IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received August 24, 2020, accepted September 13, 2020, date of publication September 21, 2020,
date of current version September 30, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.3025096

Adaptive Graph Regularized Low–Rank Matrix Factorization With Noise and Outliers for Clustering

MIN ZHAO AND JINGLEI LIU
School of Computer Science and Control Engineering, Yantai University, Yantai 264005, China
Corresponding author: Jinglei Liu (jinglei_liu@sina.com)

This work was supported in part by the Natural Science Foundation of China under Grant 61572419, Grant 61773331, Grant 61703360, and Grant 61801414.

ABSTRACT Clustering, which is a commonly used tool, has been applied in machine learning, data mining and so on, and has received extensive research. However, there are usually noise and outliers in the data, which will bring about significant errors in the clustering results. In this paper, a robust clustering model with adaptive graph regularization (RCAG) is proposed, on which, sparse error matrix is introduced to express sparse noise, such as impulse noise, dead line, stripes, and ℓ_1 norm is introduced to alleviate the sparse noise. In addition, the $\ell_{2,1}$ norm is also proposed mitigating the effects of outliers, and it has rotation invariance property. Therefore, our RCAG is insensitive to data noise and outliers. More importantly, the adaptive graph regularization is introduced into the RCAG to improve the clustering performance. Aiming at the optimization objective, we propose an iterative updating algorithm, named the Augmented Lagrangian Method (ALM), to update each optimization variable respectively. The convergence and time complexity of RCAG is also proved in theory. Finally, experimental results on fourteen datasets of four application scenarios, such as face image, handwriting recognition and UCI, elaborate the superiority of proposed method over seven existing classical clustering methods. The experimental results demonstrate that our approach achieves better clustering performance in ACC and Purity, which is a little less impressive in other ways.

INDEX TERMS Adaptive graph regularization, clustering, $\ell_{2,1}$, noise and outliers, augmented Lagrangian method.

I. INTRODUCTION

Clustering is the process of dividing the object set into multiple classes composed of similar objects. The cluster generated by clustering is a set of data objects, which are related to objects in the same cluster but distinct from objects in other clusters. Furthermore, data clustering is a valuable data analysis tool in machine learning and data mining. However, reducing the influence of noise and outliers in the clustering of data is a major research topic.

In recent years, varieties of clustering methods have been proposed, such as K–means [1], spectral clustering [2]–[4], NMF [5]. K–means aims to learn c cluster centroids that minimize the within cluster data distances. Spectral clustering is a kind of clustering method based on graph theory, which achieves the purpose of clustering the sample data by clustering the feature vectors of the Laplace matrix of the sample data. It is a low-dimension embedding of the affinity matrix between samples. There have been a lot of researches on clustering. Reference [6] proposed a new clustering method that toke sample invariance as priori. Reference [7] proposed a subspace clustering based on Structured AutoEncoder (StructAE). Reference [8] proposed to project raw data into one space in which the projection embraces the geometric consistency (GC) and the cluster assignment consistency (CAC), and didn't need to make intensive parameter selections. Reference [9] built the theoretical connection between Frobenius-norm-based representation (FNR) and nuclear-norm-based representation (NNR). This paper mainly studies the application of matrix factorization to clustering.

The associate editor coordinating the review of this manuscript and approving it for publication was Amir Masoud Rahmani.

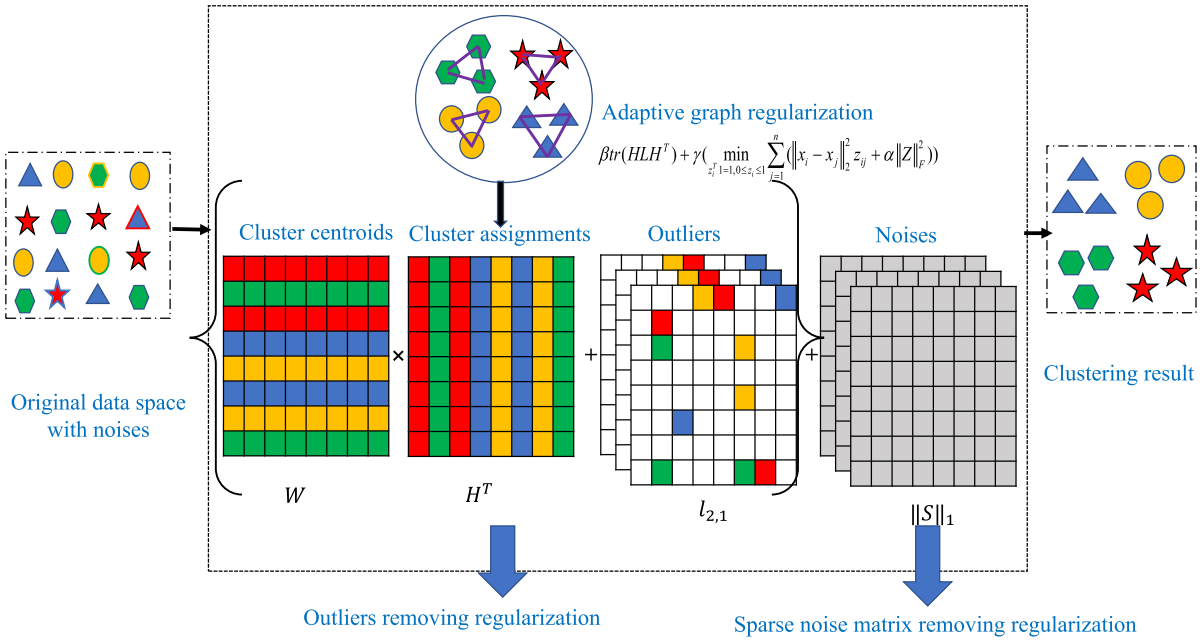


FIGURE 1. The framework of RCAG.

Some forms of matrix factorization include low–rank representation (LRR) [10], principal component analysis (PCA) [11] and singular value decomposition (SVD) [12]. In the past decades, many clustering methods based NMF have been proposed [13], [14].

Non-negative matrix decomposition was proposed by [15]. It makes all the components after decomposition non-negative, and simultaneously realizes the nonlinear dimension reduction. The general form of NMF is: $X \approx WH^T \quad s.t \quad W \geq 0, H \geq 0$ [16]. Non-negative matrix factorization is the most popular methods in this branch. In the clustering setting of NMF, $H \in R^{n\times c}$ is the clustering assignment matrix and $W \in R^{m\times c}$ is the clustering centroids, where c is the number of clusters. Noting that the result of clustering with non-negative matrix factorization can be obtained by performing K-means or other clustering methods on H.

Previous studies have proposed various regularized NMF for clustering [17]–[19]. For instance, many researches have been proposed NMF extensions with graph regularization. The graph regularization constraint of data cluster assignment matrix is applied to obtain the geometric structure of data [18], [20]–[22]. Reference [23] proposed sparse dual graph-regularized nonnegative matrix factorization, and revealed the inherent geometric structure and distinguishing structure of data space and feature space. Reference [24] introduced hypergraph Laplacian regularization to consider the intrinsic geometrical structure and introduced $\ell_{2,1}$ norm to reduce effects of the noise and outliers. Reference [25] used hypergraph regularization to preserve the high-order manifold structure. In the above work, the affinity matrix usually adopts a predefined model, which may not be optimal in practical applications. This can lead to lower-quality diagrams being built and parts of the work not dealing with noise issues.

In order to improve the performance of NMF, many variants with various regularization have been proposed and various methods are proposed to solve the noise and outliers [26]. Huber loss was proposed to handle non-Gaussian noise and outliers, sparse terms and regularization terms were introduced to enhance the sparsity of the matrix and capture the data manifold structure [27]. Reference [28] focuses on the complex noise problem by using finite mixture of exponential power (MoEP) distributions. These work to deal with noise or outliers, but also do not use adaptive graph regularization to adjust.

Nevertheless, clustering based NMF still exist the following problems: (1) The traditional matrix factorization clustering method is easy to be dominated by noise and outliers to produce large errors. (2) The quality of the original graph–based NMF will be affected if the distance between the calculated data samples is not accurate enough.

To address the problems mentioned above, we propose an adaptive graph regularization clustering (RCAG). In our model, adaptive graph regularization is introduced to improve the accuracy of clustering. In addition, we alleviate the influence of noise and outliers by taking $\ell_{2,1}$ norm function [29], and ℓ_1 norm is used to alleviate the influence of sparse corruption. Figure 1 shows the algorithm process of RCAG. The main contributions of the paper are summarized as follows:

(1) We propose a joint learning framework for clustering. By which, the adaptive graph regularization, sparse error matrix and nonnegative low–rank matrix decomposition

are integrated into a unified objective function shown by Equation 3.

(2) In order to get better clustering performance, adaptive graph regularization is introduced. It is parameter–insensitive, scale–invariant, and simple operation.

(3) To address the problem of data being corrupted by noise and outliers, we utilize the ℓ_1 norm and $\ell_{2,1}$ norm to alleviate the influence of sparse noise and outliers, such as impulse noise, dead line, stripes and image occlusion, respectively.

(4) In order to solve the optimization problem, an effective algorithm RCAG described by Algorithm 1 based on the Augmented Lagrangian Method (ALM) is developed. More specifically, the convergence analysis of the designed optimization algorithm is presented from both theoretical perspective shown in Theorem 4 and experimental perspective shown in Figure 3.

The rest of this paper is organized as following. Section II gives the definition of algorithm related symbols and adaptive graph regularization. We propose a new robust clustering frame (RCAG) for clustering and present the theoretical properties of our proposed RCAG approaches in Section III. The experimental environment, experimental procedure, experimental results and analysis are introduced in Section IV. Section V finally concludes RCAG algorithm and gives the future study direction.

II. RELATED WORK

In this section, we introduce the related work of adaptive graph regularization learning.

A. NOTATIONS

The notations description in Table 1. Matrices are written in capital letters (e.g., X), and vectors are written in bold lowercase letters. $X_{\cdot j}$ denotes the j-th column, $X_{i\cdot}$ denotes the i-th row and X_{ij} denotes the entry at the j-th column and i-th row of X.

B. ADAPTIVE GRAPH REGULARIZATION

There are many graph regularized NMFs in existence, but most do not capture the structure of the data effectively. First, most graph constructors need to calculate the distance between the data samples. But if the calculated distance is not accurate enough, then you get a picture of very poor quality. Then, once the graph is built based on the wrong calculation, it stays the same in subsequent steps. Therefore, the input graph is not optimal, and the clustering performance of NMF will be affected. Therefore, it is very necessary to build a high-quality graph.

We suppose that the probability of each sample x_i being connected to its neighbor x_j is z_{ij}, where z_{ij} is an element of the expected similarity matrix Z. Obviously, we suggest that the similar sample pair with small distance $\|x_i - x_j\|_2^2$ should be assigned a high probability z_{ij} [30]. Therefore, we have the following objective function to optimize the Z which meets our assumption:

$$\min_{z_i^T 1=1, 0\le z_i\le 1} \sum_{j=1}^{n} (\|x_i - x_j\|_2^2 z_{ij} + \alpha \|Z\|_F), \tag{1}$$

where α is the regularization parameter.

Note that the equation $\sum_{i,j} \|h_i - h_j\|_2^2 z_{ij} = 2tr(H^T L_Z H)$. Further, we add the rank constraint to it, $rank(L_Z) = n - k$. But this rank constraint is hard to solve. Because the rank constraint $rank(L_Z) = n - k$ equals to $\sum_{i=1}^{k} \sigma_i(L_Z) = 0$. Simultaneously, we have $\sum_{i=1}^{k} \sigma_i(L_Z) = \min_H tr(H^T L_Z H)$. We get Equation 1 as:

$$\begin{aligned} &\arg\min_{Z,H} \sum_{i,j}^{n} (\|x_i - x_j\|_2^2 z_{ij} + \alpha \|Z\|_F^2), \\ &s.t.\ z_i^T \mathbf{1} = 1, \quad 0 \le z_i \le 1,\ H \in R^{n\times k},\ H^T H = I,\ H \ge 0. \end{aligned} \tag{2}$$

III. PROPOSED METHOD

In this section, we introduce the proposed RCAG method. Firstly, the method is presented in three terms. Secondly, the optimization and algorithmic code of RCAG is presented. Thirdly, the time complexity and convergence analysis are presented.

A. MODEL OF RCAG

To solve the sparse noise, outliers and improve clustering performance, this paper proposes the novel robust clustering algorithm model. This model is divided into three terms. As shown in the following Equation 3.

By combining robust matrix factorization with $\ell_{2,1}$, ℓ_1 and adaptive graph regularization, the proposed RCAG can be formulated as (3), shown at the bottom of the page.

$$\begin{aligned} \arg\min_{H\ge 0, H^T H = I, W, S, Z} & \underbrace{\| \overbrace{X - WH^T - S}^{low-rank\ matrix\ factorization} \|_{2,1}}_{outliers\ removing\ regularization} \\ + & \underbrace{\lambda \|S\|_1}_{sparse\ noise\ matrix\ removing\ regularization} \\ + & \underbrace{\beta tr(HL_Z H^T) + \gamma(\min_{z_i^T \mathbf{1}=1,\ 0\le z_i\le 1} \sum_{j=1}^{n} (\|x_i - x_j\|_2^2 z_{ij} + \alpha \|Z\|_F^2))}_{adaptive\ graph\ regularization} \end{aligned} \tag{3}$$

TABLE 1. Key notational description.

Notation	Description
$X \in R^{m\times n}$	Data matrix
Z	Similarity matrix
L	Laplacian matrix
L_Z	the Laplacian matrix computed by the learned Z
D	Degree matrix
$tr(\cdot)$	Trace operator of a matrix
I	Identity matrix
c	Number of classes in the dataset
n	Number of data samples
m	Data dimension
$\|X\|_{2,1}$	$\ell_{2,1}$ norm of matrix X
$\|X\|_F$	ℓ_F norm of matrix X
$\|X\|_1$	ℓ_1 norm of matrix X

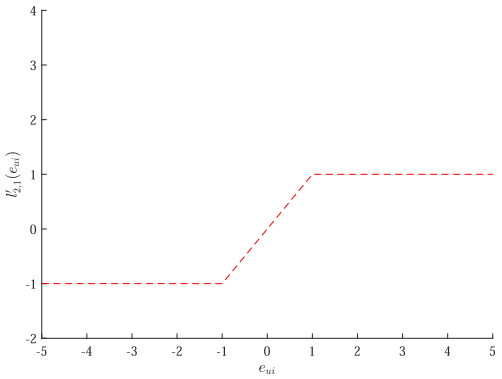


FIGURE 2. The influence function of $\ell_{2,1}$ norm.

1) OUTLIERS REMOVING REGULARIZATION

To solve some data entries that are corrupted by outliers, $X \approx WH^T + S$ is low-rank matrix factorization reconstruction loss term. The Frobenius norm is known to be sensitive to noise and outliers. In order to enhance the robustness, $\ell_{2,1}$ norm is adopted to measure the loss of matrix factorization. $\ell_{2,1}$ adds the ℓ_2 norms of all columns of a matrix. $\ell_{2,1}$ is rotational invariant for rows: $\|XR\|_{2,1} = \|X\|_{2,1}$ for any rotational matrix R [31]. Rotational invariance is a fundamental property of Euclidean space with ℓ_2 norm. Since the influence function of $\ell_{2,1}$ norm is bounded, this means that the effect of outliers on $\ell_{2,1}$ norm can be controlled [32]. But ℓ_2 norm is unbounded, so $\ell_{2,1}$ norm is more robust than ℓ_2 norm.

2) SPARSE NOISE MATRIX REMOVING REGULARIZATION

Noise matrix removing term is to solve the sparse corruption, we introduce a sparse error matrix $S \in R^{m\times n}$ which constrains by ℓ_1 norm. $\|X\|_1 = \sum_{i=1}^{m}\sum_{j=1}^{n}|x_{ij}|$ where is the sum of the absolute values of the elements in the vector. ℓ_1 norm can reduce the impact of sparse noise. This term can remove impulse noise, dead line and stripes.

3) ADAPTIVE GRAPH REGULARIZATION

The third term is adaptive graph regularization, and it is parameter-insensitive, scale–invariant, and simple operation. Because adaptive graph regularization only contains the parameter of the number of nearest neighbors. When each point is scaled, z_{ij} stays the same, which make it scale–invariant. Adaptive graph regularization only involves the basic operations of addition, subtraction, multiplication, and division, which make it simple operation [33]. Adaptive graph regularization is described in detail in Section II-B.

B. SOLUTION OF RCAG

In recent years, many methods have been proposed to solve this type of optimization problem, such as ALM [34] and LADM [35], etc.. First of all, the introduction of three auxiliary variable $E = X - WH^T - S$, $G = S$ and $F = H$. In this paper, the objective function Equation 3 can be changed to the following:

$$\begin{aligned}
&\arg\min_{E,W,H,F,S,G} \|E\|_{2,1} + \lambda\|S\|_1 + \beta tr(H^T L_Z F) \\
&\quad + \gamma\sum_{ij}^{n}(\|x_i - x_j\|_2^2 z_{ij} + \alpha\|Z\|_F^2) \\
&s.t.\ E = X - WH^T - S, \quad F = H,\ G = S, \\
&\quad H^T H = I,\ H \geq 0 \\
&\quad z_i^T \mathbf{1} = 1, \quad 0 \leq z_i \leq 1.
\end{aligned} \tag{4}$$

Then, the above objective function of the problem can be obtained by ALM [34]. The ALM of the Equation 4 is

$$\begin{aligned}
&L(E, W, H, F, Z, S, G, C_1, C_2, C_3) \\
&= \|E\|_{2,1} + \lambda\|S\|_1 \\
&\quad + \beta tr(H^T L_Z F) + \gamma\sum_{ij}^{n}(\|x_i - x_j\|_2^2 z_{ij} + \alpha\|Z\|_F^2) \\
&\quad + \langle C_1, X - WH^T - S - E\rangle + \langle C_2, H - F\rangle + \langle C_3, S - G\rangle \\
&= \|E\|_{2,1} + \lambda\|S\|_1 + \beta tr(H^T L_Z F) \\
&\quad + \frac{\mu}{2}(\|X - WH^T - S - E + \frac{C_1}{\mu}\|_F \\
&\quad + \|H - F + \frac{C_2}{\mu}\|_F + \|S - G + \frac{C_3}{\mu}\|_F),
\end{aligned} \tag{5}$$

C_1, C_2, C_3 are the Lagrange multiplier. Namely, the loss relative to one variable is minimized and other variables are fixed. There are seven variables in total, and the following is the iterative update method.

Update E**:** Fix other variables to update E by asking the following question:

$$L = \arg\min_{E} \|E\|_{2,1} + \langle C_1, X - WH^T - S - E\rangle$$
$$= \arg\min_{E} \|E\|_{2,1} + \frac{\mu}{2}\|X - WH^T - S - E + \frac{C_1}{\mu}\|_F^2 \quad (6)$$

In order to solve Equation 6, we need the following Theorem 1.

Theorem 1: ([21]) Given a matrix $A = [\mathbf{a_1}, \ldots, \mathbf{a_n}] \in R^{m\times n}$ and a positive scalar λ, hence Q^* is Equation 7 of the optimal solution,

$$\min \frac{1}{2}\|Q - A\|_F^2 + \lambda\|Q\|_{2,1}. \quad (7)$$

And the i-th column of Q^*.

$$Q^*(:, i) = \begin{cases} \frac{\|\mathbf{a_i}\| - \lambda}{\|\mathbf{a_i}\|}\mathbf{a}_i & if \quad \lambda < \|\mathbf{a_i}\|, \\ 0 & otherwise. \end{cases}$$

The Equation 6 can be written as follows:

$$\arg\min_{E} \frac{1}{2}\|E - Y\|_F^2 + \frac{1}{\mu}\|E\|_{2,1}, \quad (8)$$

where $Y = X - WH^T - S + \frac{C_1}{\mu}$.

According to Theorem 1, the solution of the Equation 6 is as follows:

$$E(:, i) = \begin{cases} \frac{\|\mathbf{y_i}\| - \lambda}{\|\mathbf{y_i}\|}\mathbf{y}_i & if \quad \frac{1}{\mu} < \|\mathbf{y_i}\|, \\ 0 & otherwise. \end{cases} \quad (9)$$

where $\mathbf{y}_i$ is the i-th column of Y.

Update Z**:** Fix other variables to update Z. Firstly,we denote $d_{ij}^x = \|x_i - x_j\|_2^2$, then Equation 5 becomes:

$$\arg\min_{Z} \sum_{i,j}(d_{ij}^x z_{ij} + \alpha z_{ij}^2) + \frac{\beta}{\gamma}tr(H^T L_Z F).$$
$$s.t \quad \forall i, \quad \sum_j z_{ij} = 1, \ 0 \le z_i \le 1 \quad (10)$$

Denote $d_{ij}^{hf} = \|f_i - h_j\|_2^2$, we can deal with following problem individually for each i:

$$\arg\min_{z_i} \sum_{j=1}^{n}(d_{ij}^x z_{ij} + \alpha z_{ij}^2 + \frac{\beta}{\gamma}d_{ij}^{hf} z_{ij})$$
$$s.t \quad \forall i, \quad \sum_j z_{ij} = 1, \ 0 \le z_i \le 1. \quad (11)$$

Denote $d_i \in R^n$ is a vector with the j-th element as $d_{ij} = d_{ij}^x + \frac{\beta}{\gamma}d_{ij}^{hf}$, then the above problem can be rewritten as follows:

$$\arg\min_{z_i} \|z_i - \frac{1}{2\alpha}d_i\|_2^2 \quad s.t \quad \forall i, \quad \sum_j z_{ij} = 1, \ 0 \le z_i \le 1. \quad (12)$$

Update H**:** Fix other variables to update H by asking the following question:

$$L = \arg\min_{H} 2\beta tr(H^T L_Z F) + \langle C_1, X - WH^T - S - E\rangle$$
$$+ \langle C_2, H - F\rangle$$
$$= \arg\min_{H} 2\beta tr(H^T L_Z F) + \frac{\mu}{2}(\|X - WH^T - S - E$$
$$+ \frac{C_1}{\mu}\|_F^2 + \|H - F + \frac{C_2}{\mu}\|_F^2). \quad (13)$$

Let

$$N = F - \frac{C_2}{\mu} + \frac{2\beta}{\mu}L_Z F + (X - S - E + \frac{C_1}{\mu})^T W. \quad (14)$$

Rewrite optimization problem:

$$\arg\min_{H} \|H - N\|_F^2. \quad (15)$$

According to the constraint conditions, the Equation 15 can be written as:

$$\arg\max_{H^T H = I} tr(H^T N). \quad (16)$$

Theorem 2 was applied to solve this problem.

Theorem 2: ([21]) Given that an equation is similar to the target function in Equation 16, the analytic solution H is defined as:

$$H = UV^T.$$

So the solution to update H is:

$$H = UV^T, \quad (17)$$

where U and V are left and right singular values of SVD decomposition of N.

Update W**:** Fix other variables to update W by asking the following question:

$$L = \arg\min_{W} \|X - WH^T - S - E + \frac{C_1}{\mu}\|_F^2. \quad (18)$$

This is a classical regression problem. Let

$$M = X - S - E + \frac{C_1}{\mu}. \quad (19)$$

The solution of W becomes

$$W = MH. \quad (20)$$

Update S**:** Fix other variables, variable S can be obtained by solving the following problem:

$$L = \arg\min_{S} \lambda\|S\|_1 + \frac{\mu}{2}(\|X - WH^T - S - E + \frac{C_1}{\mu}\|_F$$
$$+ \|S - G + \frac{C_3}{\mu}\|_F). \quad (21)$$

Define $M_1 = X - WH^T - E + \frac{C_1}{\mu}$, $M_2 = \frac{C_3}{\mu} - G$, then S can be obtained via the soft thresholding [34], [36] as follows:

$$S = \max\left(\max\left(M_1 + M_2 - \frac{\lambda}{\mu}, 0\right) + \min\left(M_1 + M_2 + \frac{\lambda}{\mu}, 0\right), 0\right). \quad (22)$$

Update F: Update F by fixing other variables to get the following problem:

$$L = \arg\min_F \beta tr(H^T LF) + \frac{\mu}{2}\|H - F + \frac{C_2}{\mu}\|_F. \quad (23)$$

Let

$$J = H + \frac{C_2}{\mu} - \frac{\beta}{\mu}H^T L. \quad (24)$$

The solution of F is:

$$F_{ij} = \max(J_{ij}, 0), \quad i = 1, \ldots, n, \ j = 1, \ldots, k. \quad (25)$$

Update G: Update G by fixing other variables to get the following problem:

$$L = \arg\min_G \frac{\mu}{2}\|S - G + \frac{C_3}{\mu}\|_F^2, \quad (26)$$

Let

$$\frac{\partial(L_G)}{\partial G} = 0.. \quad (27)$$

The solution of G is:

$$G = S + \frac{C_3}{\mu}. \quad (28)$$

Update parameter C_1, C_2, C_3, μ: After the variables are updated, these ALM method parameters need to be updated.

$$C_1 = C_1 + \mu(X - WH^T - S - E). \quad (29)$$
$$C_2 = C_2 + \mu(H - F). \quad (30)$$
$$C_3 = C_3 + \mu(S - G). \quad (31)$$
$$\mu = \rho\mu. \quad (32)$$

The procedure of the algorithm is described in Algorithm 1.

Algorithm 1: The Proposed RCAG Algorithm

Input: Data set $X \in R^{m\times n}$, the number of data clusters c, μ, ρ, maximum number of iterations T
Output: Converged W and H.
Initialize W and H, Z using K-means;
Step1 :
Computation of H.
while *not converged and iteration less than* T **do**
  1. Update E using Equation 9;
  2. Update Z using Equation 12;
  3. Update H using Equation 17;
  4. Update W using Equation 20;
  5. Update S using Equation 22;
  6. Update F using Equation 25;
  7. Update G using Equation 28;
  8. Update parameter C_1, C_2, C_3, μ using Equation 29-Equation 32;
end
return $W \in R^{m\times c}$, $H \in R^{n\times c}$;
Step2 :
Clustering for H.
Record the column number of the maximum value of each row of H.
return *Clustering result*.

C. THEORETICAL ANALYSIS OF RCAG

In this section, the validity of the method is proved by analysing the time required for calculation. We then pointed out the advantages of RCAG over other approaches.

1) COMPUTATION TIME

The computation complexity of E includes the calculation and update of Y is $O(mn + c^3)$ and $O(mn)$, respectively.

We need $O(mn^2)$ to obtain the matrix Z.

The computation complexity of W is $O(c^2)$. The computation complexity of G is $O(c^2)$.

The computation complexity of S is $O(m^3 + mnc + mc^2 + mc \cdot \max(m, n))$.

The computation complexity of F is $O(m^3 + mc \cdot \max(m, n))$ The main computation complexity of H includes the calculation of N and its SVD decomposition, which are $O(m^3)$ and $O(nc^2)$, respectively.

The overall cost for each iteration is $O(m^3 + mc^2 + mnc + mn + mc \cdot \max(m, n))$. The computational complexity of RCAG is in polynomial time.

2) CONVERGENCE ANALYSIS

The convergence of ALM has been proved in [35]. However, there are seven variables in the paper: W, H, E, Z, G, F, and S. Also, the objective function Equation 4 is not absolutely smooth. These factors do not guarantee that our method is convergent. Fortunately, Theorem 3 proves three sufficient conditions for convergence [10].

Theorem 3: Shaped like $L(x) = f(x) + \lambda h(x)$ can be solved by ALM method. The three conditions to be satisfied by ALM method convergence are as follows:

(1) Parameter λ of ALM problem is needed to be upper bounded.

(2) Original data matrix is full column rank.

(3) The optimality gap produced in each iteration step is monotonically decreasing, namely the error ϑ_k is monotonically decreasing.

Theorem 4: Algorithm 1 is convergence.

Proof: According to Theorem3, our proposed objective function solving method satisfies the above conditions:

(1) Parameter μ of Equation 5 is needed to be upper bounded.

(2) Data matrix X is full column rank.

(3) The optimality gap produced in each iteration step, i.e., $\vartheta_k = \|(E_k, W_k, S_k, H_k, F_k, Z_k, G_k) - \arg\min_{E,W,S,H,F,G} \|_F^2$, monotonically decreases, where $E_k, W_k, S_k, H_k, F_k, Z_k, G_k$ represent the value of E, W, S, H, F, Z, G at the k-th step, respectively.

The first two conditions have been met [10]. But the third condition is hard to prove in theory. Nevertheless, we can prove the third condition experimentally. The value of the objective function in the k iteration is $\vartheta_k = \|E_k - E_{k-1}\|_F^2 + \|W_k = W_{k-1}\|_F^2 + \|S_k - S_{k-1}\|_F^2 + \|H_k - H_{k-1}\|_F^2 + \|F_k - F_{k-1}\|_F^2 + \|Z_k - Z_{k-1}\|_F^2 + \|G_k - G_{k-1}\|_F^2$. It can be seen from Figure3 that the value of the objective function dramatically decreases and then converges rapidly to a stable value, indicating that the third condition has been satisfied to some extent. The convergence of the third term can be proved by [37]. In conclusion, the convergence of the algorithm is guaranteed. ■

3) ADVANTAGES OF RCAG

From a theoretical point of view, RCAG combines matrix factorization, $\ell_{2,1}$, ℓ_1 norm and adaptive graph to data clustering.

–Interpretability:

Different from other matrix factorization methods, NMF decomposes a non-negative data matrix into two non-negative matrices (one is the basis matrix and the other is the coefficient matrix) from the perspective of "individuals constitute the whole and parts constitute the whole". Since the basis matrix and coefficient matrix obtained by the non-negative matrix decomposition method are non-negative, the results of the decomposition are highly interpretable.

–Robustness:

RCAG is effective to remove spares noise in a dataset, and it can handle the influence of outliers. ℓ_1 norm is introduced to alleviate the sparse noise. $\ell_{2,1}$ norm is also introduced to handle outliers, and it has rotation invariance property. Therefore, our RCAG is insensitive to data noise and outliers.

IV. EXPERIMENTS

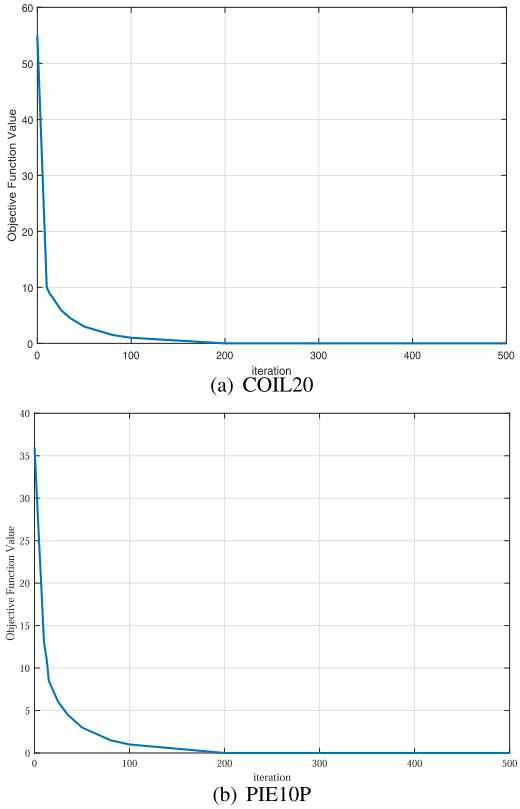
In this section, we evaluate the clustering quality of RCAG over fourteen datasets of four types of datasets, including ACC, NMI, and Purity.

A. EXPERIMENTAL ENVIRONMENT

All algorithms were implemented using Matlab R2014a. The experiment was performed on a computer with 3.2GHz Intel Core CPU, 8.0GB RAM and the Windows 7 operating system.

B. DATASETS DESCRIPTION

There are in total fourteen datasets of four types of datasets used in our experiments. Our experiment was carried out on four datasets: face image dataset, UCI dataset, handwritten recognition dataset. Table 2 summarizes the characteristics of these datasets used in the experiments.



(a) COIL20

(b) PIE10P

FIGURE 3. Convergence curves of RCAG on COIL20 and PIE10P.

FIGURE 4. warpAR10P dataset samples.

- WarpAR10P contains 130 images of 11 people with different facial expressions under different sunglasses and scarf, and sunglasses and scarf occlusions are used as two types of real world noise.
- Yale university is made up of 2,414 faces of 38 people with various occlusions incurred by different light conditions.
- ORL face database consists of a series of face images taken by Olivetti laboratory in Cambridge, UK, with 40 objects of different ages, genders and races.
- COIL20 [38] contains 20 objects, each rotated 360 degrees horizontally and photographed every 5 degrees, making a total of 72 images for each object.

TABLE 2. Description of datasets.

Dataset category	Dataset	#Instance	#Feature	#Class
Face image dataset	warpAR10P	130	2400	10
	Yale	165	4096	15
	ORL	400	1024	40
	COIL20	1440	1024	20
	PIE10P	210	2420	10
Handwritten recognition dataset	USPS	9298	256	10
	MNIST	1884	2414	58
	Digit	1797	16	10
UCI dataset	Ionosphere	351	34	2
	Wine	178	13	3
	Segment	2310	19	7
	Glass	214	9	6
Biological dataset	GLIOMA	50	4434	4
	TOX_171	171	5748	4

FIGURE 5. PIE10P dataset samples.

FIGURE 7. Yale dataset samples.

FIGURE 6. COIL20 dataset samples.

FIGURE 8. MINIST dataset samples.

- PIE10P carries 10 frontal images from 10 subjects.
- Digit database of 250 samples from 44 writers. We selected 1797 instances from 10 writers.
- MNIST contains 70,000 handwritten pen digits. We used a subset of 1,884 in our experiment.
- USPS is composed of 9,298 handwritten digit images. Each image is represented by a 256 dimensional vector.
- Ionosphere consists of 351 instances and 34 attributes.
- Wine contains 13 properties and three types of wine.
- Segment has 2310 examples, 19 dimensions.
- Glass contains 6 types of glass and defines in terms of their oxide content.
- GLIOMA consists of 50 instances and 4434 attributes.
- TOX_171 consists of 171 instances and 5748 attributes.

C. EVALUATION METRIC

The three evaluation indexes of ACC NMI and Purity were used to evaluate the performance of experiment, and the

TABLE 3. Clustering result measured by ACC.

Dataset category	Dataset	K-means	RPCA	NMF	GNMF	RMNMF	RGNMF	LRRGR	RCAG
Face image dataset	warpAR10P	0.25 ± 0.01	0.25 ± 0.01	0.14 ± 0.03	0.26 ± 0.02	0.30 ± 0.03	0.26 ± 0.02	0.40 ± 0.01	$\mathbf{0.42 \pm 0.01}$
	Yale	0.48 ± 0.01	0.53 ± 0.01	0.35 ± 0.02	0.53 ± 0.01	0.46 ± 0.02	0.60 ± 0.01	$\mathbf{0.63 \pm 0.02}$	0.61 ± 0.03
	ORL	0.56 ± 0.01	0.61 ± 0.02	0.23 ± 0.01	0.62 ± 0.01	0.62 ± 0.01	0.58 ± 0.01	0.60 ± 0.03	$\mathbf{0.62 \pm 0.01}$
	COIL20	0.14 ± 0.01	0.10 ± 0.01	0.25 ± 0.01	0.29 ± 0.01	0.30 ± 0.01	0.40 ± 0.01	0.58 ± 0.01	$\mathbf{0.65 \pm 0.03}$
	PIE10P	0.22 ± 0.01	0.25 ± 0.03	0.26 ± 0.02	0.27 ± 0.01	0.38 ± 0.01	0.39 ± 0.01	0.37 ± 0.02	$\mathbf{0.41 \pm 0.03}$
Handwritten recognition dataset	MNIST	0.18 ± 0.04	0.17 ± 0.01	0.25 ± 0.05	0.19 ± 0.03	0.12 ± 0.03	0.27 ± 0.02	0.27 ± 0.01	$\mathbf{0.28 \pm 0.01}$
	USPS	0.67 ± 0.04	0.69 ± 0.01	0.28 ± 0.05	$\mathbf{0.75 \pm 0.04}$	0.55 ± 0.03	0.54 ± 0.01	0.67 ± 0.03	0.50 ± 0.02
	Digit	0.21 ± 0.04	0.19 ± 0.01	0.35 ± 0.05	0.28 ± 0.04	0.32 ± 0.03	0.35 ± 0.03	$\mathbf{0.56 \pm 0.02}$	0.54 ± 0.02
UCI dataset	Ionosphere	0.63 ± 0.01	0.38 ± 0.01	0.33 ± 0.01	0.42 ± 0.02	0.36 ± 0.02	0.49 ± 0.03	$\mathbf{0.73 \pm 0.01}$	0.72 ± 0.01
	Wine	0.70 ± 0.02	0.71 ± 0.01	0.61 ± 0.01	0.70 ± 0.02	0.62 ± 0.02	$\mathbf{0.72 \pm 0.01}$	0.70 ± 0.01	0.71 ± 0.02
	Segment	0.23 ± 0.01	0.29 ± 0.01	0.19 ± 0.01	0.25 ± 0.01	0.40 ± 0.01	0.50 ± 0.01	0.57 ± 0.02	$\mathbf{0.59 \pm 0.01}$
	Glass	0.54 ± 0.01	0.25 ± 0.02	0.23 ± 0.01	0.45 ± 0.02	0.35 ± 0.03	0.54 ± 0.01	0.55 ± 0.01	$\mathbf{0.58 \pm 0.02}$
Biomedical dataset	TOX_171	0.39 ± 0.04	0.37 ± 0.01	0.34 ± 0.05	0.39 ± 0.04	0.42 ± 0.03	0.43 ± 0.05	0.42 ± 0.03	$\mathbf{0.43 \pm 0.02}$
	GLIOMA	0.29 ± 0.04	0.27 ± 0.01	0.33 ± 0.05	0.29 ± 0.04	0.32 ± 0.03	0.35 ± 0.02	0.36 ± 0.01	$\mathbf{0.36 \pm 0.02}$

specific definitions of these evaluation indicators are as following:

Accuracy (ACC) is defined as

$$ACC = \frac{\sum_{i=1}^{n} \delta(map(r_i), l_i)}{n}, \tag{33}$$

where $map(r_i)$ is a permutation mapping function for permuting the cluster labels r_i to match the equivalent labels in the dataset. n denotes the number of the data points, r_i is the cluster predicted label of x_i and l_i is the corresponding true cluster label. $\delta(x, y)$ is the delta function that if $x \neq y$ equals 1, otherwise it equals 0.

The normalized mutual information (NMI) measure between two index sets is defined as following:

$$NMI(Y, C) = \frac{MI(Y, C)}{\sqrt{H(Y)H(C)}}. \tag{34}$$

$H(X)$ is given by:

$$H(X) = \sum_{i=1}^{|X|} p(i) \log p(i), \tag{35}$$

where $p(i) = |X|/N$ is the probability that an object selected randomly from X falls into class X_i.

The mutual information (MI) between the grand-truth label (Y) and cluster results (C) is given by:

$$MI = \sum_{i=1}^{|Y|} \sum_{j=1}^{|C|} P(i, j) \log \frac{P(i, j)}{P(i)P(j)}, \tag{36}$$

where $P(i, j) = |Y_i \cap C_i|/N$.

The cluster Purity is defined as following:

$$Purity = \frac{1}{n} \sum_{i=1}^{c} \max_{j}(n_i^j), \tag{37}$$

where n_i^j represents the number of data points of the i input class assigned to the cluster $C_j (1 \leq j \leq c)$, c is the number of clusters.

D. EXPERIMENTAL PROCESS

In this section, we introduce the process and results of the comprehensive experiment.

1) BENCHMARK ALGORITHMS

Many studies have shown that the matrix factorization-based clustering method has strong clustering ability [18]. In order to evaluate the clustering performance of the proposed method, we compare our algorithm with the following seven clustering algorithms, including K-means [1], robust principal component analysis (RPCA) [39], NMF [15], GNMF [18], RMNMF [21], RGNMF [40], LRRGR [41] to evaluate the effectiveness of RCAG.

1) The K–means clustering method was one of the most widely used clustering methods.
2) Robust principal component analysis (RPCA) was the most widely used dimension reduction techniques.
3) Nonnegative matrix factorization (NMF).
4) Graph regularized nonnegative matrix factorization (GNMF) took into account the nonlinear structure of the data.
5) Robust manifold nonnegative matrix factorization (RMNMF) was an improved graph-based which uses $\ell_{2,1}$ norm to improve robust.
6) Robust graph regularized nonnegative matrix factorization (RGNMF) introduced a sparse error matrix and apply the ℓ_1 norm to solve unreliable regularization.
7) Low-rank representation with graph regularization (LRRGR) proposed a low-rank representation method that incorporates graph regularization.

2) EXPERIMENTAL RESULT

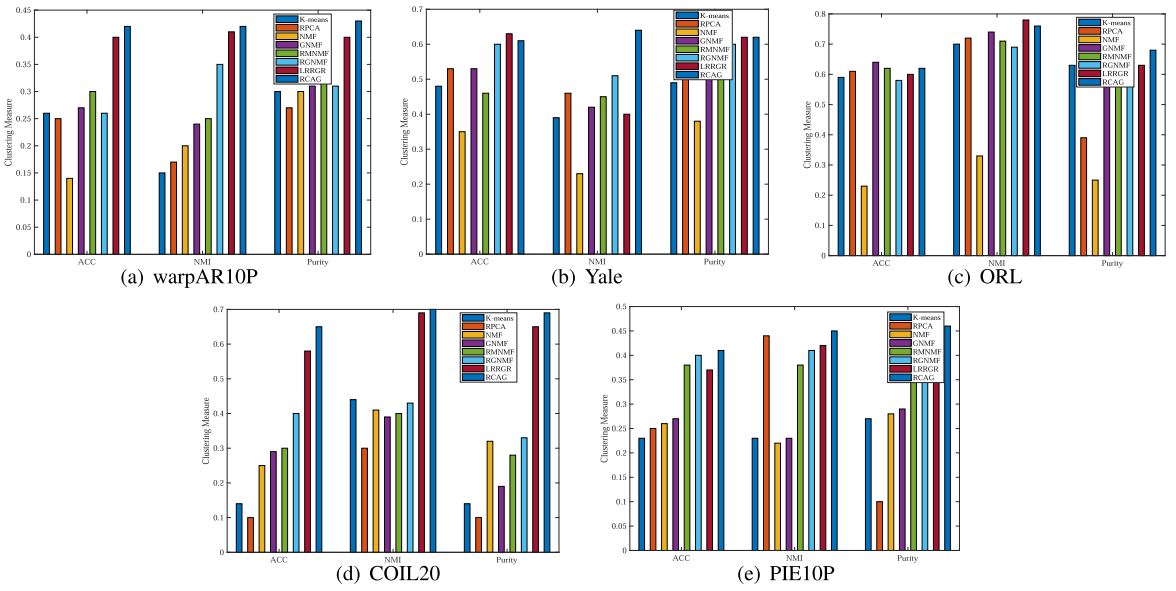
Table 3, 4, 5 tabulates the clustering results of different clustering methods. As we can see, our method achieves good performance on most datasets. For example, on the face image dataset, our method was superior to other comparison algorithms in terms of ACC, NMI and Purity. According to Figure 9 to 12, in some cases, our method dose not achieve the best performance. For example, in handwritten recognition dataset, our method is slightly lower than other comparison methods in ACC, NMI and Purty. The reason may be that the dependence of factorization may cause these steps to lose some important connection. Our method does not perform well on the Biomedical dataset, probably because the Biomedical dataset is too complex and needs preprocessing due to its high dimension.

TABLE 4. Clustering result measured by NMI.

Dataset category	Dataset	K-means	RPCA	NMF	GNMF	RMNMF	RGNMF	LRRGR	RCAG
Face image dataset	warpAR10P	0.15 ± 0.02	0.16 ± 0.01	0.20 ± 0.04	0.24 ± 0.04	0.25 ± 0.03	0.35 ± 0.03	0.41 ± 0.02	$\mathbf{0.42 \pm 0.01}$
	Yale	0.39 ± 0.02	0.46 ± 0.01	0.23 ± 0.03	0.42 ± 0.02	0.45 ± 0.01	0.51 ± 0.01	0.40 ± 0.01	$\mathbf{0.64 \pm 0.02}$
	ORL	0.68 ± 0.01	0.72 ± 0.01	0.33 ± 0.02	0.73 ± 0.02	0.71 ± 0.01	0.69 ± 0.02	$\mathbf{0.78 \pm 0.01}$	0.76 ± 0.01
	COIL20	0.44 ± 0.01	0.30 ± 0.01	0.40 ± 0.01	0.39 ± 0.01	0.40 ± 0.01	0.43 ± 0.01	0.70 ± 0.01	$\mathbf{0.70 \pm 0.03}$
	PIE10P	0.22 ± 0.02	0.44 ± 0.04	0.22 ± 0.02	0.23 ± 0.01	0.35 ± 0.01	0.41 ± 0.02	0.42 ± 0.03	$\mathbf{0.45 \pm 0.02}$
Handwritten recognition dataset	MNIST	0.10 ± 0.03	0.07 ± 0.01	0.12 ± 0.02	0.12 ± 0.04	0.18 ± 0.02	0.12 ± 0.02	0.17 ± 0.01	$\mathbf{0.18 \pm 0.05}$
	USPS	0.59 ± 0.04	0.57 ± 0.02	0.04 ± 0.01	$\mathbf{0.74 \pm 0.01}$	0.36 ± 0.03	0.33 ± 0.03	0.57 ± 0.02	0.60 ± 0.02
	Digit	0.19 ± 0.04	0.17 ± 0.01	0.24 ± 0.05	0.29 ± 0.04	0.20 ± 0.03	0.23 ± 0.03	0.26 ± 0.02	$\mathbf{0.30 \pm 0.02}$
UCI dataset	Ionosphere	0.10 ± 0.01	0.09 ± 0.01	0.08 ± 0.01	0.10 ± 0.02	0.11 ± 0.02	0.23 ± 0.03	$\mathbf{0.24 \pm 0.01}$	0.18 ± 0.03
	Wine	0.41 ± 0.01	0.40 ± 0.01	0.30 ± 0.03	0.40 ± 0.02	0.23 ± 0.02	0.38 ± 0.02	$\mathbf{0.42 \pm 0.01}$	0.36 ± 0.02
	Segment	0.34 ± 0.01	0.36 ± 0.01	0.20 ± 0.01	0.41 ± 0.01	0.30 ± 0.01	0.35 ± 0.02	0.43 ± 0.01	$\mathbf{0.45 \pm 0.01}$
	Glass	0.37 ± 0.02	0.22 ± 0.02	0.04 ± 0.01	0.32 ± 0.03	0.12 ± 0.01	$\mathbf{0.37 \pm 0.03}$	0.35 ± 0.01	0.35 ± 0.02
Biomedical dataset	TOX_171	0.06 ± 0.03	0.17 ± 0.01	0.04 ± 0.05	0.09 ± 0.04	0.11 ± 0.01	0.15 ± 0.03	0.16 ± 0.02	$\mathbf{0.19 \pm 0.02}$
	GLIOMA	0.05 ± 0.04	0.01 ± 0.001	0.006 ± 0.005	0.03 ± 0.004	0.04 ± 0.003	0.05 ± 0.002	0.07 ± 0.001	$\mathbf{0.08 \pm 0.002}$

TABLE 5. Clustering result measured by Purity.

Dataset category	Dataset	K-means	RPCA	NMF	GNMF	RMNMF	RGNMF	LRRGR	RCAG
Face image dataset	warpAR10P	0.30 ± 0.02	0.27 ± 0.01	0.30 ± 0.04	0.31 ± 0.04	0.32 ± 0.03	0.31 ± 0.03	0.40 ± 0.01	$\mathbf{0.43 \pm 0.02}$
	Yale	0.49 ± 0.02	0.54 ± 0.03	0.38 ± 0.03	0.53 ± 0.02	0.51 ± 0.01	0.60 ± 0.03	0.62 ± 0.01	$\mathbf{0.62 \pm 0.02}$
	ORL	0.63 ± 0.01	0.39 ± 0.01	0.25 ± 0.02	0.67 ± 0.02	0.65 ± 0.01	0.62 ± 0.01	0.63 ± 0.03	$\mathbf{0.68 \pm 0.02}$
	COIL20	0.14 ± 0.01	0.10 ± 0.01	0.32 ± 0.01	0.19 ± 0.01	0.28 ± 0.01	0.33 ± 0.01	0.65 ± 0.01	$\mathbf{0.69 \pm 0.03}$
	PIE10P	0.27 ± 0.01	0.10 ± 0.01	0.28 ± 0.02	0.29 ± 0.01	0.42 ± 0.02	0.42 ± 0.01	0.41 ± 0.02	$\mathbf{0.46 \pm 0.04}$
Handwritten recognition dataset	MNIST	0.20 ± 0.04	0.19 ± 0.01	0.24 ± 0.05	0.27 ± 0.04	0.28 ± 0.03	0.29 ± 0.03	0.28 ± 0.01	$\mathbf{0.31 \pm 0.02}$
	USPS	0.73 ± 0.02	0.59 ± 0.02	0.30 ± 0.01	$\mathbf{0.77 \pm 0.02}$	0.54 ± 0.04	0.56 ± 0.01	0.57 ± 0.04	0.58 ± 0.04
	Digit	0.30 ± 0.04	0.17 ± 0.01	0.34 ± 0.05	0.39 ± 0.04	0.32 ± 0.03	$\mathbf{0.43 \pm 0.03}$	0.43 ± 0.01	0.42 ± 0.01
UCI dataset	Ionosphere	0.59 ± 0.04	0.57 ± 0.01	0.64 ± 0.05	0.59 ± 0.04	0.62 ± 0.02	0.71 ± 0.03	$\mathbf{0.75 \pm 0.01}$	$.74 \pm 0.02$
	Wine	0.70 ± 0.01	0.30 ± 0.01	0.64 ± 0.02	0.69 ± 0.01	0.61 ± 0.02	0.72 ± 0.01	0.72 ± 0.02	$\mathbf{0.73 \pm 0.01}$
	Segment	0.34 ± 0.01	0.20 ± 0.01	0.42 ± 0.01	0.29 ± 0.01	0.37 ± 0.01	0.46 ± 0.02	0.53 ± 0.01	$\mathbf{0.58 \pm 0.01}$
	Glass	0.56 ± 0.01	0.32 ± 0.02	0.39 ± 0.01	0.54 ± 0.03	0.40 ± 0.01	0.59 ± 0.03	0.60 ± 0.03	$\mathbf{0.64 \pm 0.02}$
Biomedical dataset	TOX_171	0.15 ± 0.04	0.19 ± 0.01	0.13 ± 0.05	0.27 ± 0.04	0.24 ± 0.03	0.26 ± 0.01	0.25 ± 0.02	$\mathbf{0.27 \pm 0.03}$
	GLIOMA	0.30 ± 0.01	0.32 ± 0.02	0.34 ± 0.05	$\mathbf{0.40 \pm 0.02}$	0.39 ± 0.03	0.38 ± 0.03	0.36 ± 0.01	0.31 ± 0.02



(a) warpAR10P (b) Yale (c) ORL

(d) COIL20 (e) PIE10P

FIGURE 9. Clustering measure on face image dataset.

For the rest, our method achieves good performance over the comparison methods, which demonstrates the necessity and advantage of the introduced $\ell_{2,1}$ norm, ℓ_1 norm and adaptive graph regularization. $\ell_{2,1}$ norm and ℓ_1 norm makes our methods robust to outliers and noise.

The advantages of RCAG are shown in the following two aspects:

(1) The objective function uses the $\ell_{2,1}$ norm as the discrepancy measure, which alleviates the outlier problems common in other clustering methods [42]. And we also apply ℓ_1 norm to sparse error matrix to alleviate the impact of sparse noise on clustering.

(2) The adaptive graph regularization can improve the clustering precision. And it is parameter–insensitive, scale–invariant and simple operation.

3) PERFORMANCE ON CORRUPTED DATA

In this subsection, we consider the dataset with corruption, such as sparse noise and outliers. For this purpose, we use ORL dataset and artificially vanish 20%, 40% entries.

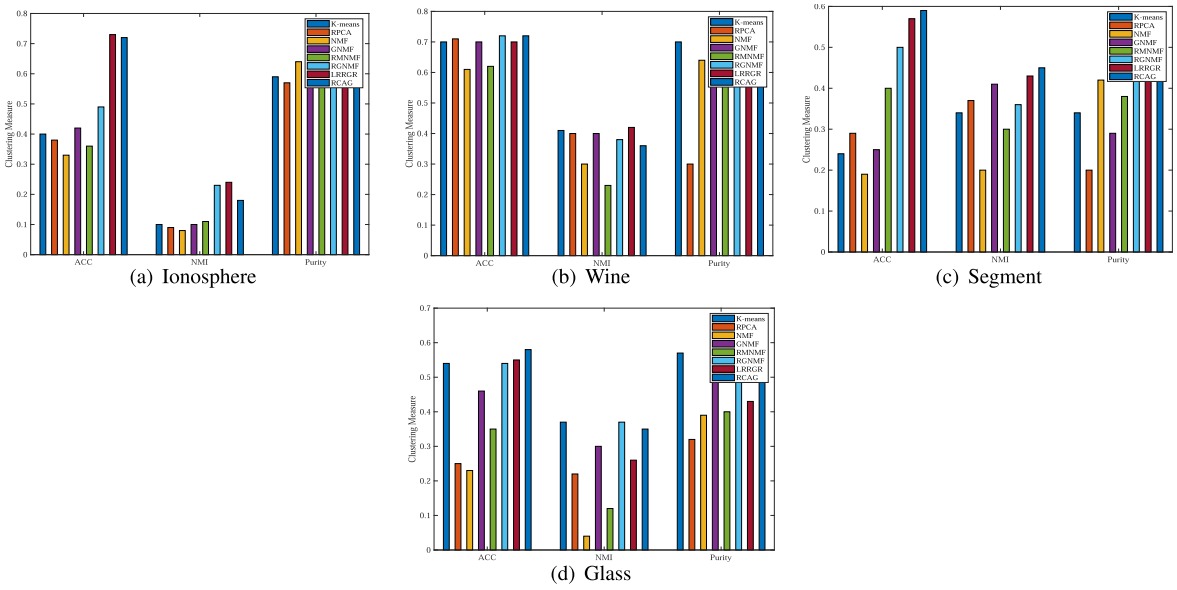


FIGURE 10. Clustering measure on UCI dataset.

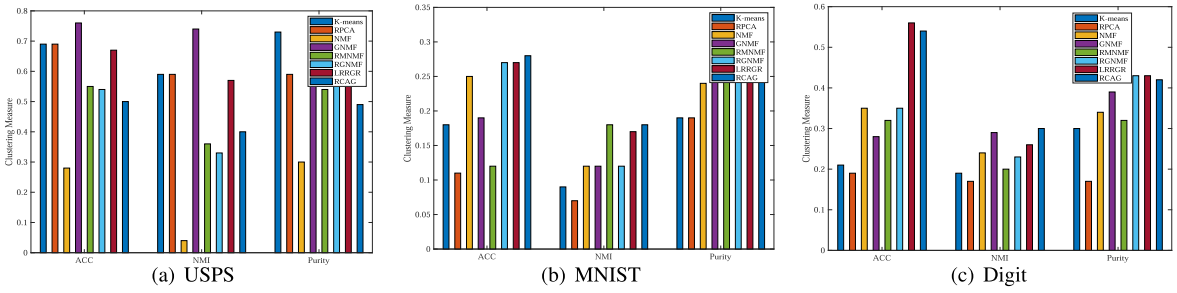


FIGURE 11. Clustering measure on handwritten recognition dataset.

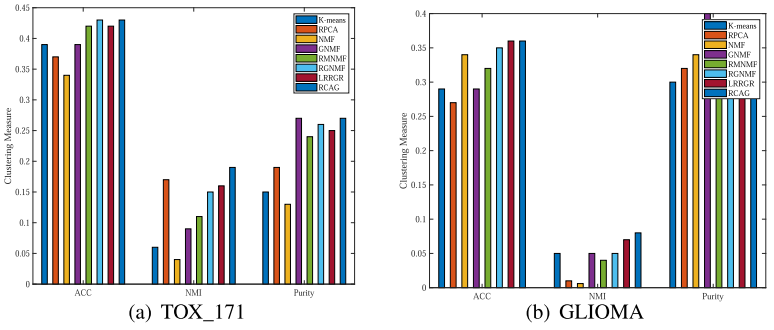


FIGURE 12. Clustering measure on Biomedical dataset.

From Table 6, 7, it is seen that RPCA achieve the best performance in few case. While the RPCA performs best in many cases with 20% corrupted data, in most cases our approach is second best. Meanwhile on 40% corrupted data, RCAG outperforms RPCA.

4) PARAMETER SENSITIVENESS

In this subsection, we test the influence of our method with respect to parameters. Parameter λ is selected by searching from [0.0001, 100], and parameter β varies from [0.0001, 100]. The choice of γ and α is based on [33].

Form Figure 13, we can observe that our method is not very sensitive to the choice of λ, and is sensitive to the choice of β. The RCAG achieves good performance when β varies from 1 to 100.

E. DISCUSSION

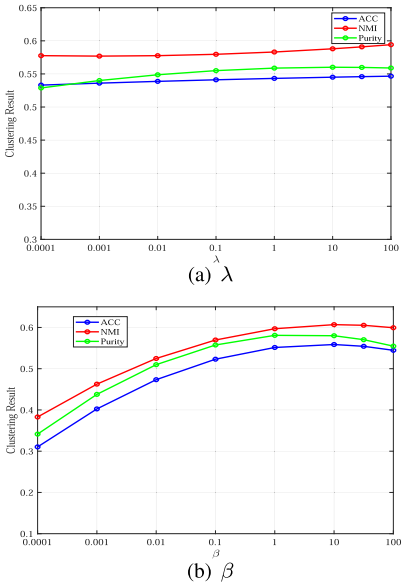
In the above subsections, several experiments on different types of datasets have been performed to show the efficiency of our proposed RCAG.

TABLE 6. Clustering result on ORL with corropted data (20% missing).

Dataset category	K-means	RPCA	NMF	GNMF	RMNMF	RGNMF	LRRGR	RCAG
ACC	0.42 ± 0.01	$\mathbf{0.67 \pm 0.01}$	0.29 ± 0.01	0.59 ± 0.01	0.49 ± 0.02	0.59 ± 0.01	0.62 ± 0.02	0.66 ± 0.01
NMI	0.35 ± 0.03	0.70 ± 0.02	0.20 ± 0.02	0.60 ± 0.02	0.45 ± 0.03	0.40 ± 0.02	0.68 ± 0.01	$\mathbf{0.70 \pm 0.03}$
Purity	0.43 ± 0.01	$\mathbf{0.70 \pm 0.01}$	0.31 ± 0.02	0.61 ± 0.02	0.50 ± 0.01	0.67 ± 0.01	0.63 ± 0.02	0.68 ± 0.02

TABLE 7. Clustering result on ORL with corropted data (40% missing).

Dataset category	K-means	RPCA	NMF	GNMF	RMNMF	RGNMF	LRRGR	RCAG
ACC	0.33 ± 0.02	0.34 ± 0.01	0.24 ± 0.03	0.52 ± 0.01	0.37 ± 0.02	0.44 ± 0.02	0.45 ± 0.01	$\mathbf{0.46 \pm 0.02}$
NMI	0.24 ± 0.03	0.30 ± 0.04	0.20 ± 0.01	0.40 ± 0.02	0.30 ± 0.01	0.38 ± 0.02	0.42 ± 0.02	$\mathbf{0.40 \pm 0.01}$
Purity	0.34 ± 0.01	0.39 ± 0.02	0.28 ± 0.02	0.44 ± 0.03	0.38 ± 0.01	0.50 ± 0.01	0.43 ± 0.03	$\mathbf{0.52 \pm 0.01}$



(a) λ

(b) β

FIGURE 13. Clustering result with varying λ and β on Yale dataset.

The adaptive graph regularization clustering method performs better than the baseline K-means. Adaptive graph regularization not only capture the global structures of data, but can preserve the local geometric structures, i.e., preserve nonlinear structure.

Compared with NMF and GNMF, RPCA, etc., our proposed model uses $\ell_{2,1}$ norm and ℓ_1 norm. $\ell_{2,1}$ norm keeps feature rotation invariance and can alleviate the influence of outliers. ℓ_1 norm reduces the influence of sparse noise includes impulse noise (salt and pepper), dead line, stripes [43].

V. CONCLUSION AND FUTURE WORK

In this paper, a low–rank matrix factorization model with noise and outliers based on adaptive graph regularization is proposed. Our model can not only solve sparse noise and outliers, but also improve clustering performance by adaptive graph regularization. We introduce spares error matrix S and ℓ_1 norm to solve sparse noise problem. By using the sparse error matrix, a large amount of data can be reconstructed to obtain robust decomposition results. In addition, the $\ell_{2,1}$ norm is applied to the matrix decomposition, and a robust solution for outliers. Therefore, RCAG approximates the clean data reconstructed from sparse outliers, constrains outliers by $\ell_{2,1}$ norm, and constrains sparse noise by ℓ_1 norm to achieve robustness. It proposes an iterative updating method to optimize problem, and proved to be convergent. Experimental results show the effectiveness of the RCAG. Nevertheless, there are still some limitations to our approach. For example, because of the shrinkage effect, ℓ_1 norm usually results in a biased estimator which affects the accuracy of the matrix rank approximation.

Next, we will carry out follow–up work to study various epidemic regulation to improve clustering performance [44], [45]. Such as normalized ε-penalty solves sparse corruption, include impulse noise, deadline and stripes [46], side information and low rank constraint [47]. Normalized ε-penalty can can replace the ℓ_1 norm, and can enhance the sparsity in both the intrinsic low–rank structure and the sparse corruptions. Semi-supervised clustering is realized by adding side information. Adding a kernel norm constraint of the objective functions.

REFERENCES

[1] J. Hartigan and M. Wong, "A k-means clustering algorithm," *Appl. Stat.*, vol. 28, no. 1, pp. 100–108, 2013.
[2] U. von Luxburg, "A tutorial on spectral clustering," *Statist. Comput.*, vol. 17, no. 4, pp. 395–416, Dec. 2007.
[3] S. Zhou, E. Zhu, X. Liu, T. Zheng, Q. Liu, J. Xia, and J. Yin, "Subspace segmentation-based robust multiple kernel clustering," *Inf. Fusion*, vol. 53, pp. 145–154, Jan. 2020.
[4] S. Zhou, X. Liu, C. Zhu, Q. Liu, and J. Yin, "Spectral clustering-based local and global structure preservation for feature selection," in *Proc. Int. Joint Conf. Neural Netw. (IJCNN)*, Jul. 2014, pp. 550–557.
[5] C. H. Q. Ding, T. Li, and M. I. Jordan, "Convex and semi-nonnegative matrix factorizations," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 32, no. 1, pp. 45–55, Jan. 2010.
[6] X. Peng, H. Zhu, J. Feng, C. Shen, H. Zhang, and J. T. Zhou, "Deep clustering with sample-assignment invariance prior," *IEEE Trans. Neural Netw. Learn. Syst.*, early access, Dec. 21, 2019, doi: 10.1109/TNNLS.2019.2958324.
[7] X. Peng, J. Feng, S. Xiao, W.-Y. Yau, J. T. Zhou, and S. Yang, "Structured AutoEncoders for subspace clustering," *IEEE Trans. Image Process.*, vol. 27, no. 10, pp. 5076–5086, Oct. 2018.
[8] X. Peng, Z. Huang, J. Lv, H. Zhu, and J. T. Zhou, "COMIC: Multi-view clustering without parameter selection," in *Proc. Int. Conf. Mach. Learn.*, vol. 97, Jun. 2019, pp. 5092–5101.
[9] X. Peng, C. Lu, Z. Yi, and H. Tang, "Connections between nuclear-norm and Frobenius-Norm-Based representations," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 29, no. 1, pp. 218–224, Jan. 2018.
[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. Mach. Intell.*, vol. 35, no. 1, pp. 171–184, Jan. 2013.

[11] I. T. Jolliffe, "Principal component analysis," *J. Marketing Res.*, vol. 87, no. 4, p. 513, 1986.
[12] A. Höcker and V. Kartvelishvili, "SVD approach to data unfolding," *Nucl. Instrum. Methods Phys. Res. A, Accel., Spectrometers, Detectors Associated Equip.*, vol. 372, no. 3, pp. 469–481, Apr. 1996.
[13] L. Wang and M. Dong, "Exemplar-based low-rank matrix decomposition for data clustering," *Data Mining Knowl. Discovery*, vol. 29, no. 2, pp. 324–357, Mar. 2015.
[14] K. Allab, L. Labiod, and M. Nadif, "A Semi-NMF-PCA unified framework for data clustering," *IEEE Trans. Knowl. Data Eng.*, vol. 29, no. 1, pp. 2–16, Jan. 2017.
[15] D. D. Lee and H. S. Seung, "Learning the parts of objects by non-negative matrix factorization," *Nature*, vol. 401, no. 6755, pp. 788–791, Oct. 1999.
[16] H. Liu, Z. Wu, D. Cai, and T. S. Huang, "Constrained nonnegative matrix factorization for image representation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 7, pp. 1299–1311, Jul. 2012.
[17] D. Cai, X. He, X. Wu, and J. Han, "Non-negative matrix factorization on manifold," in *Proc. 8th IEEE Int. Conf. Data Mining*, Dec. 2008, pp. 63–72.
[18] D. Cai, X. He, J. Han, and T. S. Huang, "Graph regularized nonnegative matrix factorization for data representation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 8, pp. 1548–1560, Aug. 2011.
[19] S. Zhou, X. Liu, M. Li, E. Zhu, L. Liu, C. Zhang, and J. Yin, "Multiple kernel clustering with neighbor-kernel subspace segmentation," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 31, no. 4, pp. 1351–1362, Apr. 2020.
[20] Q. Gu and J. Zhou, "Local learning regularized nonnegative matrix factorization," in *Proc. 21st Int. Joint Conf. Artif. Intell. (IJCAI)*, Jan. 2009, pp. 1046–1051.
[21] J. Huang, F. Nie, H. Huang, and C. Ding, "Robust manifold nonnegative matrix factorization," *ACM Trans. Knowl. Discovery Data*, vol. 8, no. 3, p. 11, Jun. 2014.
[22] F. Sun, M. Xu, X. Hu, and X. Jiang, "Graph regularized and sparse nonnegative matrix factorization with hard constraints for data representation," *Neurocomputing*, vol. 173, pp. 233–244, Jan. 2016.
[23] J. Sun, Z. Wang, F. Sun, and H. Li, "Sparse dual graph-regularized NMF for image co-clustering," *Neurocomputing*, vol. 316, pp. 156–165, Nov. 2018.
[24] N. Yu, Y.-L. Gao, J.-X. Liu, J. Wang, and J. Shang, "Hypergraph regularized NMF by $L_{2,1}$-norm for clustering and com-abnormal expression genes selection," in *Proc. IEEE Int. Conf. Bioinf. Biomed. (BIBM)*, Dec. 2018, pp. 578–582.
[25] M.-J. Wu, Y.-L. Gao, J.-X. Liu, C.-H. Zheng, and J. Wang, "Integrative hypergraph regularization principal component analysis for sample clustering and co-expression genes network analysis on multi-omics data," *IEEE J. Biomed. Health Informat.*, vol. 24, no. 6, pp. 1823–1834, Jun. 2020.
[26] H. Liu, X. Li, and X. Zheng, "Solving non-negative matrix factorization by alternating least squares with a modified strategy," *Data Mining Knowl. Discovery*, vol. 26, no. 3, pp. 435–451, May 2013.
[27] C.-Y. Wang, J.-X. Liu, N. Yu, and C.-H. Zheng, "Sparse graph regularization non-negative matrix factorization based on huber loss model for cancer data analysis," *Frontiers Genet.*, vol. 10, no. 11, pp. 1–11, Nov. 2019.
[28] X. Guo, X. Xie, G. Liu, M. Wei, and J. Wang, "Robust low-rank subspace segmentation with finite mixture noise," *Pattern Recognit.*, vol. 93, pp. 55–67, Sep. 2019.
[29] S. Zhou, X. Liu, Q. Liu, S. Wang, C. Zhu, and J. Yin, "Random Fourier extreme learning machine with $\ell_{2,1}$-norm regularization," *Neurocomputing*, vol. 174, pp. 143–153, Jan. 2016.
[30] X. He, Q. Wang, and X. Li, "Robust adaptive graph regularized nonnegative matrix factorization," *IEEE Access*, vol. 7, pp. 83101–83110, 2019.
[31] F. Nie, H. Huang, X. Cai, and C. H. Q. Ding, "Efficient and robust feature selection via joint $\ell_{2,1}$-norms minimization," in *Proc. Adv. Neural Inf. Process. Syst.*, 2010, pp. 1813–1821.
[32] F. Zhang, Y. Lu, J. Chen, S. Liu, and Z. Ling, "Robust collaborative filtering based on non-negative matrix factorization and R_1-norm," *Knowl.-Based Syst.*, vol. 118, pp. 177–190, Feb. 2017.
[33] F. He, F. Nie, R. Wang, X. Li, and W. Jia, "Fast semisupervised learning with bipartite graph for large-scale data," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 31, no. 2, pp. 626–638, Feb. 2020.
[34] Z. Lin, M. Chen, and Y. Ma, "The augmented Lagrange multiplier method for exact recovery of corrupted low-rank matrices," 2010, *arXiv:1009.5055*. [Online]. Available: http://arxiv.org/abs/1009.5055
[35] J. Yang and X. Yuan, "Linearized augmented lagrangian and alternating direction methods for nuclear norm minimization," *Math. Comput.*, vol. 82, no. 281, pp. 301–329, Mar. 2012.
[36] S. Zhan, J. Wu, N. Han, J. Wen, and X. Fang, "Unsupervised feature extraction by low-rank and sparsity preserving embedding," *Neural Netw.*, vol. 109, pp. 56–66, Jan. 2019.
[37] L. Zhang, Q. Zhang, B. Du, J. You, and D. Tao, "Adaptive manifold regularized matrix factorization for data clustering," in *Proc. 26th Int. Joint Conf. Artif. Intell.*, Aug. 2017, pp. 3399–3405.
[38] S. A. Nene, S. K. Nayar, and H. Murase, "Columbia object image library (coil-20)," Columbia Univ. Image Library, Manhattan, NY, USA, Tech. Rep. CUCS-005-96, Feb. 1996.
[39] T. Zhang and Y. Yang, "Robust PCA by manifold optimization," *J. Mach. Learn. Res.*, vol. 19, no. 80, pp. 1–39, 2018.
[40] S. Huang, H. Wang, T. Li, T. Li, and Z. Xu, "Robust graph regularized nonnegative matrix factorization for clustering," *Data Mining Knowl. Discovery*, vol. 32, no. 2, pp. 483–503, Mar. 2018.
[41] W. He, J. X. Chen, and W. Zhang, "Low-rank representation with graph regularization for subspace clustering," *Soft Comput.*, vol. 21, no. 6, pp. 1569–1581, Mar. 2017.
[42] D. Kong, C. Ding, and H. Huang, "Robust nonnegative matrix factorization using L21-norm," in *Proc. 20th ACM Int. Conf. Inf. Knowl. Manage. (CIKM)*, 2011, pp. 673–682.
[43] H. Ye, H. Li, B. Yang, F. Cao, and Y. Tang, "A novel rank approximation method for mixture noise removal of hyperspectral images," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 7, pp. 4457–4469, Jul. 2019.
[44] M. Belkin, P. Niyogi, and V. Sindhwani, "Manifold regularization: A geometric framework for learning from labeled and unlabeled examples," *J. Mach. Learn. Res.*, vol. 7, pp. 2399–2434, Nov. 2006.
[45] D. Cai, X. Wang, and X. He, "Probabilistic dyadic data analysis with local and global consistency," in *Proc. 26th Annu. Int. Conf. Mach. Learn. (ICML)*, 2009, p. 14.
[46] T. Xie, S. Li, and B. Sun, "Hyperspectral images denoising via nonconvex regularized low-rank and sparse matrix decomposition," *IEEE Trans. Image Process.*, vol. 29, pp. 44–56, 2020.
[47] K.-Y. Chiang, I. S. Dhillon, and C.-J. Hsieh, "Using side information to reliably learn low-rank matrices from missing and corrupted observations," *J. Mach. Learn. Res.*, vol. 19, no. 76, pp. 1–35, 2018.

MIN ZHAO was born in Yantai, China, in 1995. She received the Bachelor of Science degree from Yantai University, in 2018, where she is currently pursuing the M.S. degree in computer science and technology. Her main research interest includes clustering.

JINGLEI LIU received the B.S. and M.S. degrees from the Department of Automation, Taiyuan University of Technology, and the Ph.D. degree from the Department of Computer Science and Technology, Tianjin University, China. He is currently a Professor with Yantai University, China. His research interests include artificial intelligence and theoretical computer science, artificial intelligence aspect in knowledge representation and reasoning for graph model, such as CP-nets and coalition structure graph, and theoretical computer science aspect in finite model theory which focus on logical properties of algorithm, such as expressive power and complexity of specific logic on finite mathematical structure.