triangleq \varphi_\varepsilon(\mathbf{C}). \tag{8}$$

By observing the above equation, one may notice that the error between $rank(\mathbf{C})$ and its lower bound $\varphi_\varepsilon(\mathbf{C})$ would be close to 0 when ε approaches 0. Therefore, we can use $\varphi_\varepsilon(\mathbf{C})$ to approximate the rank function in a very accurate way. On the other side, $\varphi_\varepsilon(\mathbf{C})$ would vary continuously with ε. And most importantly, when σ_i is much larger than ε, $\varphi_\varepsilon(\mathbf{C})$ barely changes with ε, thereby having an ideal robustness to ε. However, the rank function represented by $\varphi_\varepsilon(\mathbf{C})$ in (8) is an implicit function about the variable *r*, which would cause an intractable obstacle to solve the optimization problem. For it, we need to make a further transformation of $\varphi_\varepsilon(\mathbf{C})$, so as to obtain the explicit function about the rank *r*. To be specific, for $\varphi_\varepsilon(\mathbf{C})$, we have

$$\begin{aligned}
\varphi_\varepsilon(\mathbf{C}) &= Tr\begin{pmatrix} \sigma_1^2(\sigma_1^2+\varepsilon)^{-1} & 0 & \cdots & \cdots & \cdots & 0 \\ 0 & \ddots & 0 & 0 & \ddots & \vdots \\ \vdots & \ddots & \sigma_r^2(\sigma_r^2+\varepsilon)^{-1} & \ddots & \ddots & \vdots \\ \vdots & \ddots & 0 & 0 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & \vdots \\ 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix} \\
&= Tr\left(\begin{pmatrix} (\sigma_1^2+\varepsilon)^{-1} & 0 & \cdots & \cdots & \cdots & 0 \\ 0 & \ddots & 0 & 0 & \ddots & \vdots \\ \vdots & \ddots & (\sigma_r^2+\varepsilon)^{-1} & \ddots & \ddots & \vdots \\ \vdots & \ddots & 0 & 0 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & \vdots \\ 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}\begin{pmatrix} \sigma_1^2 & 0 & \cdots & \cdots & \cdots & 0 \\ 0 & \ddots & 0 & 0 & \ddots & \vdots \\ \vdots & \ddots & \sigma_r^2 & \ddots & \ddots & \vdots \\ \vdots & \ddots & 0 & 0 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & \vdots \\ 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}\right) \\
&= Tr\left(\begin{pmatrix} (\sigma_1^2+\varepsilon)^{-1} & 0 & \cdots & \cdots & \cdots & 0 \\ 0 & \ddots & 0 & 0 & \ddots & \vdots \\ \vdots & \ddots & (\sigma_r^2+\varepsilon)^{-1} & \ddots & \ddots & \vdots \\ \vdots & \ddots & 0 & \varepsilon^{-1} & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & \vdots \\ 0 & 0 & 0 & 0 & 0 & \varepsilon^{-1} \end{pmatrix}\begin{pmatrix} \sigma_1^2 & 0 & \cdots & \cdots & \cdots & 0 \\ 0 & \ddots & 0 & 0 & \ddots & \vdots \\ \vdots & \ddots & \sigma_r^2 & \ddots & \ddots & \vdots \\ \vdots & \ddots & 0 & 0 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & \vdots \\ 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}\right) \\
&= Tr\left(\begin{pmatrix} diag(\sigma^2(\mathbf{C})) + \varepsilon\mathbf{I}_r & \mathbf{0} \\ \mathbf{0} & \varepsilon\mathbf{I}_{N-r} \end{pmatrix}^{-1}\begin{pmatrix} diag(\sigma^2(\mathbf{C})) & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix}\right),
\end{aligned} \tag{9}$$

where $diag(\sigma^2(\mathbf{C}))$ is a diagonal matrix with r squares of corresponding non-zero singular values of $\mathbf{C}$. Then, based on the property of unitary matrix, $\varphi_\varepsilon(\mathbf{C})$ can be further written as

$$\begin{aligned}\varphi_\varepsilon(\mathbf{C}) &= Tr\left(\begin{pmatrix} diag(\sigma^2(\mathbf{C}))+\varepsilon\mathbf{I}_r & \mathbf{0} \\ \mathbf{0} & \varepsilon\mathbf{I}_{N-r}\end{pmatrix}^{-1}\left(\mathbf{\Sigma}^T\mathbf{\Sigma}\right)\right) \\ &= Tr\left(\mathbf{U}\mathbf{\Sigma}\begin{pmatrix} diag(\sigma^2(\mathbf{C}))+\varepsilon\mathbf{I}_r & \mathbf{0} \\ \mathbf{0} & \varepsilon\mathbf{I}_{N-r}\end{pmatrix}^{-1}\mathbf{\Sigma}^T\mathbf{U}^T\right) \\ &= Tr\left(\mathbf{U}\mathbf{\Sigma}\mathbf{V}^T\mathbf{V}\begin{pmatrix} diag(\sigma^2(\mathbf{C}))+\varepsilon\mathbf{I}_r & \mathbf{0} \\ \mathbf{0} & \varepsilon\mathbf{I}_{N-r}\end{pmatrix}^{-1}\mathbf{V}^T\mathbf{V}\mathbf{\Sigma}^T\mathbf{U}^T\right) \\ &= Tr\left(\mathbf{C}\mathbf{V}\begin{pmatrix} diag(\sigma^2(\mathbf{C}))+\varepsilon\mathbf{I}_r & \mathbf{0} \\ \mathbf{0} & \varepsilon\mathbf{I}_{N-r}\end{pmatrix}^{-1}\mathbf{V}^T\mathbf{C}^T\right)\end{aligned} \tag{10}$$

Then, by resorting to the property of the inversion of unitary matrix, $\varphi_\varepsilon(\mathbf{C})$ can be further written as

$$\begin{aligned}\varphi_\varepsilon(\mathbf{C}) &= Tr\left(\mathbf{C}\left(\mathbf{V}^T\right)^{-1}\begin{pmatrix} diag(\sigma^2(\mathbf{C}))+\varepsilon\mathbf{I}_r & \mathbf{0} \\ \mathbf{0} & \varepsilon\mathbf{I}_{N-r}\end{pmatrix}^{-1}\mathbf{V}^{-1}\mathbf{C}^T\right) \\ &= Tr\left(\mathbf{C}\left(\mathbf{V}^T\right)^{-1}\left(\mathbf{\Sigma}^T\mathbf{\Sigma}+\varepsilon\mathbf{I}_N\right)^{-1}\mathbf{V}^{-1}\mathbf{C}^T\right) \\ &= Tr\left(\mathbf{C}\left(\mathbf{V}\left(\mathbf{\Sigma}^T\mathbf{\Sigma}+\varepsilon\mathbf{I}_N\right)\mathbf{V}^T\right)^{-1}\mathbf{C}^T\right)\end{aligned} \tag{11}$$

Based on (11), the closed-form expression of $\varphi_\varepsilon(\mathbf{C})$ can be derived via the inverse process of SVD of $\mathbf{C}$

$$\begin{aligned}\varphi_\varepsilon(\mathbf{C}) &= Tr\left(\mathbf{C}\left(\mathbf{V}\left(\mathbf{\Sigma}^T\mathbf{\Sigma}\right)\mathbf{V}^T+\varepsilon\mathbf{I}_N\right)^{-1}\mathbf{C}^T\right) \\ &= Tr\left(\mathbf{C}\left(\mathbf{V}\left(\mathbf{\Sigma}^T\mathbf{U}^T\mathbf{U}\mathbf{\Sigma}\right)\mathbf{V}^T+\varepsilon\mathbf{I}_N\right)^{-1}\mathbf{C}^T\right) \\ &= Tr\left(\mathbf{C}\left(\mathbf{C}^T\mathbf{C}+\varepsilon\mathbf{I}_N\right)^{-1}\mathbf{C}^T\right).\end{aligned} \tag{12}$$

So far, we obtain the novel approximation for the rank function of matrix $\mathbf{C}$ in the closed-form way.

3.2 The Integration of Multi-view Data on Spectral Structure

In the data space, since there may exist dramatic divergence across each view, integrating the data from multiple views directly may cause a large information loss. However, work [17] reveals that the data from different views still have a very similar spectral block structure. Inspired by this discovery, we turn to establish a unified spectral block structure across all views in the spectral embedding space [18]. In this way, the issue caused by the obvious inconsistency of the data representation among all views can be avoided. Specifically, to

obtain the spectral block structure on each view, we incorporate spectral clustering into the optimization, and then perform an adaptive integration based on minimizing the difference between the spectral block structure on each view and the unified spectral block structure. To be specific, the objective function is formulated as

$$\begin{aligned} &\min_{\mathbf{S},\mathbf{F}_m} \sum_{m=1}^{M} p_m \left\| \mathbf{F}_m \mathbf{F}_m^T - \mathbf{S} \right\|_F^2 + \beta_3 Tr\left(\mathbf{F}_m^T \mathbf{L}_m \mathbf{F}_m \right), \\ &s.t.\ \ \mathbf{F}_m^T \mathbf{F}_m = \mathbf{I}_k,\ rank\left(\mathbf{L}_s \right) = N - k,\ \mathbf{S1} = \mathbf{1}, \end{aligned} \tag{13}$$

where $\mathbf{L}_m = \mathbf{I}_N - \mathbf{D}_m^{(-1/2)} \mathbf{U}^{(m)} \mathbf{D}_m^{(-1/2)}$ is the Laplace matrix of the affinity matrix $\mathbf{U}^{(m)}$ with $\mathbf{D}_m$ being its degree matrix, $\mathbf{F}_m \in \mathbb{R}^{N \times k}$ denotes the spectral embedding matrix of the *m*-th view, $\mathbf{F}_m \mathbf{F}_m^T$ is the spectral block structure of the *m*-th view, $\mathbf{S}$ is the unified spectral structure across all views, $\mathbf{1}$ is a column vector that all elements are 1, β_3 denotes the balance parameter for the spectral embedding term.

So far, the optimization problem (5) can be reformulated as

$$\begin{aligned} \min_{\mathbf{C}^{(m)},\mathbf{S},\mathbf{F},\mathbf{F}_m} & \sum_{m=1}^{M} \left\{ \left\| \mathbf{X}^{(m)} - \mathbf{X}^{(m)} \mathbf{C}^{(m)} \right\|_F^2 + \beta_1 \varphi_\varepsilon \left(\mathbf{C}^{(m)} \right) + \beta_2 \left\| \mathbf{C}^{(m)} \right\|_1 \right\} \\ & + \sum_{m=1}^{M} \left\{ p_m \left\| \mathbf{F}_m \mathbf{F}_m^T - \mathbf{S} \right\|_F^2 + \beta_3 Tr\left(\mathbf{F}_m^T \mathbf{L}_m \mathbf{F}_m \right) \right\} + \eta Tr\left(\mathbf{F}^T \mathbf{L}_s \mathbf{F} \right), \\ s.t.\ \ & \mathbf{C}_{ij}^{(m)} > 0,\ diag\left(\mathbf{C}^{(m)} \right) = \mathbf{0},\ \mathbf{F}_m^T \mathbf{F}_m = \mathbf{I}_k,\ \mathbf{F}^T \mathbf{F} = \mathbf{I}_k. \end{aligned} \tag{14}$$

4. Optimization Algorithms

To solve the optimization problem (14), Alternating Direction Method of Multipliers (ADMM) is adopted here.

4.1 Update $\mathbf{C}^{(m)}$

With fixed $\mathbf{S}$, $\mathbf{F}$ and $\mathbf{F}_m$, the optimization problem can be reduced as

$$\min_{\mathbf{C}} \left\| \mathbf{X} - \mathbf{XC} \right\|_F^2 + \beta_1 \varphi_\varepsilon \left(\mathbf{C} \right) + \beta_2 \left\| \mathbf{C} \right\|_1 - \beta_3 Tr\left(\mathbf{T}_m^T \mathbf{C} \mathbf{T}_m \right),\ \ s.t.\ \ diag\left(\mathbf{C} \right) = \mathbf{0}, \tag{15}$$

where $\mathbf{T}_m = \mathbf{D}_m^{(-1/2)} \mathbf{F}_m$ and the superscript of *m* is omitted for the sake of description convenience. Then, by introducing the auxiliary variables $\{\mathbf{C}_i\}_{i=1}^3$ and $\mathbf{W}$, (15) can be reformulated as

$$\begin{aligned} &\min_{\mathbf{W},\{\mathbf{C}_i\}_{i=1}^3} \left\| \mathbf{X} - \mathbf{XW} \right\|_F^2 + \beta_1 \varphi_\varepsilon \left(\mathbf{C}_1 \right) + \beta_2 \left\| \mathbf{C}_2 \right\|_1 - \beta_3 Tr\left(\mathbf{T}_m^T \mathbf{C}_3 \mathbf{T}_m \right), \\ &s.t.\ \ \mathbf{W} = \mathbf{C}_2 - diag\left(\mathbf{C}_2 \right),\ \mathbf{W} = \mathbf{C}_1,\ \mathbf{W} = \mathbf{C}_3. \end{aligned} \tag{16}$$

Moreover, its augmented Lagrangian is

$$\begin{aligned} L\left(\mathbf{W},\{\mathbf{C}_i\}_{i=1}^3,\{\mathbf{\Omega}_i\}_{i=1}^3\right) &= \left\|\mathbf{X}-\mathbf{X}\mathbf{W}\right\|_F^2+\beta_1\varphi_\varepsilon\left(\mathbf{C}_1\right)+\beta_2\left\|\mathbf{C}_2\right\|_1-\beta_3 Tr\left(\mathbf{T}_m^T\mathbf{C}_3\mathbf{T}_m\right) \\ &+\frac{v}{2}\left(\left\|\mathbf{W}-\mathbf{C}_2+diag\left(\mathbf{C}_2\right)\right\|_F^2+\left\|\mathbf{W}-\mathbf{C}_1\right\|_F^2+\left\|\mathbf{W}-\mathbf{C}_3\right\|_F^2\right) \\ &+Tr\left(\mathbf{\Omega}_1^T\left(\mathbf{W}-\mathbf{C}_2+diag\left(\mathbf{C}_2\right)\right)\right) \\ &+Tr\left(\mathbf{\Omega}_2^T\left(\mathbf{W}-\mathbf{C}_1\right)\right)+Tr\left(\mathbf{\Omega}_3^T\left(\mathbf{W}-\mathbf{C}_3\right)\right), \end{aligned} \tag{17}$$

where v is the penalty parameter and $\{\mathbf{\Omega}_i\}_{i=1}^3$ are the Lagrange dual variables.

4.1.1 Update W

By setting the partial derivative of (17) with respect to $\mathbf{W}$ as 0, $\mathbf{W}$ can be updated directly as

$$\mathbf{W}=\left(2\mathbf{X}^T\mathbf{X}+3v\mathbf{I}_N\right)^{-1}\left(2\mathbf{X}^T\mathbf{X}+v\left(\mathbf{C}_1+\mathbf{C}_2+\mathbf{C}_3\right)-\mathbf{\Omega}_1-\mathbf{\Omega}_2-\mathbf{\Omega}_3\right). \tag{18}$$

4.1.2 Update $\mathbf{C}_1$

Based on the gradient descent method, $\mathbf{C}_1$ can be updated iteratively as

$$\mathbf{C}_1^{(t+1)}=\mathbf{C}_1^{(t)}-\eta_{\mathbf{C}_1}\nabla_{\mathbf{C}_1}, \tag{19}$$

where t represents the number of iterations, $\eta_{\mathbf{C}_1}$ denotes the learning rate, and $\nabla_{\mathbf{C}_1}$ represents the gradient about $\mathbf{C}_1$. However, to obtain $\nabla_{\mathbf{C}_1}$, we need to calculate the partial derivative of (17) with respect to $\mathbf{C}_1$. Firstly, we need to calculate the partial derivative of $\varphi_\varepsilon\left(\mathbf{C}_1\right)$ about $\mathbf{C}_1$. Specifically, based on the derivative rule about the trace of matrices, we have

$$\partial\left(\varphi_\varepsilon\left(\mathbf{C}_1\right)\right)/\partial\left(\mathbf{C}_1\right)=2\mathbf{C}_1\left(\left(\mathbf{C}_1\right)^T\mathbf{C}_1+\varepsilon\mathbf{I}_N\right)^{-1}\left(\mathbf{I}_N-\left(\mathbf{C}_1\right)^T\mathbf{C}_1\left(\left(\mathbf{C}_1\right)^T\mathbf{C}_1+\varepsilon\mathbf{I}_N\right)^{-1}\right). \tag{20}$$

In addition, we need to calculate the partial derivative of $\left\|\mathbf{W}-\mathbf{C}_1\right\|_F^2$ with respect to $\mathbf{C}_1$. To be specific, by resorting to the derivative rule about the squared Frobenius norm of the matrix, we can obtain

$$\partial\left(\left\|\mathbf{W}-\mathbf{C}_1\right\|_F^2\right)/\partial\left(\mathbf{C}_1\right)=-2\mathbf{W}+2\mathbf{C}_1. \tag{21}$$

Finally, we need to calculate the partial derivative of $Tr\left(\mathbf{\Omega}_2^T\left(\mathbf{W}-\mathbf{C}_1\right)\right)$ about $\mathbf{C}_1$. Specifically, based on the derivative rule about the trace of the matrix, we have

$$\partial\left(Tr\left(\mathbf{\Omega}_2^T\left(\mathbf{W}-\mathbf{C}_1\right)\right)\right)/\partial\left(\mathbf{C}_1\right)=-\mathbf{\Omega}_2. \tag{22}$$

In the end, based on (20), (21) and (22), the gradient $\nabla_{\mathbf{C}_1}$ in (19) can be further written as

$$\nabla_{\mathbf{C}_1}=v\mathbf{C}_1-v\mathbf{W}-\mathbf{\Omega}_2+2\mathbf{C}_1\left(\mathbf{C}_1^T\mathbf{C}_1+\varepsilon\mathbf{I}_N\right)^{-1}\left(\mathbf{I}_N-\mathbf{C}_1^T\mathbf{C}_1\left(\mathbf{C}_1^T\mathbf{C}_1+\varepsilon\mathbf{I}_N\right)^{-1}\right). \tag{23}$$

4.1.3 Update $\mathbf{C}_2$

From [19] [20], the update rule of $\mathbf{C}_2$ is

$$\mathbf{C}_2 = \mathbf{C}_2' - diag\left(\mathbf{C}_2'\right), \tag{24}$$

where $\mathbf{C}_2' = \pi_{\frac{\beta_2}{\nu}}(\mathbf{W} + \mathbf{\Omega}_1 / \nu)$, and $\pi_{\frac{\beta_2}{\nu}}(.)$ is the soft-thresholding operation applied entry-wise to $(\mathbf{W} + \mathbf{\Omega}_1 / \nu)$ [20].

4.1.4 Update $\mathbf{C}_3$

Based on the gradient descent method, $\mathbf{C}_3$ can be updated iteratively as

$$\mathbf{C}_3^{(t+1)} = \mathbf{C}_3^{(t)} - \eta_{\mathbf{C}_3} \nabla_{\mathbf{C}_3}, \tag{25}$$

where the gradient $\nabla_{\mathbf{C}_3} = \nu\mathbf{C}_3 - \nu\mathbf{W} - \beta_3 \mathbf{T}_m \mathbf{T}_m^T - \mathbf{\Omega}_3$.

4.1.5 Update $\{\mathbf{\Omega}_i\}_{i=1}^3$

Given $\mathbf{W}$ and $\{\mathbf{C}_i\}_{i=1}^3$, by the update rules for the dual variables in [5], dual variables $\{\mathbf{\Omega}_i\}_{i=1}^3$ can be updated as

$$\begin{aligned} \mathbf{\Omega}_1^{t+1} &= \mathbf{\Omega}_1^t + \nu\left(\mathbf{W} - \mathbf{C}_2\right), \\ \mathbf{\Omega}_2^{t+1} &= \mathbf{\Omega}_2^t + \nu\left(\mathbf{W} - \mathbf{C}_1\right), \\ \mathbf{\Omega}_3^{t+1} &= \mathbf{\Omega}_3^t + \nu\left(\mathbf{W} - \mathbf{C}_3\right). \end{aligned} \tag{26}$$

4.2 Update S

With fixed $\mathbf{C}^{(m)}$, $\mathbf{F}$ and $\mathbf{F}_m$, the optimization problem can be reduced as

$$\min_{\mathbf{S}} \sum_{m=1}^{M} \left\{ p_m \left\| \overline{\mathbf{F}}_m - \mathbf{S} \right\|_F^2 \right\} + \eta Tr\left(\mathbf{F}^T \mathbf{L}_s \mathbf{F}\right), \ s.t. \ \mathbf{F}^T\mathbf{F} = \mathbf{I}_k, \tag{27}$$

where $\overline{\mathbf{F}}_m \triangleq \mathbf{F}_m \mathbf{F}_m^T$. By assuming the j-th entry of $\mathbf{q}_i \in \mathbb{R}^{N\times 1}$ and it can be set as $\left\|\mathbf{F}(i,:) - \mathbf{F}(j,:)\right\|_2^2$, the optimization problem (27) with regard to the i-th column of $\mathbf{S}$ can be formulated as

$$\min_{\mathbf{S}(:,i)} \sum_{m=1}^{M} p_m \left\| \overline{\mathbf{F}}_m(:,i) - \mathbf{S}(:,i) \right\|_2^2 + \eta \mathbf{q}_i^T \mathbf{S}(:,i). \tag{28}$$

By taking the partial derivative about $\mathbf{S}(,:)$ on (28) as 0, we can obtain

$$\mathbf{S}(:,i) = \left(\sum\nolimits_m p_m \overline{\mathbf{F}}_m(:,i) - \frac{\eta \mathbf{q}_i}{2} \right) \Big/ \sum\nolimits_m p_m. \tag{29}$$

Thus, $\mathbf{S}$ can be obtained by implementing (29) on the each column of it.

4.3 Update F

With fixed $\mathbf{C}^{(m)}$, $\mathbf{S}$ and $\mathbf{F}_m$, the optimization problem can be reduced as

$$\min_{\mathbf{F}} Tr\left(\mathbf{F}^T\mathbf{L}_s\mathbf{F}\right),\ s.t.\ \mathbf{F}^T\mathbf{F}=\mathbf{I}_k. \tag{30}$$

Obviously, the optimal solution of $\mathbf{F}$ is the matrix consisting of the eigenvectors of $\mathbf{L}_s$ corresponding to the k smallest eigenvalues.

4.4 Update $\mathbf{F}_m$

With fixed $\mathbf{C}^{(m)}$, $\mathbf{S}$ and $\mathbf{F}$, the optimization problem can be reduced as

$$\min_{\mathbf{F}_m} \sum_{m=1}^{M} p_m \left\|\mathbf{F}_m\mathbf{F}_m^T-\mathbf{S}\right\|_F^2+\beta_3 Tr\left(\mathbf{F}_m^T\mathbf{L}_m\mathbf{F}_m\right),\ s.t.\ \mathbf{F}_m^T\mathbf{F}_m=\mathbf{I}_k. \tag{31}$$

Similarly, $\mathbf{F}_m$ can be updated iteratively as

$$\mathbf{F}_m^{(t+1)}=\mathbf{F}_m^{(t)}-\eta_{\mathbf{F}_m}\nabla_{\mathbf{F}_m}, \tag{32}$$

where the gradient can be written as

$$\nabla_{\mathbf{F}_m}=\sum_{m=1}^{M}2p_m\left(\mathbf{F}_m-\mathbf{S}^T\mathbf{F}_m-\mathbf{S}\mathbf{F}_m\right)+\beta_3\left(\mathbf{L}_m\mathbf{F}_m+\mathbf{L}_m^T\mathbf{F}_m\right). \tag{33}$$

So far, the update rules about all variables are provided above, these rules are then implemented repeatedly until the convergence or the achievement of maximum number of iteration. Furthermore, the update rules descripted above are summarized in the following algorithm.

Algorithm description ADMM

Input: $\mathbf{X}=\left\{\mathbf{X}\right\}_{i=1}^{M},\left\{\beta_i\right\}_{i=1}^{3},p_m,\eta,N,k,v$

Outputs: Assignment of the data points to k clusters

1: Initialize: $\left\{\mathbf{C}_i\right\}_{i=1}^{3}=\mathbf{0},\mathbf{W}=\mathbf{0},\left\{\mathbf{\Omega}_i\right\}_{i=1}^{3}=\mathbf{0},i=1,2,...,M$

2: **while** not converged **do**

3: **for** m=1 to M **do**

4: Fix others and update $\mathbf{W}$ by solving (18)

5: Fix others and update $\mathbf{C}_1$ by solving (19)

6: Fix others and update $\mathbf{C}_2$ by solving (24)

7: Fix others and update $\mathbf{C}_3$ by solving (25)

8: Fix others and update $\left\{\mathbf{\Omega}_i\right\}_{i=1}^{3}$ by solving (26)

9: Fix others and update $\mathbf{S}$ by solving (29)

10: Fix others and update $\mathbf{F}$ by solving (30)

11: Fix others and update $\mathbf{F}_m$ by solving (32)

12: **end for**

13: Update v

14:**end while**

15: Combine $\left\{\mathbf{C}^{(m)}\right\}_{m=1}^{M}$ by the method of spectral structure fusion

4.5 Complexity Analysis

The main computation consumption of the algorithm is the update of $\mathbf{C}^{(m)}$ and $\mathbf{F}$. To be specific, the complexity of updating $\mathbf{C}^{(m)}$ is $O(N^3)$, since the matrix inversion and multiplication. The complexity of updating $\mathbf{F}$ is $O(N^3)$, due to the implementation of SVD.

5. Experiments

5.1 Dataset Descriptions

To evaluate the performance of the proposed approach, four real-world multi-view datasets are used, which are **Reuters** [21], **3-sources**[1], **Prokaryotic** [22] and **UCI Digit**[2]. **Reuters** is a dataset containing documents in 5 languages, where 600 documents with 6 clusters are randomly sampled for this experiment. **3-sources** is a dataset of news articles collected from three online news sources, where 169 articles with 3 views and 6 clusters are adopted for this experiment. **Prokaryotic** is a dataset describing 551 prokaryotic species in a heterogeneous multi-view way including text and several genomic representations, where 551 samples with 3 views and 4 clusters are adopted for this experiment. **UCI Digit** is a dataset consisting of handwritten digits (0-9), where 2000 examples with 3 views and 10 clusters are chosen for this experiment.

5.2 Experiment Setting

We evaluate clustering performance using five different metrics, such as recall, precision, F-score, normalized mutual information (NMI) and adjusted rand index (Adj-RI). For all these metrics, the higher value implies better performance.

Besides, to evaluate the results of the experiments, we compare the proposed approach with the state-of-the-art solutions, such as Robust Multi-view Spectral Clustering (RMSC) [4], Convex Sparse Multi-view Spectral Clustering (CSMSC) [3], Pairwise Multi-view Low-rank Sparse Subspace Clustering (Pairwise MLRSSC) [5], Centroid-based Multi-view Low-rank Sparse Subspace Clustering (Centroid MLRSSC) [5] and Multi-graph Fusion for Multi-view Spectral Clustering (GFSC) [9]. Apart from that, we also verify the effectiveness of the proposed rank-norm approximation. For the sake of fairness, the same framework under different rank-norm approximations are compared. In other words, the framework of unifying multi-view data on the spectral structure under the nuclear norm (MVSS (Nuclear)), the Gamma norm (MVSS (Gamma)) and the proposed rank-norm approximation (MVSS (Ours)) are compared. Especially in MVSS (Nuclear), to compare with the cutting-edge method, the weighted nuclear norm [23], the latest nuclear norm based method, is exploited.

Moreover, all parameters of existing approaches mentioned above are based on the respective parameter-searching strategy provided by the authors. Besides, all balance parameters in the proposed method are tuned from $\{50^{-3}, 10^{-3}, 50^{-2}, 10^{-2}, 50^{-1}, 10^{-1}, 1, 5, 10\}$. In addition, the parameter for the proposed rank-norm approximation ε is tuned from $\{50^{-3}, 10^{-3}, 50^{-2}, 10^{-2}, 50^{-1}\}$, while the parameter for the Gamma norm γ is varied from

[1] http://mlg.ucd.ie/datasets/3sources.html
[2] http://archive.ics.uci.edu/ml/datasets/Multiple+Features

50^{-3} to 10^{-1} with step 50^{-3}. Apart from that, the learning rate is set as 0.05, and the maximum number of iteration is 300.

5.3 Experiment Results

In **Table 1**, the clustering performance is compared sufficiently. As shown in **Table 1**, the performance of MVSS series (including MVSS (Nuclear), MVSS (Gamma), MVSS (Ours)) generally tend to outperform the performance of other algorithms. Hence, the effectiveness of the proposed integration of multi-view data on the spectral structure can be verified.

Table 1. Clustering performance of different algorithms

Datasets	Method	Precision	Recall	F-score	NMI	Adj-RI
Returns	RMSC	0.325	0.412	0.361	0.297	0.217
	CSMSC	0.327	0.420	0.365	0.295	0.220
	Pairwise MLRSSC	0.389	**0.486**	0.428	0.390	0.300
	Centroid MLRSSC	0.395	0.482	0.432	0.394	0.306
	GFSC	0.396	0.447	0.417	0.381	0.286
	MVSS (Nuclear)	0.398	0.471	0.431	0.389	0.310
	MVSS (Gamma)	0.412	0.477	0.442	**0.404**	0.315
	MVSS (Ours)	**0.414**	0.481	**0.444**	0.400	**0.317**
3-sources	RMSC	0.515	0.453	0.477	0.517	0.330
	CSMSC	0.518	0.464	0.482	0.518	0.335
	Pairwise MLRSSC	0.707	0.619	0.659	0.594	0.565
	Centroid MLRSSC	0.696	0.619	0.654	0.595	0.557
	GFSC	0.695	0.612	0.647	0.586	0.566
	MVSS (Nuclear)	0.757	0.675	0.713	0.617	0.585
	MVSS (Gamma)	**0.767**	0.679	**0.720**	0.623	0.594
	MVSS (Ours)	0.762	**0.682**	0.719	**0.625**	**0.597**
Prokaryotic	RMSC	0.567	0.369	0.447	0.315	0.198
	CSMSC	0.565	0.391	0.462	0.269	0.206
	Pairwise MLRSSC	0.624	0.566	0.591	0.322	0.345
	Centroid MLRSSC	0.530	**0.756**	0.574	0.202	0.258
	GFSC	0.626	0.583	0.600	0.325	0.348
	MVSS (Nuclear)	0.631	0.635	0.632	0.328	0.351
	MVSS (Gamma)	**0.638**	0.644	0.640	0.331	0.354
	MVSS (Ours)	0.633	0.653	**0.642**	**0.333**	**0.357**
UCI digit	RMSC	0.728	0.757	0.742	0.778	0.713
	CSMSC	0.725	0.836	0.775	0.819	0.748
	Pairwise MLRSSC	0.809	0.854	0.830	0.851	0.810
	Centroid MLRSSC	0.819	0.854	0.835	0.854	0.817
	GFSC	0.821	0.851	0.831	0.862	0.812
	MVSS (Nuclear)	0.825	0.858	0.841	0.861	0.819
	MVSS (Gamma)	0.831	0.863	0.846	**0.867**	**0.828**
	MVSS (Ours)	**0.832**	**0.865**	**0.848**	0.862	0.823

Moreover, one may observe that, the performance of MVSS series is much better than that of others in 3-sources, while the performance of MVSS series is just slightly better than that of others in Reuters. For example, F-score of MVSS series is obviously higher than that of the other approaches in 3-sources, while F-score of MVSS series is slightly higher than that of the other approaches in Reuters. This is because that, the divergence across each view of data in 3-sources is more obvious than that in Reuters. Besides, in the proposed MVSS series, one

may find that, the clustering performance of MVSS (Ours), which is based on the proposed rank-norm approximation, is almost the same as it of MVSS (Gamma). Hence, the proposed rank-norm approximation can achieve almost the same performance of the rank function approximation via the Gamma norm. Nevertheless, to achieve the same performance, the method based on the Gamma norm may require finer-grained grid search for the parameter γ, compared with the parameter ε in the proposed rank-norm approximation. This is because that, better robustness can be derived by the proposed rank-norm approximation.

Table 2. F-score under different *f*

Datasets	Approach	F-score under different *f*			
		f=0%	*f*=5%	*f*=10%	*f*=15%
Returns	MVSS (Gamma)	0.442	0.423	0.408	0.387
	MVSS (Ours)	**0.444**	**0.435**	**0.425**	**0.413**
3-sources	MVSS (Gamma)	**0.720**	0.701	0.688	0.671
	MVSS (Ours)	0.719	**0.710**	**0.701**	**0.692**
Prokaryotic	MVSS (Gamma)	0.640	0.623	0.605	0.567
	MVSS (Ours)	**0.642**	**0.635**	**0.621**	**0.606**
UCI digit	MVSS (Gamma)	0.846	0.826	0.808	0.785
	MVSS (Ours)	**0.848**	**0.839**	**0.828**	**0.818**

Table 3. NMI under different *f*

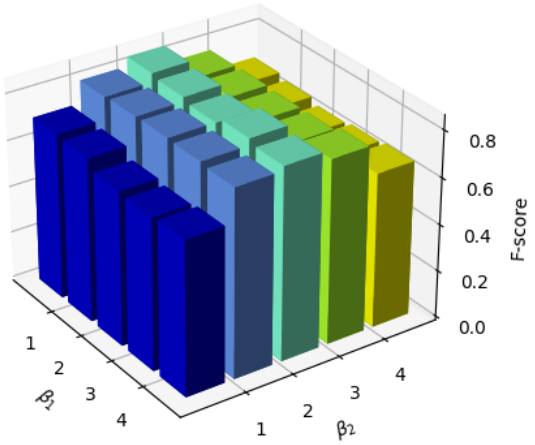
Datasets	Approach	NMI under different *f*			
		f=0%	*f*=5%	*f*=10%	*f*=15%
Returns	MVSS (Gamma)	**0.404**	0.387	0.371	0.364
	MVSS (Ours)	0.400	**0.395**	**0.388**	**0.379**
3-sources	MVSS (Gamma)	0.623	0.612	0.601	0.588
	MVSS (Ours)	**0.625**	**0.619**	**0.611**	**0.605**
Prokaryotic	MVSS (Gamma)	0.331	0.318	0.305	0.292
	MVSS (Ours)	**0.333**	**0.325**	**0.317**	**0.311**
UCI digit	MVSS (Gamma)	**0.867**	**0.858**	0.841	0.832
	MVSS (Ours)	0.862	0.856	**0.850**	**0.845**

Table 4. Adj-RI under different *f*

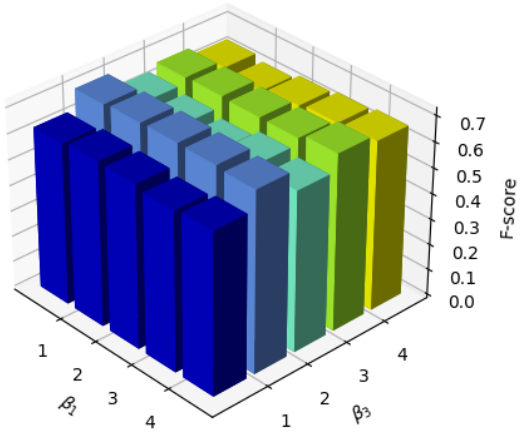
Datasets	Approach	Adj-RI under different *f*			
		f=0%	*f*=5%	*f*=10%	*f*=15%
Returns	MVSS (Gamma)	0.315	0.304	0.293	0.284
	MVSS (Ours)	**0.317**	**0.311**	**0.302**	**0.296**
3-sources	MVSS (Gamma)	0.594	0.584	0.573	0.564
	MVSS (Ours)	**0.597**	**0.592**	**0.586**	**0.579**
Prokaryotic	MVSS (Gamma)	0.354	0.347	0.338	0.327
	MVSS (Ours)	**0.357**	**0.350**	**0.344**	**0.338**
UCI digit	MVSS (Gamma)	**0.828**	**0.817**	0.805	0.796
	MVSS (Ours)	0.823	0.816	**0.811**	**0.804**

To further compare the robustness of the Gamma norm with that of the proposed method, **Table 2-4** show how the change of the parameters of rank-norm approximation methods (i.e., ε in the proposed rank-norm approximation or γ in the Gamma norm) impacts on clustering performance. As the length limit, the comprehensive performance metric, F-score, NMI, Adj-RI, are used here only. In **Table 2-4**, the change of the parameters of rank-norm approximation methods is denoted by *f* which means the percentage change of the parameters

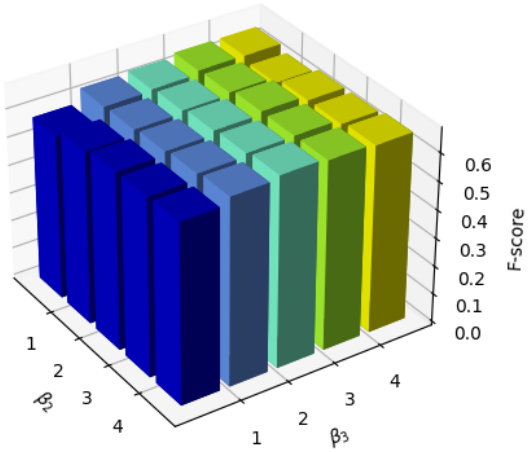
compared with the value chosen in **Table 1**. For example, $f = 5\%$ means that, the parameters are increased and decreased by 5% from the value chosen in **Table 1**, and the corresponding metric in **Table 2-4** is the highest one among the two F-score values. It can be seen that, the impact of the variation of f on the performance of MVSS (Ours) is smaller than its counterpart on the performance of MVSS (Gamma). So, it can be seen that the Gamma norm is more sensitive to the parameter γ. Therefore, we can conclude that, the proposed rank-norm approximation can provide a better robustness than the Gamma norm.



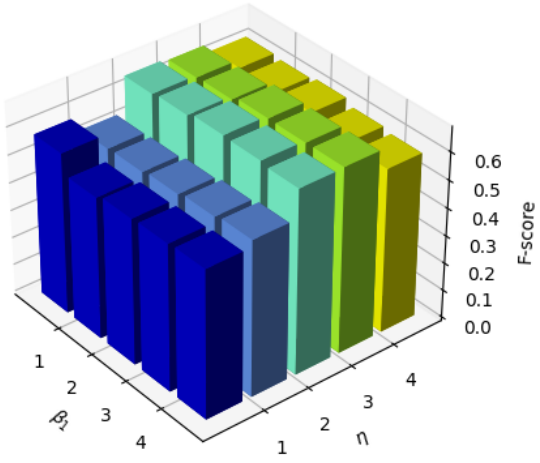
(a)



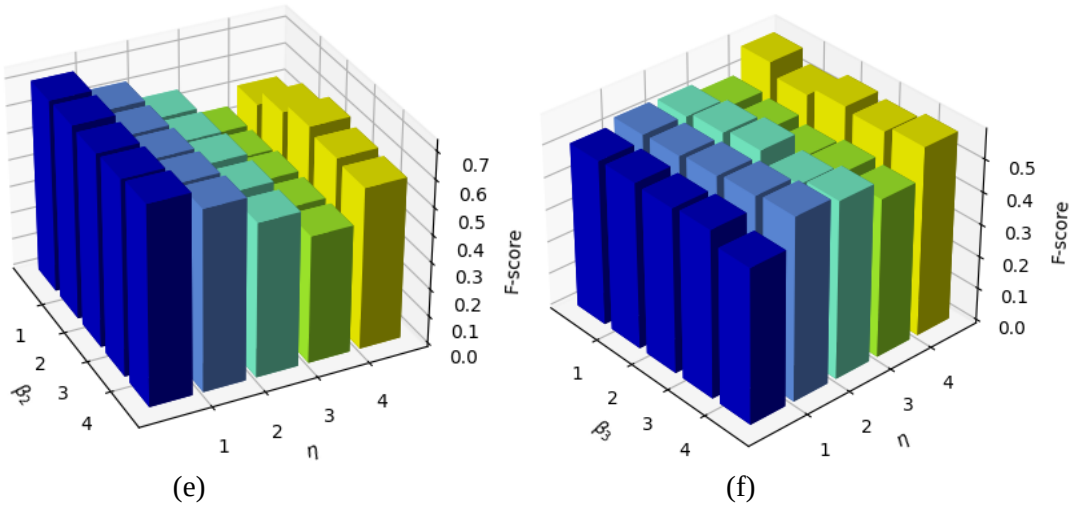
(b)



(c)



(d)



(e) (f)

Fig. 1. Illustration of the parameter sensitivity of the proposed method on UCI digit dataset (a) β_1 vs β_2. (b) β_1 vs β_3. (c) β_2 vs β_3. (d) β_1 vs η. (e) η vs β_2. (f) η vs β_3.

The analysis of parameter sensitivity is shown in **Fig. 1**, where different combinations of two balance parameters chosen from [1, 2, 3, 4] are conducted the analysis and the F-score is taken as the performance metric. As shown in **Fig. 1**, under different parameter perturbations, the proposed method can obtain stable performance. In contrast, the performance of the proposed method is relatively sensitive to the parameter combination of η and β_2.

6. Conclusion

In the multi-view clustering, to solve the performance deterioration problem caused by the huge diversity of the data representation in each view, we propose to construct the unified representation in the spectral embedding domain. In this way, the complementary information between perspectives can be unified very well. In addition, to better capture the global structure information of data in subspace clustering, we propose a novel low-rank constraint relaxation method via the tight lower bound on the rank function and develop an optimization algorithm to solve the multi-view clustering problem. Finally, experimental results demonstrate that the proposed method has better tradeoff between clustering effect and robustness.

References

[1] C. Xu, D. Tao and C. Xu, "A survey on multi-view learning," *arXiv preprint arXiv:1304.5634*, Apr. 2013. Article (CrossRef Link)

[2] G. Chao, S. Sun, and J. Bi, "A survey on multi-view clustering," *arXiv preprint arXiv:1712.06246*, Dec. 2017. Article (CrossRef Link)

[3] C. Lu, S. Yan and Z. Lin, "Convex sparse spectral clustering: Single-view to multi-view," *IEEE Trans. Image Process.*, vol. 25, no. 6, pp. 2833-2843, Apr. 2016. Article (CrossRef Link)

[4] R. K. Xia, P. Yan, D. Lei and Y. Jian, "Robust multi-view spectral clustering via low-rank and sparse decomposition," in *Proc of the AAAI conference on artificial intelligence*, Québec, Canada, vol. 28, no. 1, pp. 2149-2155, Jun. 2014. Article (CrossRef Link)

[5] M. Brbić, I. Kopriva, "Multi-view low-rank sparse subspace clustering," *Pattern Recognit.*, vol. 73, pp. 247-258, Jan. 2018. Article (CrossRef Link)
[6] X. Cao, C. Zhang, H. Fu, S. Liu and H. Zhang, "Diversity-induced multi-view subspace clustering," in *Proc of the IEEE conference on computer vision and pattern recognition*, pp. 586-594, 2015. Article (CrossRef Link)
[7] X. Wang, X. Guo, Z. Lei, C. Zhang and S. Z. Li, "Exclusivity-consistency regularized multi-view subspace clustering," in *Proc of the IEEE conference on computer vision and pattern recognition*, pp. 923-931, 2017. Article (CrossRef Link)
[8] F. Nie, J. Li and X. Li, "Self-weighted Multiview Clustering with Multiple Graphs," in *Proc of the Twenty-Sixth International Joint Conference on Artificial Intelligence*, pp. 2564-2570, 2017. Article (CrossRef Link)
[9] Z. Kang, G. Shi, S. Huang, W. Chen, X. Pu, J. T. Zhou and Z. Xu, "Multi-graph fusion for multi-view spectral clustering," *Knowl. Based Syst.*, vol. 189, pp. 105102-105110, Feb. 2020. Article (CrossRef Link)
[10] 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.*, vol. 35, no. 1, pp. 171-184, Jan. 2013. Article (CrossRef Link)
[11] S. Liu, X. Liu, S. Wang and K. Muhammad, "Fuzzy-aided solution for out-of-view challenge in visual tracking under iot-assisted complex environment," *Neural. Comput. Appl.*, vol. 33, pp.1055–1065, Feb. 2021. Article (CrossRef Link)
[12] K. Fan, "On a theorem of Weyl concerning eigenvalues of linear transformations I," in *Proc of the National Academy of Sciences of the United States of America*, vol. 35, no. 11, pp. 652-655, Nov. 1949. Article (CrossRef Link)
[13] J. Cao, Y. Fu, X. Shi and B. W. Ling, "Subspace Clustering Based on Latent Low Rank Representation with Schatten-*p* Norm," in *Proc of the 2020 2nd World Symposium on Artificial Intelligence*, Guangzhou, China, pp. 58-62, Jul. 2020. Article (CrossRef Link)
[14] H. Zhang, J. Yang, J. Xie, J. Qin and B. Zhang, "Weighted sparse coding regularized nonconvex matrix regression for robust face recognition," *Inf. Sci.*, vol. 394, pp. 1-17, Jul. 2017. Article (CrossRef Link)
[15] S. Wang, D. Liu, and Z. Zhang, "Nonconvex relaxation approaches to robust matrix recovery," in *Proc of the Twenty-Third International Joint Conference on Artificial Intelligence*, pp. 1764-1770, Aug. 2013. Article (CrossRef Link)
[16] S. Liu, S. Wang, X. Liu, A. H. Gandomi, M. Daneshmand, K. Muhammad and V. H. C. De Albuquerque, "Human memory update strategy: a multi-layer template update mechanism for remote visual monitoring," *IEEE Trans. Multimedia*, vol. 23, pp. 2188-2198, Mac. 2021. Article (CrossRef Link)
[17] Y. Wang, L. Wu, X. Lin and J. Gao, "Multiview spectral clustering via structured low-rank matrix factorization," *IEEE Trans. Neural Networks Learn. Syst.*, vol. 29, no. 10, pp. 4833-4843, 2018. Article (CrossRef Link)
[18] S. Liu, S. Wang, X. Liu, C. T. Lin and Z. Lv, "Fuzzy detection aided real-time and robust visual tracking under complex environments," *IEEE Trans. Fuzzy Syst.*, vol. 29, pp. 90-102, Jun. 2020. Article (CrossRef Link)
[19] I, Daubechies, M. Defrise and C. De Mol, "An iterative thresholding algorithm for linear inverse problems with a sparsity constraint," *Commun. Pure Appl. Math.*, vol. 57, no. 11, pp. 1413-1457, Aug. 2004. Article (CrossRef Link)
[20] D. L. Donoho, "De-noising by soft-thresholding," *IEEE Trans. Inf. Theory*, vol. 41, no. 3, pp. 613-627, May 1995. Article (CrossRef Link)
[21] D. D. Lewis, Y. Yang, T.G. Rose and F. Li, "Rcv1: A new benchmark collection for text categorization research," *J. Mach. Learn. Res.*, vol. 5, pp. 361-397, Apr. 2004. Article (CrossRef Link)
[22] M. Brbić, M. Piškorec, V. Vidulin, A. Kriško, T. Šmuc and F. Supek, "The landscape of microbial phenotypic traits and associated genes," *Nucleic Acids Res.*, vol. 44, no. 21, pp. 10074-10090, Dec. 2016. Article (CrossRef Link)

[23] S. Gu, Q. Xie, D. Meng, W. Zuo, X. Feng and L. Zhang, "Weighted nuclear norm minimization and its applications to low level vision," *Int. J. Comput. Vis.*, vol. 121, no. 2, pp. 183-208, Jan. 2017. Article (CrossRef Link)

Yin Long. He received his Ph.D degree in Electronic Engineering from University of Electronic Science and Technology of China, China, in 2018.
He was a visiting Ph.D student in the University of Alberta, Canada from 2015-2017. He is currently an associate professor in the School of Computer Science and Technology at Southwest University of Science and Technology, China. His primary research is in the area of subspace learning, manifold learning and data mining.

Xiaobo Liu. He is currently a M.S. graduate student in the School of Computer Science and Technology at Southwest University of Science and Technology, China, and majored in Software Engineering. His primary research is in the area of subspace learning, manifold learning and data mining.

Simon Murphy. He received his Ph.D degree in Applied Mathematics from McMaster University, Canada, in 2015. He was a Postdoctoral Fellow in the University of Waterloo, Canada from 2015-2018. He is currently a Postdoctoral Fellow in the School of Computer Science at University of Alberta, Canada. His primary research is in the area of machine learning and data mining.