2021 IEEE International Conference on Data Mining (ICDM)

Outlier-Robust Multi-View Subspace Clustering with Prior Constraints

Mehrnaz Najafi
LinkedIn
San Francisco, CA, USA
mnajafi@linkedin.com

Lifang He
Department of Computer Science & Engineering
Lehigh University
Bethlehem, PA, USA
lih319@lehigh.edu

Philip S. Yu
Department of Computer Science
University of Illinois at Chicago
Chicago, IL, USA
psyu@uic.edu

2021 IEEE International Conference on Data Mining (ICDM) | 978-1-6654-2398-4/21/$31.00 ©2021 IEEE | DOI: 10.1109/ICDM51629.2021.00055

***Abstract*—Data may have multiple modalities, known as multi-view data. With the assumption that multi-view data often lie on a latent subspace, multi-view subspace clustering finds the underlying subspace by leveraging multiple views and clusters the data accordingly. Due to inevitable system errors, multi-view data may contain outliers and it may not therefore strictly follow subspace structure. Besides, prior information such as pairwise constraints describing relations between data instances is often available. These constraints provide a valuable guide on learning. Unfortunately, standard multi-view subspace clustering methods do not simultaneously exploit high order correlations among views and prior constraints with low computational complexity. In this paper, we propose a novel Robust Multi-View Subspace Clustering method, named as RMVSC, which is capable of taking advantage of high order correlations among views and prior constraints for outlier-robust multi-view subspace clustering with low computational complexity. The key idea is to use a low-rank tensor along with a constraint to integrate information from views and prior constraints for more comprehensive learning. We regard underlying clean subspace of singular vectors of views (leveraging views) which also represent projection coefficient of cluster membership vectors in data space (utilizing prior constraints) as a tensor. By decomposing singular vector of each view into its underlying clean subspace and a structured-sparse error (outlier) term, we characterize outliers explicitly. To solve the challenging optimization problem, we develop an algorithm based on Augmented Lagrangian Multiplier. Experimental results on real-world datasets show the superiority of the proposed method and its robustness against outliers.**

***Keywords*-Outlier, Multi-view, Subspace clustering**

I. Introduction

Data is often determined and observed by different views/perspectives, known as *multi-view data*. Multi-view data often come from diverse and different sources. Each individual view may be comprised of arbitrary type and number of features [1]. For instance, an image may be described by different feature types such as color, edge and texture where each feature type is a view of data. While these individual views could be sufficient for learning, they often provide complementary and consistent information to each other [1]. Thus, it is advantageous to combine multiple views for more comprehensive learning. Besides, due to sensor failures and system errors, each view may be *erroneous*, where error refers to the deviation between model assumption and data [2]. Error could exhibit as *outliers*, which refer to large deviations from real data values.

Multi-view clustering is one of the important tasks in multi-view learning. The key problem is how to exploit complementary and consistent information from multiple views rather than using a single view to boost clustering performance [1], [3], [4]. One of the most common algorithms used for multi-view clustering is *multi-view subspace clustering*, which is based on the assumption that multi-view data lie on a latent subspace [5]–[9]. Following this assumption, the key idea is to first uncover underlying subspace and then conduct clustering accordingly. To capture high order correlations underlying multiple views for more comprehensive learning, one of the most promising approaches is to use *tensor* [7]–[9]. Tensor is a generalization of vectors (first order tensors) and matrices (second order tensors). Zhang et al. equipped a tensor established by stacking subspace matrices of all views with a low-rank constraint, while decomposing each individual view into the underlying clean subspace and error matrix that encodes the error in the view [7]. Different from [7], Xie et al. imposed more efficient and effective tensor low-rank constraint to exploit high order correlations in data while following the same decomposition [8].

Lack of label information to guide the learning process makes clustering task much harder. Thus, it is important to utilize *prior information* such as *prior pairwise constraints* that describe relationship between data instances when learning representations. Prior pairwise constraints discover valuable information about structure of data, i.e., similarity between data instances and therefore act as a powerful guidance for learning. Prior pairwise constrains often consist of two parts: *must-link* and *cannot-link* constraints. A must-link constraint imposes the same cluster (or class) membership on the pair of data instances, while a cannot-link constraint specifies that pair of data instances should not be assigned to the same cluster (or class). For example, for object clustering, we may know as prior information that two data instances must belong to the same cluster and we thus the pairwise similarity between them equals to 1.

To incorporate prior pairwise constraints into multi-view clustering, Zhao et al. proposed an approach named as

978-1-6654-2398-4/21/$31.00 ©2021 IEEE
DOI 10.1109/ICDM51629.2021.00055

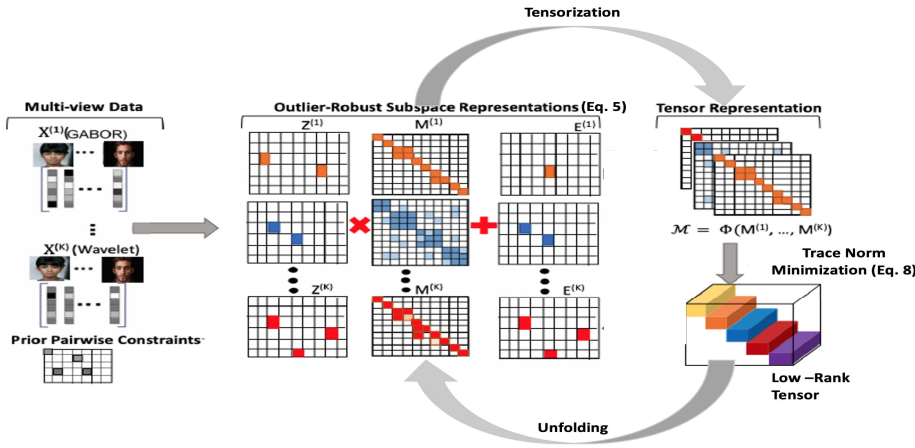


Figure 1: Overview of the proposed RMVSC method: Given a collection of data instances with multiple views $\mathbf{X}^{(1)}$, ..., $\mathbf{X}^{(K)}$ where K indicates number of views, RMVSC integrates all underlying clean subspaces $\mathbf{M}^{(1)}$, ..., $\mathbf{M}^{(K)}$ which is constructed by separating error terms $\mathbf{E}^{(1)}$, ..., $\mathbf{E}^{(K)}$ from singular vectors of views $\mathbf{Z}^{(1)}$, ..., $\mathbf{Z}^{(K)}$ into a low-rank tensor $\mathcal{M}$ to capture high order correlations underlying data under prior pairwise constraints to jointly promote the subspace representations.

MVMC [4]. Initially, MVMC builds pairwise similarity matrices for views and enforces prior pairwise constraints i.e., must-link and cannot-link constraints on the corresponding entries in the similarity matrices. It then learns shared similarity matrix across all views by casting multi-view clustering task into similarity matrix completion problem. MVMC suffers from the following drawbacks: (1) it fails to exploit high order correlations underlying the multi-view data because it only captures pairwise correlations in data, (2) MVMC concentrates only on clean data, which makes it inadmissible for outlier-corrupted multi-view data to a large extent. Existing multi-view clustering methods that take prior pairwise constraints into account focus only on clean multi-view data and do not explore high order correlations in data. Besides, existing tensorized multi-view subspace clustering methods do not utilize clues from prior pairwise constraints when learning representations for clustering. These approaches have high computational complexity i.e., $O(N^3)$ where N refers to the number of data instances. It is thus unrealistic to apply them on large multi-view datasets. All the challenges lead to the problem of outlier-robust multi-view clustering with prior pairwise constraints.

To utilize clues from prior pairwise constraints for more comprehensive multi-view subspace clustering, we propose a novel method for outlier-Robust Multi-View Subspace Clustering with prior pairwise constraints (RMVSC) that only requires low computational complexity. Unlike the existing tensorized multi-view subspace clustering methods that regard all subspace representations of views as a tensor, RMVSC aims to reduce computational complexity while preserving clustering performance by establishing a tensor with subspace of singular vectors of views, as shown in Fig. 1. The subspace of singular vectors of views requires low computational complexity compared with subspace of views, yielding a more compact representation of multi-view data. Inspired by the idea of robust subspace recovery [2] that decomposes possibly erroneous data into clean data as well as error term that encodes error in data, these subspaces are learned by a joint decomposition of singular vector of each individual view into the underlying clean subspace and structured-sparse error (outlier) term with enforcement of low-rank constraint on the tensor built on underlying clean subspaces. The low-rank constraint also decreases the redundancy of the learned subspaces.

To integrate information from views and prior constraints for comprehensive learning, we use the subspaces as dot product of projection coefficient of cluster membership vectors in data space to find clustering solution. We also introduce a constraint into the objective function to enforce prior constraints. The proposed RMVSC method formulates the problem of multi-view subspace clustering with prior pairwise constraints as a joint tensor rank minimization with $\ell_{2,1}$ regularization term along with view decomposition. To further improve robustness of RMVSC against outliers, we impose $\ell_{2,1}$ norm on the error matrix. Since error matrix with outliers has sparse row structure, we use $\ell_{2,1}$ norm to characterize this sparsity property.

Since multi-view data is often collected from different sources, error matrices of views may have inconsistent magnitude values and can therefore significantly drop clustering performance. As suggested by [2], we impose a constraint on column of error matrices to jointly have consistent magnitude values by vertical concatenation of error matrices of all views. Similar to prevalent clustering approaches [4], [8], we use a two stage framework for clustering: we first learn affinity matrix for similarity matrices of views

Table I: List of notations

Symbol	Definition and description
X	each uppercase letter represents a scale
$\mathbf{x}$	each boldface lowercase letter represents a vector
$\mathbf{X}$	each boldface uppercase letter represents a matrix
$\mathcal{X}$	each calligraphic letter represents a tensor
$\mathbf{X}_{i,j}$	the (i,j)-th entry of $\mathbf{X}$
$\mathbf{x}_i$	the i-th entry of $\mathbf{x}$
$\mathcal{X}(:,:,i)$	the i-th frontal slice of $\mathcal{X}$
$\|\|\mathbf{X}\|\|_2$	ℓ_2-norm of matrix $\mathbf{X}$
$\|\|\mathbf{X}\|\|_F$	Frobenius norm of matrix $\mathbf{X}$, defined as $\|\|\mathbf{X}\|\|_F = \sqrt{\sum_{i,j} A_{i,j}^2}$
$\|\|\mathbf{X}\|\|_{2,1}$	$\ell_{2,1}$ norm of matrix $\mathbf{X}$, defined as $\|\|\mathbf{X}\|\|_{2,1} = \sum_j \|\|\mathbf{X}(:,j)\|\|_2$
$\|\|\mathbf{X}\|\|_{tr}$	The trace norm of matrix $\mathbf{X}$, defined as sum of the singular values of $\mathbf{X}$
$\|\|\mathcal{X}\|\|_{tr}$	$\|\|\mathcal{X}\|\|_{tr} = \sum_v \|\|\mathcal{X}(:,:,v)\|\|_{tr}$

and then apply a traditional clustering algorithm such as k-means to obtain the final clustering solution. Notably, importance of clustering based on similarity between data instances has been well recognized in the literature [10]. Decomposing each individual view into underlying clean subspace and error term as well as imposing $\ell_{2,1}$ norm on the error matrix make objective function of the proposed RMVSC method challenging to optimize. We present a new efficient optimization algorithm based on Augmented Lagrangian Multiplier [11], [12] to solve it. We also provide computational complexity and convergence analysis of the optimization solution. Our main contributions are as follows:

- To the best of our knowledge, RMVSC is the first work for outlier-robust multi-view subspace clustering that ties together high order correlations underlying multi-view data as well as prior information and harnesses them under a unified framework with low computational complexity. Different from existing tensorized multi-view subspace clustering algorithms, the proposed RMVSC method aims to reduce computational complexity by stacking underlying clean subspace of singular vectors of all views as a tensor.
- We propose a new efficient optimization algorithm to solve the objective function along with convergence and computational complexity analysis.
- Through extensive experiments on real-world datasets, we show the proposed RMVSC method outperforms several state-of-the-art multi-view clustering algorithms and is robust against error due to outliers in data.

II. Preliminaries

Table I summarizes the notations. The N^{th} order tensor $\mathcal{X}$ is defined as $\mathcal{X} \in \mathbb{R}^{I_1 \times I_2 \times ... \times I_N}$ (a tensor of size $I_1 \times I_2 \times ... \times I_N$). A *slice* of a tensor is a 2D section formulated by fixing all but two indices where $\mathcal{X}(:,:,i)$ denotes i-th frontal slices. The trace norm of matrix $\mathbf{X}$ is $||\mathbf{X}||_{tr} = \sum_i \sigma_i(\mathbf{X})$ where $\sigma_i(\mathbf{X})$ indicates the i-th largest singular value. The trace norm of tensor $\mathcal{X}$ is $||\mathcal{X}||_{tr} = \sum_i ||\mathcal{X}(:,:,i))||_{tr}$.

III. Outlier-Robust Multi-View Subspace Clustering with Prior Information

We are given N distinct data instances with K related views, the possibly erroneous views are denoted as $\mathbf{X}^{(1)} \in \mathbb{R}^{N \times I_1}$, ..., $\mathbf{X}^{(K)} \in \mathbb{R}^{N \times I_K}$ for which I_i indicates number of features in i-th view such that $1 <= i <= K$ and all views are generated from latent subspace. Let C indicate number of clusters. Prior pairwise constraints of type of must-link $\mathbf{A} \in \mathbb{R}^{N \times N}$ where $\mathbf{A}_{i,j} = 1$ if data instances i and j must belong to the same cluster, and cannot-link $\mathbf{B} \in \mathbb{R}^{N \times N}$ where $\mathbf{B}_{i,j} = 1$ if data instances i and j should not be assigned to the same cluster are also given. Our goal is to partition all the N data instances into C clusters while recovering underlying clean subspace of all views.

We assume that the features in each individual view are sufficient for obtaining most of the clustering information (*complementary*). Also, the true underlying clustering would assign corresponding data instances in each view to the same cluster (*consistency*). Each view might be erroneous. Besides, cluster membership vectors lie in the subspace of the first P left singular vectors of feature matrices of all views where $P >= C$. The last assumption used in [4], [5], [13] essentially implies that all the cluster membership assignments can be accurately predicted by a linear combination of feature vectors of instances. Span of vectors is defined as a set of all linear combinations of them. Mathematically, span of all the cluster membership assignments is subset of span of the first P left singular vectors of feature matrices of all views.

Our aim is to address non-overlapping clusters due to high popularity of multi-view datasets with such characteristics. We assume each data instance belongs to *only* one cluster, but our method could be easily adjusted for the overlapping clusters by using traditional similarity-based overlapping clustering approaches [14]. In the following, we describe how to derive the proposed objective function.

A. Pairwise Similarity Matrix Construction

Prior pairwise constraints provide a valuable guide on learning the subspace because they reveal information on the structure of the data, i.e., relationship between data instances. In order to utilize clues from prior pairwise constraints, we establish pairwise similarity matrix with respect to each individual view based on must-link and cannot-link constraints $\mathbf{A}$ and $\mathbf{B}$. Let $\mathbf{u}_c^{(v)} \in \mathbb{R}^N$ be the membership vector of the c-th cluster in the v-th view where $\mathbf{u}_{c,i}^{(v)} = 1$, if data instance i is assigned to the c-th cluster and zero, otherwise. The pairwise similarity matrix $\mathbf{S}^{(v)} \in \mathbb{R}^{N \times N}$ can be then formulated as $\mathbf{S}^{(v)} = \sum_{c=1}^{C} \mathbf{u}_c^{(v)} (\mathbf{u}_c^{(v)})^\top$ where $\mathbf{S}_{i,j}^{(v)} = 1$, if data instances i and j belong to the same cluster in v-th view, and zero, otherwise. Specifically, $\mathbf{S}_{i,j}^{(v)} = 1$ if $\mathbf{A}_{i,j} = 1$ and $\mathbf{S}_{i,j}^{(v)} = 0$ if $\mathbf{B}_{i,j} = 1$.

Let $\mathbf{S}_{gt}$ denote groundtruth pairwise similarity matrix, $\mathbf{S}_{ob}$ indicate partial observations of $\mathbf{S}_{gt}$ defined as follows:

$$(i,j)\text{-th entry of } \mathbf{S}_{ob} = \begin{cases} 1 & \text{if } \mathbf{A}_{i,j} = 1, \\ 0 & \text{if } \mathbf{B}_{i,j} = 1. \end{cases} \quad (1)$$

$\mathbf{\Omega} \in \mathbb{R}^{N\times N}$ is a binary matrix (observation matrix) with the same size as $\mathbf{S}_{gt}$, and its entries indicate whether each corresponding entry in $\mathbf{S}_{gt}$ is observed or not based on the prior pairwise constraints. In other words, $\mathbf{\Omega}_{i,j} = 1$ if the corresponding entry in $\mathbf{S}_{ob}$ is observed and $\mathbf{\Omega}_{i,j} = 0$, otherwise. * for matrices denotes the element-wise product. $\mathbf{S}_{ob}$ can be obtained by $\mathbf{\Omega} * \mathbf{S}_{gt} = \mathbf{S}_{ob}$.

To efficiently exploit feature matrix $\mathbf{X}^{(v)}$ for each individual view v, we define $\mathbf{Z}^{(v)} = [\mathbf{z}_1^{(v)}, ..., \mathbf{z}_P^{(v)}] \in \mathbb{R}^{N\times P}$ to denote the first P left singular vectors of $\mathbf{X}^{(v)}$ corresponding to the P largest singular values, where $P >= C$. As mentioned earlier, we assume that the cluster membership vectors $\mathbf{u}_1^{(v)}, \cdots$ and $\mathbf{u}_C^{(v)}$ lie in the subspace of the first P left singular vectors of feature matrix $\mathbf{Z}^{(1)}, \cdots$ and $\mathbf{Z}^{(K)}$. We can then derive Span$(\mathbf{u}_1^{(v)}, \cdots, \mathbf{u}_C^{(v)}) \subseteq$ Span$(\mathbf{z}_1^{(v)}, \cdots, \mathbf{z}_P^{(v)})$.

According to the span definition, cluster membership vector is defined as linear combination of left singular vectors. $\forall i = 1, ..., C$, $\mathbf{u}_i^{(v)} = \mathbf{Z}^{(v)}\boldsymbol{\theta}_i^{(v)}$ where $\boldsymbol{\theta}_i^{(v)} \in \mathbb{R}^P$ denotes projection coefficients of cluster membership vectors in view v space. The similarity matrix $\mathbf{S}^{(v)}$ can be thus written as follows:

$$\mathbf{S}^{(v)} = \sum_{c=1}^{C} \mathbf{u}_c^{(v)}(\mathbf{u}_c^{(v)})^\top = \sum_{c=1}^{C} \mathbf{Z}^{(v)}\boldsymbol{\theta}_c^{(v)}(\mathbf{Z}^{(v)}\boldsymbol{\theta}_c^{(v)})^\top = \mathbf{Z}^{(v)}\mathbf{M}^{(v)}(\mathbf{Z}^{(v)})^\top \quad (2)$$

where $\mathbf{M}^{(v)} = \sum_{c=1}^{C} \boldsymbol{\theta}_c^{(v)}(\boldsymbol{\theta}_c^{(v)})^\top \in \mathbb{R}^{P\times P}$. Since $\mathbf{M}^{(v)}$ is a symmetric positive semidefinite matrix which ensures that all of its eigenvalues are non-negative, it can be interpreted as covariance matrix that provides direction of linear relationship between projection coefficients of cluster membership vectors in each individual view space.

B. Pairwise Similarity Matrix Completion

The importance of considering similarity matrix or relationship among data instances has been well recognized by previous literature and clustering using similarity matrix has caught significant attention [10]. To approximate similarity matrices $\mathbf{S}^{(v)}$ while reducing their complexity, we minimize $rank(\mathbf{S}^{(v)})$, where $rank(\cdot)$ denotes the low-rank approximation of a true density matrix (i.e., low-rankness criterion). However, since rank($\mathbf{S}^{(v)}$) is non-convex, we replace rank($\mathbf{S}^{(v)}$) with the trace norm $||\mathbf{S}^{(v)}||_{tr}$ as a convex envelope to the matrix rank [15]. As $\mathbf{Z}^{(v)}$ and $(\mathbf{Z}^{(v)})^\top$ are orthogonal matrices, from Eq. (2), we can thus write the following equation for each $\mathbf{S}^{(v)}$ [13]:

$$||\mathbf{S}^{(v)}||_{tr} = ||\mathbf{Z}^{(v)}\mathbf{M}^{(v)}(\mathbf{Z}^{(v)})^\top||_{tr} = ||\mathbf{M}^{(v)}||_{tr} \quad (3)$$

To explore complementary and high order correlations underlying multiple views as well as information of each view, we merge $\mathbf{M}^{(v)}$ to a 3^{rd} order tensor with size $P \times P \times K$ defined as $\mathcal{M} \in \mathbb{R}^{P\times P\times K}$ (i.e., $\mathcal{M}(:,:,v) = \mathbf{M}^{(v)}$). Likewise, we merge all $\mathbf{S}^{(v)}$ to a 3^{rd} order tensor $\mathcal{S} \in \mathbb{R}^{N\times N\times K}$ and $\mathbf{Z}^{(v)}$ to a 3^{rd} order tensor $\mathcal{Z} \in \mathbb{R}^{N\times P\times K}$ such that $\mathcal{S}(:,:,v) = \mathbf{S}^{(v)}$ and $\mathcal{Z}(:,:,v) = \mathbf{Z}^{(v)}$, respectively. To enforce low-rankness criterion on $\mathcal{S}$, we minimize rank($\mathcal{S}$). Since rank($\mathcal{S}$) is non-convex, we replace rank($\mathcal{S}$) with $||\mathcal{S}||_{tr}$ as a convex envelope to tensor rank. We can further obtain a novel equality as follows:

$$||\mathcal{S}||_{tr} = ||\mathcal{M}||_{tr} \quad (4)$$

The main benefit of Eq. (4) is that it enables us to exploit high order correlations underlying multi-view data by using tensor in addition to information of each individual view.

Theorem. Let $\mathcal{S}(:,:,v) = \mathbf{S}^{(v)} \in \mathbb{R}^{N\times N}$, $\mathcal{Z}(:,:,v) = \mathbf{Z}^{(v)} \in \mathbb{R}^{N\times P}$ and $\mathcal{M}(:,:,v) = \mathbf{M}^{(v)} \in \mathbb{R}^{P\times P}$ for $1 <= v <= K$, and $\mathbf{S}^{(v)} = \mathbf{Z}^{(v)}\mathbf{M}^{(v)}(\mathbf{Z}^{(v)})^\top$. Based on the tensorization, we can derive $||\mathcal{S}||_{tr} = ||\mathcal{M}||_{tr}$.

Proof. Based on tensor trace norm, we have the following:

$$||\mathcal{S}||_{tr} = \frac{1}{K}\sum_{v=1}^{K} ||\mathbf{S}(:,:,v)||_{tr} = \frac{1}{K}\sum_{v=1}^{K} ||\mathbf{S}^{(v)}||_{tr} = \frac{1}{K}\sum_{v=1}^{K} ||\mathbf{Z}^{(v)}\mathbf{M}^{(v)}(\mathbf{Z}^{(v)})^\top||_{tr} = \frac{1}{K}\sum_{v=1}^{K} ||\mathbf{M}^{(v)}||_{tr} \quad (5)$$

Since $\frac{1}{K}\sum_{v=1}^{K} ||\mathbf{M}^{(v)}||_{tr} = ||\mathcal{M}||_{tr}$, we proved that $||\mathcal{S}||_{tr} = ||\mathcal{M}||_{tr}$ holds.

C. Outlier-Robust Multi-View Subspace Clustering

With the aim of improving robustness against error, we separate error from each individual view by decomposing it into underlying clean subspace of the view and error (outlier) term that encodes error in the view as follows:

$$\mathbf{Z}^{(v)} = \mathbf{Z}^{(v)}\mathbf{M}^{(v)} + \mathbf{E}^{(v)} \quad (6)$$

where $\mathbf{M}^{(v)}$ represents the same variable in Eq. (2) which acts as the subspace of singular vectors of view v, and $\mathbf{E}^{(v)}$ indicates error in view v.

Our proposed decomposition in Eq. (6) is novel because it seamlessly connects pairwise similarity completion and error removal from views via $\mathbf{M}^{(v)}$. More precisely, according to Eqs. (2) and (6), $\mathbf{M}^{(v)}$ has more responsibility than just storing information on direction of linear relationship between projection coefficients for singular vectors of views. $\mathbf{M}^{(v)}$ captures clean underlying subspace of singular vectors as well as projection coefficients of cluster membership vectors. The role of $\mathbf{M}^{(v)}$ in Eq. (2) as subspace of singular vectors in Eq. (6) is motivated by the assumption that cluster membership vectors lie in the space of singular vectors of views. According to Eqs. (6) and (4), we derive the following objective function to learn the underlying clean subspace of singular vectors of views while utilizing prior information by optimizing $\mathbf{M}^{(v)}$ and $\mathbf{E}^{(v)}$:

$$\min_{\mathbf{M}^{(v)},\mathbf{E}^{(v)}} ||\mathcal{M}||_{tr}$$
$$s.t.\ \mathbf{Z}^{(v)} = \mathbf{Z}^{(v)}\mathbf{M}^{(v)} + \mathbf{E}^{(v)}, \mathcal{M} = \Phi(\mathbf{M}^{(1)}, \cdots, \mathbf{M}^{(K)}) \tag{7}$$

where Φ denotes the merge operator to put $\mathbf{M}^{(v)}$ along third mode and construct $\mathcal{M}$. Since multi-view data may come from multiple sources with inconsistent magnitude values, as suggested by [2], we vertically concatenate error matrices, i.e., $\mathbf{E} = [\mathbf{E}^{(1)}; \cdots; \mathbf{E}^{(K)}] \in \mathbb{R}^{KN\times P}$ to have the consistent magnitude values across columns of $\mathbf{E}^{(v)}$ jointly. To further improve robustness of the proposed RMVSC method against outliers, we impose $\ell_{2,1}$ norm on $\mathbf{E}$. This norm characterizes outliers elegantly because error matrix with outliers has sparse row support [2]. The objective function can be then written as follows (λ is the nonnegative trade-off parameter):

$$\min_{\mathbf{M}^{(v)},\mathbf{E}^{(v)}} ||\mathcal{M}||_{tr} + \lambda||\mathbf{E}||_{2,1}\ s.t.\ \mathbf{Z}^{(v)} = \mathbf{Z}^{(v)}\mathbf{M}^{(v)} + \mathbf{E}^{(v)},$$
$$\mathcal{M} = \Phi(\mathbf{M}^{(1)}, \cdots, \mathbf{M}^{(K)}), \mathbf{E} = [\mathbf{E}^{(1)}; \cdots; \mathbf{E}^{(K)}] \tag{8}$$

D. Outlier-Robust Multi-View Subspace Clustering with Prior Constraints

According to $\mathbf{S}_{ob} = \mathbf{\Omega}*\mathbf{S}_{gt}$, we introduce a constraint into Eq. (8) to preserve observed entries based on prior pairwise constraints: $\mathbf{\Omega}*\mathbf{Z}^{(v)}\mathbf{M}^{(v)}(\mathbf{Z}^{(v)})^\top = \mathbf{S}_{ob}$. The final objective function for the proposed RMVSC method is as follows:

$$\min_{\mathbf{M}^{(v)},\mathbf{E}^{(v)}} ||\mathcal{M}||_{tr} + \lambda||\mathbf{E}||_{2,1}\ s.t.\ \mathbf{Z}^{(v)} = \mathbf{Z}^{(v)}\mathbf{M}^{(v)} + \mathbf{E}^{(v)},$$
$$\mathcal{M} = \Phi(\mathbf{M}^{(1)}, \cdots, \mathbf{M}^{(K)}),$$
$$\mathbf{E} = [\mathbf{E}^{(1)}; \cdots; \mathbf{E}^{(K)}], \mathbf{\Omega} * \mathbf{Z}^{(v)}\mathbf{M}^{(v)}(\mathbf{Z}^{(v)})^\top = \mathbf{S}_{ob} \tag{9}$$

Initially, we optimize $\mathcal{M}$. We then unfold $\mathcal{M}$ to obtain $\mathbf{M}^{(v)}$ using $\mathcal{M}(:,:,v) = \mathbf{M}^{(v)}$. Next, we establish $\mathbf{S}^{(v)}$ as $\mathbf{S}^{(v)} = \mathbf{Z}^{(v)}\mathbf{M}^{(v)}(\mathbf{Z}^{(v)})^\top$. Finally, we obtain affinity matrix $\mathbf{S}$ by combining all similarity matrices using average operator as follows: $\mathbf{S} = \frac{1}{K}\sum_{v=1}^{K}\mathbf{Z}^{(v)}\mathbf{M}^{(v)}(\mathbf{Z}^{(v)})^\top$.

IV. Optimization Procedure

We use *Augmented Lagrange Multiplier (ALM)* to solve the objective function in Eq. (9). By introducing the auxiliary tensor variable $\mathcal{G}$, we make the objective function separable. We thus solve the following optimization problem:

$$\min_{\mathbf{M}^{(v)},\mathbf{E}^{(v)},\mathcal{G}} ||\mathcal{G}||_{tr} + \lambda||\mathbf{E}||_{2,1}$$
$$+\sum_{v=1}^{K}(< \mathbf{Y}^{(v)}, \mathbf{Z}^{(v)} - \mathbf{Z}^{(v)}\mathbf{M}^{(v)} - \mathbf{E}^{(v)} >)$$
$$+ < \mathcal{W}, \mathcal{M} - \mathcal{G} > + \frac{\mu}{2}\sum_{v=1}^{K}(||\mathbf{Z}^{(v)} - \mathbf{Z}^{(v)}\mathbf{M}^{(v)} - \mathbf{E}^{(v)}||_F^2)$$
$$+\frac{\rho}{2}||\mathcal{M} - \mathcal{G}||_F^2 + \epsilon\sum_{v=1}^{K}||\mathbf{\Omega} * (\mathbf{Z}^{(v)}\mathbf{M}^{(v)}(\mathbf{Z}^{(v)})^\top) - \mathbf{S}_{ob}||_F^2 \tag{10}$$

$\mathbf{M}^{(v)}$-subproblem. When $\mathbf{E}^{(v)}$ and $\mathcal{G}$ are fixed, we need to solve the following subproblem to update $\mathbf{M}^{(v)}$:

$$\min_{\mathbf{M}^{(v)}}\sum_{v=1}^{K}(< \mathbf{Y}^{(v)}, \mathbf{Z}^{(v)} - \mathbf{Z}^{(v)}\mathbf{M}^{(v)} - \mathbf{E}^{(v)} >)$$
$$+ < \mathbf{W}^{(v)}, \mathbf{M}^{(v)} - \mathbf{G}^{(v)} >$$
$$+\frac{\mu}{2}\sum_{v=1}^{K}(||\mathbf{Z}^{(v)} - \mathbf{Z}^{(v)}\mathbf{M}^{(v)} - \mathbf{E}^{(v)}||_F^2) + \frac{\rho}{2}||\mathbf{M}^{(v)} - \mathbf{G}^{(v)}||_F^2$$
$$+\epsilon\sum_{v=1}^{K}||\mathbf{\Omega} * (\mathbf{Z}^{(v)}\mathbf{M}^{(v)}(\mathbf{Z}^{(v)})^\top) - \mathbf{S}_{ob}||_F^2 \tag{11}$$

We use gradient descent to solve the above subproblem. The gradient for $\mathbf{M}^{(v)}$ ($\bigtriangledown_{\mathbf{M}^{(v)}}$) can be computed as follows:

$$\bigtriangledown_{\mathbf{M}^{(v)}} = 2\epsilon(\mathbf{Z}^{(v)})^\top(\mathbf{\Omega} * (\mathbf{Z}^{(v)}\mathbf{M}^{(v)}(\mathbf{Z}^{(v)})^\top - \mathbf{S}_{ob}))\mathbf{Z}^{(v)}$$
$$+\rho(\mathbf{M}^{(v)} - \mathbf{G}^{(v)} + \frac{1}{\rho}\mathbf{W}^{(v)})$$
$$-\frac{\mu}{2}(\mathbf{Z}^{(v)})^\top(\mathbf{Z}^{(v)} - \mathbf{Z}^{(v)}\mathbf{M}^{(v)} - \mathbf{E}^{(v)} + \frac{1}{\mu}\mathbf{Y}^{(v)}) \tag{12}$$

where $\mathbf{Y}^{(v)}$, $\mathcal{W}$ are two Lagrange multipliers. Then, $\mathbf{M}^{(v)}$ can be updated using $\mathbf{M}^{(v)} \leftarrow \mathbf{M}^{(v)} - \eta\bigtriangledown_{\mathbf{M}^{(v)}}$.

$\mathbf{E}^{(v)}$-subproblem. When other variables are fixed, the following subproblem should be solved to update $\mathbf{E}^{(v)}$:

$$\min_{\mathbf{E}}\lambda||\mathbf{E}||_{2,1} + \sum_{v=1}^{K}(< \mathbf{Y}^{(v)}, \mathbf{Z}^{(v)} - \mathbf{Z}^{(v)}\mathbf{M}^{(v)} - \mathbf{E}^{(v)} >)$$
$$+\frac{\mu}{2}\sum_{v=1}^{K}(||\mathbf{Z}^{(v)} - \mathbf{Z}^{(v)}\mathbf{M}^{(v)} - \mathbf{E}^{(v)}||_F^2)$$
$$= \lambda||\mathbf{E}||_{2,1} + \frac{\mu}{2}||\mathbf{E} - \mathbf{D}||_F^2 \tag{13}$$

where $\mathbf{D}$ is constructed by vertically concatenating the matrices $\mathbf{Z}^{(v)} - \mathbf{Z}^{(v)}\mathbf{M}^{(v)} + (\frac{1}{\mu})\mathbf{Y}^{(v)}$ together along column. According to [2], the solution for the subproblem can be obtained as follows:

Algorithm 1 outlier-Robust Multi-View Subspace Clustering with prior pairwise constraints (RMVSC)

Input: $\mathbf{X}^{(1)}, \cdots, \mathbf{X}^{(K)}$, $\mathbf{A}$, $\mathbf{B}$, C
Parameter: λ, P
Output: $\mathbf{M}^{(1)}, \cdots, \mathbf{M}^{(K)}$, $\mathbf{E}$, clustering results

```
E = 0, η = 0.1, ϵ = 10, μ = 10^{-4}, ρ = 10^{-5}, tol = 10^{-6}, η_l = 1.9
Y^(v) = 0, G = W = 0, ρ_max = μ_max = 10^{10}
Construct S_ob from A and B using S_ob = Σ_{c=1}^{C} u_c^(v) (u_c^(v))^⊤
for v = 1, ···, K do
   repeat
      Update M^(v) ← M^(v) − η∇_{M^(v)}
      Update E using Eq. (14)
      Update G^(v) using Eq. (16)
      Update W ← W + ρ(M − G)
      Update Y^(v) ← Y^(v) + μ(Z^(v) − Z^(v) M^(v) − E^(v))
      μ = min(η_l μ, μ_max), ρ = min(η_l ρ, ρ_max)
   until ||Z^(v) − Z^(v) M^(v) − E^(v)||_∞ < tol and ||M^(v) − G^(v)||_∞ < tol
end for
S = (1/K) Σ_{v=1}^{K} Z^(v) M^(v) (Z^(v))^⊤
Apply k-means on the shared similarity matrix S
```

$$\mathbf{e}_i = \begin{cases} \frac{||\mathbf{d}_i||_2 - \frac{\lambda}{\mu}}{||\mathbf{d}_i||_2}\mathbf{d}_i & \text{if} \quad ||\mathbf{d}_i||_2 > \frac{\lambda}{\mu} \\ 0 & \text{otherwise} \end{cases} \tag{14}$$

where $\mathbf{e}_i$ denotes i-th column of $\mathbf{E}$, $\mathbf{d}_i$ represents i-th column of $\mathbf{D}$, respectively.

$\mathcal{G}$**-subproblem.** When other variables are fixed, to update variable $\mathcal{G}$, $\min_{\mathcal{G}} ||\mathcal{G}||_{tr} + \frac{\rho}{2}||\mathcal{M} - \mathcal{G} + \frac{1}{\rho}\mathcal{W}||_F^2$ should be solved. The solution for the subproblem is given by the singular value decomposition for the frontal slices of $\mathcal{G}$ i.e., $\mathcal{G}(:,:,v) = \mathbf{G}^{(v)}$ [16]. In other words, $\mathcal{G}$ can be updated slice-by-slice as follows:

$$\min_{\mathbf{G}^{(v)}} ||\mathbf{G}^{(v)}||_{tr} + \frac{\rho}{2}||\mathbf{G}^{(v)} - \mathbf{M}^{(v)} - \frac{1}{\rho}\mathbf{W}^{(v)}||_F^2 \tag{15}$$

The above subproblem has a closed-form solution by using singular value threshold (SVT) method [17]. More precisely, let $\mathbf{U}^{(v)}\sum^{(v)}(\mathbf{V}^{(v)})^\top$ denote the SVD form of $(\mathbf{M}^{(v)} + \frac{1}{\rho}\mathbf{W}^{(v)})$. We update $\mathbf{G}^{(v)}$ as follows:

$$\mathbf{G}^{(v)} = \mathbf{U}^{(v)}\mathbf{S}_{1/\rho}(\sum\nolimits^{(v)})(\mathbf{V}^{(v)})^\top \tag{16}$$

where $\mathbf{S}_\sigma = max(\mathbf{X} - \sigma, 0) + min(\mathbf{X} + \sigma, 0)$ is shrinkage operator [11].

$\mathcal{W}$**-subproblem.** Lagrange multiplier $\mathcal{W}$ is updated as $\mathcal{W} \leftarrow \mathcal{W} + \rho(\mathcal{M} - \mathcal{G})$.

$\mathbf{Y}^{(v)}$**-subproblem.** We update it as $\mathbf{Y}^{(v)} \leftarrow \mathbf{Y}^{(v)} + \mu(\mathbf{Z}^{(v)} - \mathbf{Z}^{(v)}\mathbf{M}^{(v)} - \mathbf{E}^{(v)})$.

The optimization procedure is described in Algorithm 1. The convergence condition is two infinity norms on optimality gap for subspace representations.

V. Computational Complexity and Convergence Analysis

Table II: Complexity of tensorized multi-view subspace clustering algorithms

Methods	Computational Complexity
LTMSC	$O(DN^2 + N^3)$ (D : largest dimension of view features)
TMSRL	$O(DN^2 + N^3)$ (D : largest dimension of view features)
t-SVD-MSC	$O(N^3) + O(K(2N^2Klog(N)))$
RMVSC	$O(NPK) + O(P^3K)$

The computational bottleneck of RMVSC only lies in solving the subproblems of $\mathbf{E}$ and $\mathcal{G}$. The computational complexity of $\mathbf{E}$ at each iteration is $O(NPK)$. For $\mathcal{G}$ subproblem, computing K SVDs of $P \times P$ matrices at each iteration takes $O(P^3K)$. Overall, the per iteration complexity is $O(NPK) + O(P^3K)$, which is significantly less than computational complexity of existing tensorized multi-view subspace clustering methods in [7]–[9], named as t-SVD-MSC, TMSRL and LTMSC, as reported in Table II. It is worth mentioning that $P << N$ where N is the number of data instances which can be arbitrarily large. P is usually some multiple of C, the number of clusters. C is much smaller than N.

According to [2], [8], two requirements are sufficient for convergence of Algorithm 1: (1) $\mathbf{Z}^{(v)}$ is of full column rank, (2) the optimality gap i.e., $||(\mathbf{M}_{iter}^{(v)}, \mathbf{E}_{iter}^{(v)}) - \arg\min_{\mathbf{M}^{(v)},\ \mathbf{E}^{(v)}} \text{Eq.}(9)||_F^2$ at each iteration $iter$ is monotonically decreasing. To satisfy the first condition, similar to [2], we can state that $\mathbf{M}^{(v)} \in \text{span}((\mathbf{Z}^{(v)})^\top)$. We can therefore express $\mathbf{M}^{(v)}$ as linear combination of $(\mathbf{Z}^{(v)})^\top$. In other words, $\mathbf{M}^{(v)} = \mathbf{P}^{(v)}\hat{\mathbf{M}^{(v)}}$ where $\mathbf{P}^{(v)}$ is derived by making the columns of $(\mathbf{Z}^{(v)})^\top$ orthogonal. The decomposition part of Eq. (9) can be equivalently stated as follows:

$$\mathbf{Z}^{(v)} = \mathbf{Z}^{(v)}\mathbf{P}^{(v)}\hat{\mathbf{M}^{(v)}} + \mathbf{E}^{(v)} \tag{17}$$

It is evident that $\mathbf{I} = \mathbf{Z}^{(v)}\mathbf{P}^{(v)}$, which is of full column rank. The second condition is met to some extent due to the convexity of the Lagrange function. The convergence of Algorithm 1 could be thus well expected.

VI. Experimental Evaluation

We empirically evaluate the performance of the proposed RMVSC method on several challenging real-world datasets.

A. Real-World Datasets

Statistics of the following real-world datasets are summarized in Table III. **Coil (CO)**[1]: a generic object dataset that

[1]http://www.cs.columbia.edu/CAVE/software/softlib/coil-20.php

contains images of object categories taken from different angels. Similar to [7], all images are normalized to 32×32. In addition to intensity (view 1), Gabor features (view 2) are extracted with one scale $\lambda = 4$ at four orientations $\theta = \{0°, 45°, 90°, 135°\}$. The dimension of Gabor is 4096. **Caltech-Middle (CM)**: a generic object dataset which is subcategory of Caltech-All dataset [2]. The views are (1) Hog features, (2) LBP features, and (3) concatenation of Cenhist, Gist, Wavelet moments and Gabor features. **Caltech-All (CA)**[3]: a generic object dataset which is supercategory of Caltech-Middle dataset. The views are (1) Hog features, (2) LBP features, and (3) concatenation of Cenhist, Gist, Wavelet moments and Gabor features. **NUS [18]**: a generic object dataset with six types of features. View 1 has LAB-226, Edge direction histogram, while view 2 includes LAB-64, HSV and wavelet features.

Table III: Statistics of the real-world multi-view datasets

Dataset	# data instances	# views	# clusters
Coil (CO)	1440	2	20
Caltech-Middle (CM)	2386	3	20
Caltech-All (CA)	9144	3	102
NUS	30000	2	31

B. Baselines

We compare RMVSC with the following 10 baselines. **Best Single View (BSV)** does clustering on the most informative view. **Feature Concatenation (FC)** concatenates features of all views. **Robust Multi-View Spectral Clustering via Low-Rank and Sparse Decomposition (RMSC)** is based on integration of matrix low-rank constraint and ℓ_1 norm for robust multi-view clustering in Markov Chain [19]. λ is tuned from $\{10^{-3}, 10^{-2}, \cdots, 10^2, 10^3\}$. **Error-Robust Multi-View Clustering (EMVC)** learns low-rank shared transition probability matrix across all views in Markov chain and enforces $\ell_{2,1}$ and group ℓ_1 norms to characterize various error types [20]. Each regularization hyperparameter is tuned from $\{10^{-9}, 10^{-8}, \cdots, 10^8, 10^9\}$. **t-SVD-MSC** learns subspace representations in favor of efficient low-rank constraint on tensor by stacking subspace representations of all views as well as $\ell_{2,1}$ sparsity on the corresponding error matrices for multi-view subspace clustering [8]. The parameter λ is tuned from $\{0, 0.2, 0.4, ..., 2\}$. **Multi-View Matrix Completion for Clustering with Side Information (MVMC)** a multi-view clustering method that ties together feature representations of all views and instance-level metadata via matrix completion for multi-view clustering. Parameter k is tuned from $\{100, 200, 300, 400, 500\}$. **Low-Rank Representation (LRR)** applies low-rank constraint on concatenated views to recover underlying subspace representation and imposes regularization to improve robustness against error. It then employs PCA to reduce learned representation dimensionality to 500 and modifies it with prior must-link pairwise constraints [2]. Parameter λ is tuned from $\{10^{-3}, 10^{-2}, \cdots, 10^2, 10^3\}$. **Low-Rank Tensor Constrained Multi-View Subspace Clustering (LTMSC)** regards subspace representations of all views as a tensor and formulates the multi-view clustering problem as a tensor rank minimization problem with $\ell_{2,1}$ regularization term [7]. **Multimodal Sparse and Low-rank Subspace Clustering (MSSC)** learns a shared subspace representation across all views, while decomposing each view into underlying subspace representation and other terms to encode error in the view [21]. λ related parameters are tuned from $\{10^{-3}, 10^{-2}, \cdots, 10^2, 10^3\}$. **Tensorized Multi-View Subspace Representation Learning (TMSRL)** incorporates label-based prior constraints into tensorized multi-view subspace clustering. There is no mature approach to convert prior pairwise constraints to label-based prior constraints for TMSRL [9]. We include it without any prior information.

C. Experimental Settings

For RMVSC, we apply line search to find optimal value for λ from $\{0.001, 0.01, 0.1, 1, 10, 100, 1000\}$. We consider the range $P_{set} = \{100, 200, 300, 400, 500\}$ for parameter P such that the optimal value is chosen as $min(P_{set}, d^{(v)})$ where $d^{(v)}$ denotes number of features in view v. The maximum number of iterations for all methods is set to 50. We use standard **Normalized Mutual Information (NMI)** and **Purity** to measure the clustering performance. The higher value of each metric indicates the better performance. Each experiment is repeated for five times, and the mean of each metric in each dataset is reported. We then use k-means to obtain final clustering solution. We run k-means 20 times on each dataset and report the average results.

Similar to [22], we generate prior pairwise constraints by randomly selecting a pair of data instances. The $ratio$ of prior constraints is varied from the range $\{0.1, 0.2\}$ such that $\beta = ratio \times N^2$ (β denotes quantity of prior constraints). RMVSC1, MVMC1, and LRR1 denote the methods when using $ratio = 0.1$ prior pairwise constraints, while RMVSC2, MVMC2 and LRR2 indicate the corresponding algorithms with $ratio = 0.2$ prior pairwise constraints. We randomly select a small portion of data instances $\{2\%, 4\%\}$ and replace their feature values in all views by random values, which is similar to generation of attribute outliers in [23]. The outlier-contaminated data instances are chosen as uniformly distributed random numbers.

D. Experimental Results

Tables IV and V report clustering performance (NMI and purity) on NUS and CO datasets, while Tables VI and VII on CM and CA datasets. The bold numbers highlight the best results. The first column for each dataset denotes the data without any additive outliers, while the other columns indicate dataset with $\{2\%, 4\%, 6\%\}$ outliers in all views.

[2]http://www.vision.caltech.edu/Image_Datasets/Caltech101/
[3]http://www.vision.caltech.edu/Image_Datasets/Caltech101/

Table IV: Comparison results on (erroneous) datasets (NMI) - Part 1

Method\Dataset	NUS	NUS2%	NUS4%	NUS6%	CO	CO2%	CO4%	CO6%
BSV	14.4	13.9	13.7	11.0	77.0	75.7	74.1	74.0
FC	10.3	9.9	8.1	7.7	77.4	75.3	73.4	71.1
RMSC	17.2	16.1	14.5	13.2	82.9	81.0	78.0	75.9
EMVC	18.0	16.9	15.1	14.3	87.0	87.0	83.5	80.4
LTMSC	18.3	18.2	15.1	13.2	86.0	79.1	76.4	72.6
TMSRL	19.6	18.0	15.5	13.9	81.0	79.5	76.5	73.0
t-SVD-MSC	20.3	18.5	17.6	15.2	88.4	82.5	82.3	82.1
MSSC	10.3	9.9	8.0	6.9	53.5	50.5	47.9	46.2
LRR1	20.6	19.2	17.4	15.3	91.5	90.5	90.0	89.1
LRR2	21.2	20.0	18.2	16.5	91.1	90.7	90.1	89.1
MVMC1	20.0	18.0	16.5	15.0	94.8	91.6	89.3	87.6
MVMC2	22.0	19.1	17.2	16.6	95.2	92.3	89.6	87.6
RMVSC1	23.5	23.1	21.2	21.0	99.4	98.2	97.5	96.7
RMVSC2	**32.6**	**32.3**	**32.1**	**32.2**	**99.5**	**98.9**	**97.9**	**97.6**

Table V: Comparison results on (erroneous) datasets (Purity) - Part 1

Method\Dataset	NUS	NUS2%	NUS4%	NUS6%	CO	CO2%	CO4%	CO6%
BSV	17.8	16.0	14.7	13.9	68.4	67.2	65.8	66.2
FC	15.6	13.2	11.3	10.0	69.9	66.2	64.4	62.1
RMSC	19.3	18.9	17.4	16.2	76.2	73.6	70.6	68.4
EMVC	25.0	21.0	20.0	18.9	78.0	75.9	73.5	71.6
LTMSC	30.1	29.0	27.2	25.6	81.0	73.2	70.1	69.5
TMSRL	30.5	30.0	28.3	26.9	74.5	73.6	70.5	70.0
t-SVD-MSC	31.1	30.5	30.0	28.3	82.0	75.2	74.9	75.0
MSSC	16.5	14.0	12.7	11.0	50.0	43.4	41.3	38.7
LRR1	32.3	31.5	30.2	28.3	80.0	79.0	78.1	78.3
LRR2	33.0	32.0	31.8	30.0	79.7	79.2	80.2	79.1
MVMC1	32.1	31.0	29.1	27.0	88.7	84.8	82.3	81.0
MVMC2	32.5	31.3	29.0	28.0	89.2	85.4	82.3	80.7
RMVSC1	35.6	35.0	35.0	33.0	96.2	95.9	94.1	91.0
RMVSC2	**38.8**	**38.5**	**37.0**	**35.5**	**98.1**	**97.9**	**96.9**	**94.1**

The proposed RMVSC2 method under 6% of error outperforms all baseline without additive error on both NMI and purity over all four datasets. That means robustness to error of the proposed method. Even for RMVSC1 with 6% error, it still outperforms the baselines without additive error on purity over all datasets, while on NMI except NUS.

Also, the proposed RMVSC method achieves superior performance to tensorized multi-view subspace clustering approaches, i.e., t-SVD-MSC, LTMSC and TMSRL, that exploit high order intrinsic structure in data, while failing to incorporate prior pairwise constraints during learning. The clustering performance becomes better when using prior pairwise constraints and high order correlations underlying multi-view data. RMVSC leads to improvement with a large average margin of at least 10%. Our method outperforms RMSC, EMVC, LRR and MSSC, i.e., robust multi-view clustering schemes that do not capture high order correlations in data. Compared to LRR, our method gains significant improvement in average at least 4%. Overall, RMVSC achieves in average improvement of 6%.

The proposed RMVSC method achieves superior performance to LTMSC, t-SVD-MSC, TMSRL i.e., multi-view clustering algorithms that separate error from data while clustering the data instances. On CO dataset in Tables IV and V, our method obtains a nearly perfect result (in average greater than 96%). Based on Tables IV and VI, on CM, CA and NUS, the improvement of our method over other tensorized multi-view clustering approaches LTMSC, TMSRL and t-SVD-MSC is noticeable because our method utilizes prior constraints to enhance clustering solution. Our method exhibits robustness against outliers.

E. Hyperparameter Analysis

RMVSC has two major parameters, (λ, P) where P denotes number of first left singular vectors of feature matrices of all views. Fig. 2 shows the clustering performance for λ and P for datasets without additive error when the ratio of metadata is 0.1. On NUS, we cannot report the results for $P = 400$ and $P = 500$ because one of the views has less than 400 features. The performance is fairly stable. RMVSC gives more promising results when λ is in range of [0.01, 0.1] and P has possible maximum value.

VII. Related Work

Multi-view learning has been extensively studied in recent years [3]. Existing approaches for multi-view clustering, which are mostly related to our work, can be roughly classified into three categories [3], [8]: (1) graph-based approaches, (2) co-training or co-regularized methods, (3) subspace learning algorithms. Graph-based approaches use multiple graph fusion strategy to capture complementary and consistent information across all views as well as information of each individual view [19], [20], [24]. To handle various error types in multi-view data, Najafi et al. presented a Markov chain based multi-view clustering method, named as EMVC, to obtain a shared transition

Table VI: Comparison results on (erroneous) datasets (NMI) - Part 2

Method\Dataset	CM	CM2%	CM4%	CM6%	CA	CA2%	CA4%	CA6%
BSV	59.9	57.7	55.4	54.1	53.0	49.6	48.6	47.9
FC	38.4	37.6	36.7	36.0	37.0	34.0	33.5	33.0
RMSC	50.7	49.1	47.9	46.5	57.5	56.1	54.2	53.0
EMVC	57.9	57.6	56.9	55.9	59.2	57.2	55.0	54.1
LTMSC	49.2	48.5	47.5	46.9	64.1	62.0	61.0	60.1
TMSRL	69.5	69.0	67.1	65.8	63.8	62.5	61.6	60.5
t-SVD-MSC	60.1	59.8	58.3	57.1	69.0	68.0	64.1	63.2
MSSC	50.2	48.9	45.4	44.5	44.2	41.0	40.0	39.2
LRR1	79.0	77.1	76.0	74.1	70.5	69.2	67.0	65.2
LRR2	83.0	81.8	79.0	77.1	70.8	70.0	70.0	68.4
MVMC1	74.1	69.3	65.1	63.9	68.0	65.0	62.1	60.4
MVMC2	75.7	71.0	69.4	67.0	68.7	66.9	63.2	62.0
RMVSC1	90.3	86.1	85.4	83.0	72.0	71.1	70.1	69.2
RMVSC2	**90.5**	**88.6**	**86.5**	**84.1**	**74.0**	**73.4**	**71.9**	**71.0**

Table VII: Comparison results on (erroneous) datasets (Purity) - Part 2

Method\Dataset	CM	CM2%	CM4%	CM6%	CA	CA2%	CA4%	CA6%
BSV	76.3	74.5	72.3	71.4	51.5	47.8	46.7	46.0
FC	61.1	60.2	59.4	58.9	35.1	31.4	30.9	30.4
RMSC	69.7	67.7	67.1	66.6	60.1	58.0	56.3	54.0
EMVC	79.1	78.0	76.5	74.9	61.0	59.0	58.2	56.5
LTMSC	60.1	59.5	58.6	59.1	64.7	64.2	62.0	60.5
TMSRL	70.9	70.4	68.2	66.2	65.0	64.5	62.3	61.0
t-SVD-MSC	75.8	75.0	74.3	73.4	68.1	67.0	65.9	63.2
MSSC	54.2	53.5	50.1	50.4	55.1	53.2	52.1	40.5
LRR1	81.3	80.7	78.0	76.1	65.0	62.8	60.8	58.4
LRR2	86.1	84.6	82.3	80.0	67.1	65.0	63.1	61.2
MVMC1	86.5	81.0	79.2	77.0	68.2	66.1	64.3	62.7
MVMC2	86.7	83.1	81.4	79.9	69.1	68.2	67.1	65.2
RMVSC1	95.5	92.6	90.6	89.4	73.1	72.7	71.2	70.0
RMVSC2	**95.8**	**94.0**	**92.4**	**91.5**	**75.1**	**74.2**	**73.2**	**71.5**

probability matrix among all views via low-rank and sparse decomposition of each view [20]. The sparsity is introduced by integration of $\ell_{2,1}$ and group ℓ_1 norms. The proposed RMVSC method differs from this category of approaches in a way that it uses subspace as data structure that indeed facilitates representation and processing.

Co-training or co-regularized algorithms capture pairwise correlation between views, while focusing on each individual view simultaneously [1], [4]. Zhao et al. proposed a method that casts multi-view clustering with prior pairwise constraints into matrix completion problem while minimizing distance between shared similarity matrix across all views and similarity matrix of each individual view [4]. Different from this category, our method fully utilizes pairwise and high order correlations in multi-view data. The goal of multi-view subspace clustering is to seek underlying clean subspace representation of views and perform clustering of data instances accordingly [7]–[9], [21]. Zhang et al. formulated multi-view subspace clustering as tensor rank minimization constrained with $\ell_{2,1}$ regularization term, while decomposing each view into the underlying clean subspace representation and error term to encode error in the view [7]. The tensor is established by stacking subspace representation of all views. None of the existing methods utilize clues from prior pairwise constraints for enhanced clustering solution with low computational complexity while exploring high order correlations underlying multi-view data. RMVSC belongs to the third category.

VIII. Conclusion

In this paper, we presented a novel method, named as RMVSC, for outlier-robust multi-view subspace clustering with prior pairwise constraints. To exploit high order complementary and consistent information across all views as well as reduce computational complexity significantly, RMVSC regards subspace of singular vectors of views as a tensor. To improve robustness against outliers, each view is decomposed into underlying clean subspace and error term. A constraint term is devised to enforce prior constraints.

IX. Acknowledgement

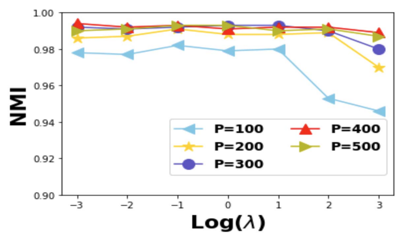
This work is supported in part by NSF under grants III-1763325, III-1909323, III-2106758, and SaTC-1930941. Lifang He was supported by NSF ONR N00014-18-1-2009 and Lehigh's accelerator grant S00010293.

References

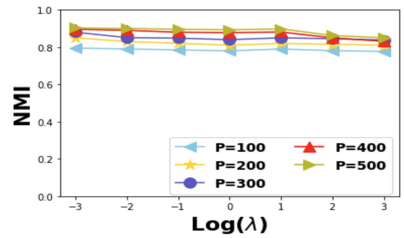
[1] S. Bickel and T. Scheffer, "Multi-view clustering," in *ICDM*, 2004, p. 19–26.

[2] G. Liu, Z. Lin, S. Yan, J. Sun, Y. Yu, and Y. Ma, "Robust recovery of subspace structures by low-rank representation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 1, pp. 171–184, 2013.

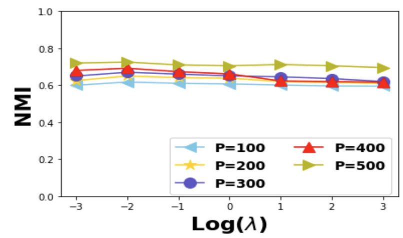
[3] Y. Yang and H. Wang, "Multi-view clustering: A survey," *Big Data Mining and Analytics*, vol. 1, no. 2, pp. 83–107, 2018.



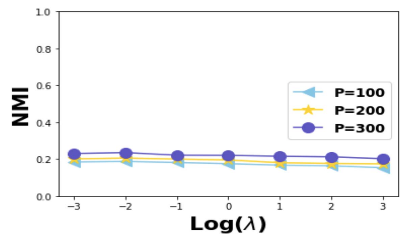
(a) CO



(b) CM



(c) CA



(d) NUS

Figure 2: Sensitivity analysis of hyperparameters

[4] P. Zhao, Y. Jiang, and Z.-H. Zhou, "Multi-view matrix completion for clustering with side information," in *Advances in Knowledge Discovery and Data Mining*, 2017, pp. 403–415.

[5] M. White, Y. Yu, X. Zhang, and D. Schuurmans, "Convex multi-view subspace learning," in *NIPS*, 2012, p. 1673–1681.

[6] H. Gao, F. Nie, X. Li, and H. Huang, "Multi-view subspace clustering," in *ICCV*, 2015, pp. 4238–4246.

[7] C. Zhang, H. Fu, S. Liu, G. Liu, and X. Cao, "Low-rank tensor constrained multiview subspace clustering," in *ICCV*, 2015, pp. 1582–1590.

[8] 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," *International Journal of Computer Vision*, vol. 126, no. 11, pp. 1157–1179, 2018.

[9] C. Zhang, H. Fu, J. Wang, W. Li, X. Cao, and Q. Hu, "Tensorized multi-view subspace representation learning," *IJCV*, pp. 1573–1405, 2020.

[10] A. S. Shirkhorshidi, S. Aghabozorgi, T. Y. Wah, and A. R. Dalby, "A comparison study on similarity dissimilarity measures in clustering continuous data," *PLoS*, vol. 10, 2015.

[11] Z. Lin, M. Chen, and Y. Ma, "The augmented lagrange multiplier method for exact recovery of corrupted low-rank matrices," *arXiv preprint arXiv:1009.5055*, 2010.

[12] R. T. Rockafellar, "Lagrange multipliers optimality," *SIAM*, vol. 35, no. 2, pp. 183–238, 1993.

[13] M. Xu, R. Jin, and Z.-H. Zhou, "Speedup matrix completion with side information: Application to multi-label learning," in *NIPS*, 2013, p. 2301–2309.

[14] F. Bonchi, A. Gionis, and A. Ukkonen, "Overlapping correlation clustering," in *ICDM*, 2011, pp. 51–60.

[15] O. Shamir and S. Shalev-Shwartz, "Matrix completion with the trace norm: Learning, bounding, and transducing," *J. Mach. Learn. Res.*, vol. 15, no. 1, p. 3401–3423, 2014.

[16] C. Lu, J. Feng, Y. Chen, W. Liu, Z. Lin, and S. Yan, "Tensor robust principal component analysis with a new tensor nuclear norm," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2018.

[17] J.-F. Cai, E. J. Candès, and Z. Shen, "A singular value thresholding algorithm for matrix completion," *SIAM Journal on Optimization*, vol. 20, no. 4, pp. 1956–1982, 2010.

[18] T.-S. Chua, J. Tang, R. Hong, H. Li, Z. Luo, and Y. Zheng, "Nus-wide: A real-world web image database from national university of singapore," in *ACM International Conference on Image and Video Retrieval*, 2009.

[19] R. Xia, Y. Pan, L. Du, and J. Yin, "Robust multi-view spectral clustering via low-rank and sparse decomposition," in *AAAI*, 2014, pp. 2149–2155.

[20] M. Najafi, L. He, and P. S. Yu, "Error-robust multi-view clustering," *IEEE Big Data*, pp. 736–745, 2017.

[21] M. Abavisani and V. M. Patel, "Multimodal sparse and low-rank subspace clustering," *Information Fusion*, vol. 39, pp. 168 – 177, 2018.

[22] H. Zeng and Y. Cheung, "Semi-supervised maximum margin clustering with pairwise constraints," *TKDE*, vol. 24, no. 5, pp. 926–939, 2012.

[23] H. Zhao and Y. Fu, "Dual-regularized multi-view outlier detection," in *IJCAI*, 2015.

[24] F. Nie, J. Li, and X. Li, "Parameter-free auto-weighted multiple graph learning: A framework for multiview clustering and semi-supervised classification," in *IJCAI*, 2016, p. 1881–1887.