Nonconvex Multiview Subspace Clustering Framework with Efficient Method Designs and Theoretical Analysis

Zhi Wang, Zhuo Liu, Dong Hu and Tao Jia

College of Computer and Information Science, Southwest University, Chongqing, China
chiw@swu.edu.cn, zhuoliu.swu@gmail.com, donghu@email.swu.edu.cn, tjia@swu.edu.cn

Abstract
Multi-view subspace clustering (MvSC) is one of the most effective methods for understanding and processing high-dimensional data. However, existing MvSC methods still have two shortcomings: (1) they adopt the nuclear norm as the low-rank constraint, which makes it impossible to fully exploit the mutually complementary subspace information, and (2) they do not handle disjoint and confounding points carefully, which may degrade the purity and distinctiveness of cross-view fusion. To address these issues, in this paper we propose a novel MvSC model with nonconvex ℓ_q regularization. Specially, our proposed model can not only effectively capture the intrinsic global low-rank structure, but also accurately cluster disjoint and confounding data samples into corresponding subspaces. Then, an efficient algorithm is developed with convergence guarantee. Furthermore, we prove that the sequence generated by our proposed algorithm converges to the desirable Karush-Kuhn-Tucker (KKT) critical point. Extensive experiments on various datasets verify the superiority of our proposed model. MATLAB code is available at https://github.com/wangzhi-swu/NLRSC-MvSC.

1 Introduction
In many real-world applications, one needs to consider how to effectively and efficiently process high-dimensional data, which is a key issue in machine learning [Slepčev and Thorpe, 2019; Cai et al., 2011; Tang et al., 2023], computer vision [Ho et al., 2003; Yang et al., 2008; Vidal et al., 2008], and pattern recognition [Basri and Jacobs, 2003; Ma et al., 2008]. However, such problems are extremely challenging because of the “curse of dimensionality” [Muja and Lowe, 2014]. Fortunately, some researchers have pointed out that high-dimensional data are drawn from low-dimensional subspaces. Therefore, how to efficiently recover the low-dimensional structures in datasets becomes a crucial step.

To address this issue, subspace clustering (SC) has been proposed and extensively studied, and has become the most important and fundamental tool for high-dimensional data analysis [Vidal, 2011; Liu et al., 2010]. Such technique aims to separate data samples into their exclusive groups, where each group corresponds to a subspace with highly similar characteristics. The most prevalent methods of self-representation based subspace clustering are sparse subspace clustering (SSC) [Elhamifar and Vidal, 2013], low-rank representation (LRR) [Liu et al., 2012], and some of their variants [Qu et al., 2022; Liu et al., 2023c]. Specifically, SSC has the ability to seek the sparsest representation coefficients corresponding to the data from the same subspace. On the other hand, LRR is to seek a low-rank subspace representation. Despite their success, these approaches only focus on single-view features.

In practical scenarios, however, data are often collected from different sources and can be represented by various types of features [Fang et al., 2023]. For example, a color image consists of colors, textures, and shapes, a video is a combination of visual frames and audio tracks, and a web page may contain text, images and their corresponding link. The features of different views usually have their own inherent attributes, which make it possible to provide consistent and complementary information to each other. Therefore, integrating these different heterogeneous and complementary features into clustering methods can significantly improve their performance.

To exploit complementary and consistent information from different views, many Multi-view SC (MvSC) methods have been proposed [Gao et al., 2015; Zhang et al., 2017; Li et al., 2019a; Chen et al., 2022]. Different from single-view subspace clustering methods, MvSC uses multiple features to accomplish the multi-view data clustering tasks. Therefore, the MvSC methods have superiority compared with single-view subspace clustering methods. Although these methods have achieved remarkable results in MvSC tasks, they still suffer from the following shortcomings. First, they adopt the convex nuclear norm as the low-rank constraint, which ignores significant differences between different singular values [Wang et al., 2021; Wang et al., 2022; Shan et al., 2023]. As a result, they cannot accurately capture low rankness, resulting in inferior performance. Second, the union of underlying n subspaces \( \{S_i\}_{i=1}^n \) is commonly disjoint [Tang et al., 2014]. As shown in Figure 1, for the data points that nearby distributed at local intersections, it is
hard to identify the subspaces they really belong to. Unfortunately, these confounding points have still not been carefully handled, which may degrade the purity and distinctiveness of cross-view fusion.

To overcome above drawbacks, in this paper we propose a new MvSC model namely Nonconvex Low-Rank and Structure-Constrained MvSC (NLRSC-MvSC). Comparing with the aforementioned models, NLRSC-MvSC has two prominent advantages. First, by using the $\ell_q$ norm [Marjanovic and Solo, 2012] in our model, the intrinsic low-rank structure in multi-view data can be efficiently captured [Wang et al., 2019; Wang et al., 2020]. Thus, the mutually complementary subspace embedding information in each view can be fully exploited, so as to provide a more comprehensive and in-depth description with regard to the underlying subspaces. Second, with the aim of clustering the disjoint and confounding data samples into corresponding subspaces, a shared cross-view structure restriction is naturally imposed on each coefficient matrix. To be concrete, the correlation between inter-cluster samples can be significantly reduced with a soft penalty strategy, whereas the informative intra-cluster coefficients are sufficiently preserved. As a result, by fusing the view-specific compact representations, a more discriminative consensus fusion is yield, which can offer a better characterization of true subspace structure. In summary, the main highlights of this paper are as follows.

- We propose a novel MvSC model that simultaneously imposes low-rank and structural constraints on view-specific representations by leveraging the nonconvex $\ell_q$ norm and soft penalty strategy. It can comprehensively explore the complementary and consistent connections among multiple coefficient matrices, thereby further facilitating the effectual cross-view fusion.
- We establish an efficient algorithm to solve the corresponding optimization problem. Furthermore, rigorous mathematical proof shows that the sequence generated by our proposed algorithm converges to a KKT point.
- Extensive experiments on various multi-view datasets validate the effectiveness and superiority of our proposed model over several state-of-the-art multi-view clustering models.

## 2 Related Work

### 2.1 Multi-view Subspace Clustering

As a fundamental tool for processing multi-view data, MvSC has become an active topic in machine learning and computer vision in recent years. Generally speaking, most existing MvSC methods have three steps, including i) learn the coefficient matrices of the data samples via “self-expressiveness”; ii) construct the affinity graph by fusing the learned coefficient representations; iii) apply the spectral clustering method [Ng et al., 2001] on affinity graph to get the final clustering result. How to accurately learn a coefficient matrix is crucial for these methods. To this end, the LRR strategy is commonly used to recover the underlying subspaces, which can fully capture the global structure of the data. However, the resultant optimization problem is NP-hard due to the discreteness of the rank function. To alleviate this problem, a common approach is to use nuclear norm as its convex relaxation. In real-world, the multi-view data may contain errors (e.g., noise and outliers). Thus, the $\ell_{2,1}$ norm can be used to reduce their interference. Along this line of research, a large number of MvSC models have been proposed [Zhang et al., 2017; Luo et al., 2018; Chen et al., 2022]. In [Zhang et al., 2017], Zhang et al. proposed a latent MvSC (LMvSC) model to explore the complementarity of multiple views. In [Chen et al., 2022], Chen et al. proposed a multi-view LRR (MvLRR) model using symmetric LRR fusion, which can more comprehensively explore the underlying low-rank subspace. In [Luo et al., 2018], Luo et al. proposed a model that can learn consistent and specific representations simultaneously. Although these methods have achieved satisfactory results in MvSC tasks, there are still some shortcomings. On the one hand, the nuclear norm as the low-rank constraint may result in a severe loss of multi-view information. On the other hand, these methods enforce the elements across representations to be consistent, which neglects the difference in value magnitudes between and within views.

### 2.2 Nonconvex $\ell_q (0 < q < 1)$ Regularization

In the region of compressed sensing [Chartrand, 2007; Chartrand and Yin, 2008], the nonconvex regularization methods related to the $\ell_q$ norm [Marjanovic and Solo, 2012] have received extensive attention, and a large number of theoretical results and algorithms have been proposed to solve the corresponding optimization problems. Comparing with its convex counterparts, using the $\ell_q$ norm can obtain more sparse solution and requires fewer measurements. In general, the minimization problem associated with $\ell_q$ norm can be solved by the following non-trivial optimization problem

$$
\min_{x \geq 0} \left\{ f(x) = \lambda x^q + \frac{1}{2} (x - y)^2 \right\},
$$

where $\lambda > 0$ is a parameter. However, it is in general difficult to provide a thorough theoretical understanding for finding a global solution. Besides, how to choose a proper $q$ to yield the best result is also an open issue. Fortunately, recent researches [Xu et al., 2012; Cao et al., 2013] have shown that the efficiency of $\ell_{1/2}$ and $\ell_{2/3}$ regularizations is very high, and their corresponding closed-form thresholding formulas can be easily obtained.
Lemma 1 (Xu et al., 2012). For any $\lambda > 0$, the optimal solution to 
\[
\min_{x \geq 0} \lambda x^2 + \frac{1}{2} (x - y)^2, \tag{2}
\]
is given by \[h_{\lambda, \frac{2}{3}}(x) = \begin{cases} \frac{2}{\lambda} x (1 + \cos(\frac{2\pi}{3} - \frac{2\psi(x)}{3})), & x > \gamma \\ 0, & \text{otherwise}, \end{cases} \tag{3}\]
where $\psi(x) = \arccos((\lambda/4)(x/3)^{-3/2})$ and $\gamma = 3\sqrt{2}/(4\lambda)^{2/3}$.

Lemma 2 (Cao et al., 2013). For any $\lambda > 0$, the optimal solution to 
\[
\min_{x \geq 0} \lambda x^2 + \frac{1}{2} (x - y)^2, \tag{4}
\]
is given by 
\[
h_{\lambda, \frac{2}{3}}(x) = \begin{cases} \left(\frac{\varphi(x) + \Phi(x)}{2}\right)^3, & x > \gamma \\ 0, & \text{otherwise}, \end{cases} \tag{5}\]
where $\Phi(x) = (2x/\varphi(x) - \varphi(x)x^2)^{1/2}$ with $\varphi(x) = (2/\sqrt{3})(\lambda/2)^{1/4}(\cosh(\arccosh(27x^2(2\lambda)^{-3/2}/(6))^3))^{1/2}$ and $\gamma = 2/3(3\lambda)^{1/4}$.

Accordingly, to improve the efficiency of our proposed method without sacrificing performance, we only consider the nonconvex $\ell_q$ regularization with $q \in \{1/2, 2/3\}$.

3 Methodology

This section firstly formulates the proposed NLRSC-MvSC model by utilizing the cross-view and underlying structure information comprehensively. Then, an effective algorithm with convergence guarantee is devised to address the resultant optimization problem.

3.1 Problem Definition

Given multi-view data $X = \{X_1, X_2, \ldots, X_V\}$, where each $X_v \in \mathbb{R}^{d_v \times n}$ denotes the $v$-th feature matrix. The purpose of MvSC is to divide the given samples into $K$ groups. To improve the performance of MvSC method, the two general prior principles, i.e., Complementarity and Consistency, should be fully considered.

Complementarity. In most practices, one data object can be described by different subspace information from multiple views, which are commonly complementary with each other. Therefore, it is of prime importance to maximize the diversity of joint representations so as to pursue a comprehensive information fusion. To this end, by leveraging the superior $\ell_q$ norm and subspace self-expressiveness property, a nonconvex low-rank MvSC model is proposed
\[
\min_{C_v, E_v} \sum_{v=1}^{V} \|C_v\|_q + \beta \sum_{v=1}^{V} \|E_v\|_{2,q}, \tag{6}\]
s.t. $X_v = X_v C_v + E_v, \forall v \in \{1, 2, \ldots, V\}$, where $C_v$ denotes the $v$-th view’s learned subspace representation, $E_v$ denotes the error matrix for characterizing the noise and outliers, $\beta > 0$ denotes a balance parameter, and the $\ell_{2,q}$ norm is a sparser error characteristic than $\ell_{2,1}$ due to the embedding of $\ell_q$ penalty.

Consistency. It is well known that all the view representations exhibit a common block diagonal structure with different value magnitudes. Instead of forcing each pair of them to be similar, we advocate to enhance the consistent block diagonal property across representation matrices, while retaining the diversity within them. Thus, a shared structural constraint on each subspace representation is introduced into (6), and the final NLRSC-MvSC model can be formulated as 
\[
\min_{C_v, E_v} \sum_{v=1}^{V} \|C_v\|_q + \lambda \sum_{v=1}^{V} \|W \odot C_v\|_q + \beta \sum_{v=1}^{V} \|E_v\|_{2,q}, \tag{7}\]
s.t. $X_v = X_v C_v + E_v, \forall v \in \{1, 2, \ldots, V\}$, where $\lambda$ is a trade-off parameter for balancing each term, $\odot$ denotes the Hadamard product [Zhang, 2017], and $W_{i,j} \in [0, 1]$ represents the soft weight on each coefficient. Ideally, the coefficients of inter-cluster samples should be assigned the largest weight 1, while the smallest 0 is weighted on intra-cluster samples. However, the label information of data is unknown in unsupervised MvSC tasks. Aim to obtain an expected weight, an efficacious strategy is given as follows 
\[
W_{i,j} = 1 - \exp \left(-\frac{1 - |\langle H_i, H_j \rangle|}{\xi}\right), \tag{8}\]
where $\langle \cdot, \cdot \rangle$ is the inner product, $H = [\hat{X}_1; \hat{X}_2; \ldots; \hat{X}_V]^T \in \mathbb{R}^{D \times n}$ with $D = d_1 + d_2 + \cdots + d_V$ denotes the vertical concatenation of all normalized feature sets $X_v$, and $\xi$ is set to the average of all $1 - |\langle H_i, H_j \rangle|$ empirically. In this way, the angle information among confounding data points can be fully exploited for accurate subspace segmentation.

3.2 Optimization

By drawing on the powerful alternating direction method of multipliers (ADMM) [Kuybeda et al., 2013; Liu et al., 2023b] framework, the nonconvex and multi-variable problem arising by model (7) can be effectively solved with the following designed algorithm.

According to the ADMM method, we introduce two auxiliary variables into each view-specific subproblem for the separability of $C_v$, and then (7) is equivalent to the following form
\[
\min_{P_v, S_v, C_v, E_v} \|P_v\|_q + \lambda \|W \odot S_v\|_q + \beta \|E_v\|_{2,q}, \tag{9}\]
s.t. $X_v = X_v C_v + E_v, C_v = P_v, C_v = S_v$. 

Now, the Augmented Lagrangian function of constrained problem (9) can be given with three dual multipliers $M_{v,1}, M_{v,2}$ and $M_{v,3}$, i.e.,
\[
\mathcal{L}_{\mu}(P_v, S_v, C_v, E_v, M_{v,1}, M_{v,2}, M_{v,3}) = \|P_v\|_q + \lambda \|W \odot S_v\|_q + \beta \|E_v\|_{2,q} + \Omega(M_{v,1}, C_v - P_v) + \Omega(M_{v,2}, C_v - S_v) + \Omega(M_{v,3}, X_v - X_v C_v - E_v), \tag{10}\]
where $\Omega(A, B) = \langle A, B \rangle + (\mu/2)\|B\|_F^2, \mu > 0$ is a penalty scalar. Then, the unconstrained problem (10) can be optimized by solving the subproblem of each variable alternately.
Update $P_v$: Fixing the other variables, the subproblem involving with $P_v$ can be written as
\[
\arg \min_{P_v} \| P_v \|_q^q + \Omega(M_{v,1}, C_v - P_v)
\]
\[
= \arg \min_{P_v} \frac{1}{\mu} \| P_v \|^q + \frac{1}{2} \| P_v - (C_v + \frac{M_{v,1}}{\mu}) \|^2.
\]
(11)

For simplicity, let $Y = C_v + M_{v,1}/\mu$, and the SVD of $Y$ be $U\text{Diag}(\sigma(Y))V^T$. Its solution can be achieved by the closed-form thresholding formulas in Lemma 1 and 2
\[
P_v^* = U\text{Diag}(\frac{\mu}{\mu-q}(\sigma(Y)))V^T,
\]
(12)
where $\sigma_{\mu-q}(x_i) = (h_{\mu,q}(x_1), h_{\mu,q}(x_2), \ldots, h_{\mu,q}(x_n))$.

Update $S_v$: Fixing the other variables, the subproblem involving with $S_v$ can be written as
\[
\arg \min_{S_v} \lambda \| W \odot S_v \|_q^q + \Omega(M_{v,2}, C_v - S_v)
\]
\[
= \arg \min_{S_v} \frac{1}{\mu} \| W \odot S_v \|^q + \frac{1}{2} \| S_v - (C_v + \frac{M_{v,2}}{\mu}) \|^2.
\]
(13)

Let $Y = C_v + M_{v,2}/\mu$ and $B = (\lambda/\mu)W$. Similar to (11), we obtain
\[
[S_v^*]_{i,j} = h_{B_{i,j},q}(Y_{i,j}).
\]
(14)

Update $C_v$: Fixing the other variables, the subproblem involving with $C_v$ can be written as
\[
\arg \min_{C_v} \Omega(M_{v,1}, C_v - P_v) + \Omega(M_{v,2}, C_v - S_v)
\]
\[
+ \Omega(M_{v,3}, X_v - X_v C_v - E_v).
\]
(15)
The convex quadratic equation (15) can easily be solved by taking the derivative with respect to $C_v$, and setting it to zero, we get
\[
C_v^* = (2\mu I + \mu X_v^T X_v)^{-1} \tilde{C}_v,
\]
(16)
where $\tilde{C}_v = X_v^T (M_{v,3} + \mu X_v - \mu E_v) + \mu P_v + \mu S_v - M_{v,1} - M_{v,2}$.

Update $E_v$: Fixing the other variables, the variable $E_v$ can be achieved by the following subproblem
\[
\arg \min_{E_v} \beta \| E_v \|_2, q + \Omega(M_{v,3}, X_v - X_v C_v - E_v),
\]
\[
= \arg \min_{E_v} \frac{\beta}{\mu} \| E_v \|_2, q + \frac{1}{2} \| E_v - \tilde{E}_v \|^2,
\]
(17)
where $\tilde{E}_v = X_v - X_v C_v + (1/\mu)M_{v,3}$. Then, we have
\[
[E_v^*]_{i,j} = h_{B_{i,j},q}(Y_{i,j}),
\]
(18)
where $Y$ can be obtained as [Liu et al., 2012]
\[
[Y_{i,j}] = \frac{\| [E_v^*]_{i,j} \|_2 - \frac{\beta}{\mu} + \| \tilde{E}_v \|_2}{\| [E_v^*]_{i,j} \|_2}.
\]
(19)

Algorithm 1 Algorithm for solving problem (9)
\begin{algorithm}
\caption{Algorithm for solving problem (9)}
\begin{algorithmic}[1]
\State \textbf{Input:} Multi-view data $\{X_1, X_2, \ldots, X_V\}$, weight matrix $W$ and parameters $\lambda, \beta, \gamma$.
\State \textbf{Initialize:} $C_v = P_v = S_v = 0$, $E_v = 0$, $M_{v,1} = M_{v,2} = M_{v,3} = 0$, $\forall v \in \{1, 2, \ldots, V\}$; $\mu = 10^{-2}$, $\mu_{\text{max}} = 10^{10}$, $\rho = 2$; $\epsilon = 10^{-6}$.
\State \textbf{Output:} $\{C^*_v\}_{v=1}^V$.
\Function{Update}{P_v} \Comment{while not converged}
\State $v \gets 1$; \textbf{while} not converged \textbf{do}
\State Update $P_v$ by (12);
\State Update $S_v$ by (14);
\State Update $C_v$ by (16);
\State Update $E_v$ by (18);
\State Update $M_{v,1}, M_{v,2}, M_{v,3}$ and $\mu$ by (20)-(23);
\EndFunction
\State \textbf{end for}
\State \textbf{Check} the convergence criterion: $\max\{\|C_v - P_v\|_\infty, \|C_v - S_v\|_\infty, \|X_v - X_v C_v - E_v\|_\infty\} < \epsilon$
\EndFunction
\end{algorithmic}
\end{algorithm}

Update $M_{v,1}, M_{v,2}, M_{v,3}$ and $\mu$: The multipliers and penalty scalar can be updated by
\[
M_{v,1}^* = M_{v,1} + \mu(C_v - P_v),
\]
(20)
\[
M_{v,2}^* = M_{v,2} + \mu(C_v - S_v),
\]
(21)
\[
M_{v,3}^* = M_{v,3} + \mu(X_v - X_v C_v - E_v),
\]
(22)
\[
\mu^* = \min(\rho \mu, \mu_{\text{max}}),
\]
(23)
where $\rho > 1$ denotes the step size for convergence. The details of our proposed optimization scheme are stated in Algorithm 1.

With the obtained set of optimal subspace representations $\{C^*_v\}_{v=1}^V$, the compact affinity graph $G$ can be constructed by the following fusion mechanism
\[
G = \frac{1}{V} \sum_{v=1}^V |C^*_v| + |C^*_v|^T.
\]
(24)

Then, the clustering labels for MvSC can be achieved by conducting the spectral clustering [Ng et al., 2001] on $G$.

3.3 Theoretical Analysis

Computational Complexity. The time complexity of our proposed algorithm is mainly to solve the subproblems involving with $P_v, S_v, C_v$, and $E_v$, respectively. To be specific, updating $P_v$ takes $O(n^3)$ due to SVD operation. The complexity of updating $S_v$ is $O(n^2)$. For variable $C_v$, its updating relies on the operations of SVD, matrix multiplication and inverse, thus the complexity is $O(n^3 + n^2 d_v)$. Besides, updating $E_v$ with matrix multiplication costs $O(n^2 d_v)$ time. Consequently, the Algorithm 1 has the computational complexity of $O(KV n^3)$ (assuming $d_{\text{max}} \leq n$), where $d_{\text{max}} = \max\{d_1, d_2, \ldots, d_V\}$, $K$ denotes the number of iterations.

Convergence Analysis. In addition to the terrific performance, it is also indispensable to establish the rigorous convergence guarantee of the proposed algorithm. Thus, a reliable analysis is theoretically provided in following Theorem 1, which demonstrates that the sequence generated by
In our experiments, five popular datasets under different clustering scenarios are selected. Their brief descriptions are reported as follows, and detailed statistical data are shown in Table 1.

• **Yale** is a face dataset which contains 165 data samples from 15 subjects in each view. And three types of features are extracted, i.e., Intensity, LBP and Gabor.
• **MSRC-v1** is a object dataset. It consists of 7 subjects with 210 images in each view. Six image features are adopted, that is, HOG, SIFT, LBP, CENT, GIST and CMT.
• **BBC4view** is a news text dataset with four different views. Each includes 685 documents from five topics.
• **100leaves** is a object dataset consists of 1600 leaf samples belonging to 100 plant species. The shape, margin and texture features of each leaf are employed as three different views.

### 4 Experiments

In this section, we evaluate our proposed NLRSC-MvSC model on five benchmark datasets in comparison with ten representative clustering models. Then, the clustering performance and robustness of NLRSC-MvSC are analyzed.

#### 4.1 Datasets and Experimental Setting

**Datasets.** In our experiments, five popular datasets under different clustering scenarios are selected. Their brief descriptions are reported as follows, and detailed statistical data are shown in Table 1.

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Clu./Sam.</th>
<th>Dimension</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yale</td>
<td>15/165</td>
<td>4096/3304/6750</td>
</tr>
<tr>
<td>MSRC-v1</td>
<td>7/210</td>
<td>1302/512/256/210/100/48</td>
</tr>
<tr>
<td>BBC4view</td>
<td>5/685</td>
<td>4659/4633/4665/4684</td>
</tr>
<tr>
<td>100leaves</td>
<td>100/1600</td>
<td>64/64/64</td>
</tr>
<tr>
<td>ProteinFold</td>
<td>27/694</td>
<td>694/694/694/...</td>
</tr>
</tbody>
</table>

Table 1: Statistical details of benchmark datasets. Clu. and Sam. denote the Cluster and Sample, respectively.

Our proposed algorithm converges to a stationary KKT critical point. The detailed proof is provided in the supplementary material.

**Theorem 1.** Let the sequence \( \{\Theta^k\}^\infty_{k=1} \) be generated by Algorithm 1. Then, we have

1) The sequence \( \{\Theta^k\}^\infty_{k=1} \) is bounded, and has at least one accumulation point.

2) Any accumulation point \( \Theta^* \) satisfies

\[
\begin{align*}
C_v^* - P_v^* &= 0, \quad (25) \\
C_v^* - S_v^* &= 0, \quad (26) \\
X_v - X_v C_v^* - E_v^* &= 0, \quad (27) \\
M_{v,1}^* \in \partial_{P_v} \| P_v^* \|^q_q, \quad (28) \\
M_{v,2}^* \in \lambda \partial_{S_v} \| W \odot S_v^* \|^q_q, \quad (29) \\
M_{v,3}^* \in \beta \partial_{E_v} \| E_v^* \|^2_q. \quad (30)
\end{align*}
\]

### 4.2 Compared Models

- **EOMvSC-CA**
- **MvLPL**
- **MvLRSSC-C**
- **LRR**
- **MMvGC**
- **Ours**

### 4.3 Clustering Metrics

Four metrics are employed as the comprehensive evaluation of clustering performance, including accuracy (ACC), normalized mutual information (NMI), adjusted rand index (ARI) and F1-score (F1). The higher the scores of metrics, the better the clustering performance. All experiments are conducted by MATLAB R2020b on a workstation with Intel(R) Xeon(R) Gold 6230 CPU@2.10GHz and 256GB RAM.
4.2 Performance Evaluation

The clustering performance on five datasets of all competing methods are listed in Table 2 - Table 6. The highest and second highest scores are marked in bold and underline, respectively. Then, the experimental results are analyzed as follows:

- It can be seen that our proposed methods can always achieve the highest scores on all metrics and datasets. For instance, on Yale dataset, NLRSC-MvSC \((q = 2/3)\) significantly improves the scores in terms of ACC, NMI, ARI and F1 by 3.64\%, 4.89\%, 6.67\% and 6.29\% compared to others. This indicates that the proposed model can fully utilize the consistency and complementarity of multi-view features, thereby contributing to learn a more comprehensive subspace representation.

- In order to verify the performance improvement of the consistent structural constraint in NLRSC-MvSC, we visualize the affinity graphs produced by different methods in Figure 2. It can be observed intuitively that the affinity graph constructed by our method exhibits a more distinct block diagonal structure with less deceptive pairwise correlations, as marked in MvLRR and MMvGC graphs by red boxes. Thus, our graph can provide more discriminative subspace information for clustering process. In addition, the block saliency of obtained graphs can be quantitatively measured by the Contrast Index (CI), which is defined as [Li et al., 2019b]

\[
\text{CI} = \frac{D_{\text{sum}}}{D_{\text{sum}} + N_{\text{sum}}} = \frac{D_{\text{sum}}}{\|G\|_1}, \tag{31}
\]

where \(D_{\text{sum}}\) and \(N_{\text{sum}}\) denote the summation of affinity values belonging to the diagonal and nondiagonal blocks, respectively. As shown in Table 7, NLRSC-MvSC achieves the highest score in comparison with several state-of-the-art methods. This can be attributed to the accurate subspace segmentation of confounding samples under the structural constraint.

- Furthermore, we apply the t-distributed stochastic neighbor embedding (t-SNE) technique [van der Maaten and Hinton, 2008] to compare the low rankness of fusion representations learned by NLRSC-MvSC and other nuclear norm-based methods. Specifically, we first obtain...
the cluster indicator matrix $\mathbf{F}$ by solving the following Laplacian embedding problem [Gao et al., 2015]

$$
\min_{\mathbf{F}} \text{Tr}(\mathbf{F}^T \mathbf{L}_G \mathbf{F}) \quad \text{s.t.} \quad \mathbf{F}^T \mathbf{F} = \mathbf{I},
$$

(32)

where $\mathbf{L}_G = \mathbf{A} - \mathbf{G}$ is Laplacian matrix, $\mathbf{G}$ can be constructed by (24), and $\mathbf{A}$ denotes a diagonal matrix with elements $A_{i,i} = \sum_j G_{i,j}$. The optimal $\mathbf{F} \in \mathbb{R}^{n \times c}$ can be yielded by the eigenvectors of $\mathbf{L}_G$ corresponding to the smallest $c$ eigenvalues, where $c$ is the number of clusters. Then, a 2-D visualization of the clustering quality can be plotted by t-SNE with input $\mathbf{F}$. It can be seen in Figures 3e and 3f that there exist more larger gaps between the clusters dyed in different colours, while the points in same cluster are closer to each other. This means that

4.3 Discussions

Parameter Sensitivity. Here, we tune the model parameters $\lambda$ and $\beta$ in a wide range, i.e., $\{10^{-6}, 10^{-5}, \ldots, 1, 10\}$ and show their impact on the clustering performance of two datasets in Figure 4. It can be seen that the ACC metric of NLRSC-MvSC stably remains at a relatively higher score, which implies that our method has the robust capability to cope with various clustering perturbations. Generally, the better clustering results can be obtained with the moderate $\lambda$ and $\beta$, such as $\{10^{-3}, 10^{-2}, 10^{-1}\}$.

Numerical Convergence. In spite of the rigorous convergence analysis of the proposed algorithm given in Theorem 1, we also experimentally record the Maximum Error (ME) value during iterations, which is defined as $\text{ME} = \max \{\text{Err}_1, \text{Err}_2, \ldots, \text{Err}_v\}$, where $\text{Err}_v = \max \{\|\mathbf{C}_v - \mathbf{F}_v\|_\infty, \|\mathbf{C}_v - \mathbf{S}_v\|_\infty, \|\mathbf{X}_v - \mathbf{X}_v \mathbf{C}_v - \mathbf{E}_v\|_\infty\}$. As presented in Figure 5, the curves on three datasets rapidly drop to 0 within about 5 iterations and remain smooth thereafter, which demonstrates that our algorithm can achieve convergence efficiently and reliably.

5 Conclusion

In this paper, we proposed a novel NLRSC-MvSC model by employing the nonconvex $\ell_q$ regularization for multi-view data clustering. Thanks to the proposed model, the intrinsic global low-rank structure in multi-view data can be efficiently captured, and confounding points can be carefully handled. As a result, a more comprehensive subspace representation can be learned. Extensive experiment results on various datasets show the superiority of our proposed model in clustering performance.
Acknowledgments
This work was supported in part by the Natural Science Foundation of China under Grant 72374173, in part by the Fundamental Research Funds for the Central Universities under Grant SWU-XDJH202303, and in part by Chongqing Innovative Research Groups under Grant CXQT21005.

References
[Li et al., 2019b] Xuelong Li, Quanmao Lu, Yongsheng Dong, and Dacheng Tao. Robust subspace clustering by cauchy loss function. IEEE Transactions on Neural Networks and Learning Systems, 2019.


