Coupled low rank representation and subspace clustering

Stanley Ebhohimhen Abhadiomhen[1,2] · ZhiYang Wang[1] · XiangJun Shen[1]

Abstract

Subspace clustering is a technique utilized to find clusters within multiple subspaces. However, most existing methods cannot obtain an accurate block diagonal clustering structure to improve clustering performance. This drawback exists because these methods learn the similarity matrix in advance by utilizing a low dimensional matrix obtained directly from the data, where two unrelated data samples can stay related easily due to the influence of noise. This paper proposes a novel method based on coupled low-rank representation to tackle the above problem. First, our method constructs a manifold recovery structure to correct inadequacy in the low-rank representation of data. Then it obtains a clustering projection matrix that obeys the k-block diagonal property to learn an ideal similarity matrix. This similarity matrix denotes our clustering structure with a rank constraint on its normalized Laplacian matrix. Therefore, we avoid k-means spectral post-processing of the low dimensional embedding matrix, unlike most existing methods. Furthermore, we couple our method to allow the clustering structure to adaptively approximate the low-rank representation so as to find more optimal solutions. Several experiments on benchmark datasets demonstrate that our method outperforms similar state-of-the-art methods in Accuracy, Normalized Mutual Information, F-score, Recall, Precision, and Adjusted Rand Index evaluation metrics.

Keywords Low-rank representation · Subspace clustering · Adaptive clustering structure · Block diagonal

1 Introduction

Subspace clustering methods find clusters within multiple subspaces of unlabeled data. As a result, four major groups of studies on subspace clustering have persisted over the years, namely applied statistics [25, 30, 42], factorization [8, 13, 29, 31, 45], algebraic [32, 33, 35], and spectral clustering [4, 14, 15, 38, 50]. However, the spectral-type, which requires constructing a similarity matrix using data's low dimensional representation, has become increasingly popular in the last decade. Thus, low-rank representation

(LRR) [19] and sparse subspace clustering (SSC) [10] approaches are known to be efficient in constructing the similarity matrix [1, 20]. The reason is that both of them utilize the self-expressiveness property of data to represent each data sample as a linear combination of others. Therefore, several studies inspired by SSC and LRR have surfaced more recently. In [5], Chen et al. proposed a collaborative representation method using the LRR recovery technique. In [46], Zhang et al. proposed a hyperspectral remote sensing image representation method based on SSC. Other studies, such as [22, 37] presented various subspace methods, combining LRR and SSC approaches. Similarly, Shen and Li [28] proposed the NLRR method by combining LRR and matrix factorization.

Notwithstanding, the above are two-phase methods, which require post-processing of the similarity matrix via spectral clustering algorithm [26]. That means the clustering performance of these methods depends on the accuracy of the predefined similarity matrix. Regarding this limitation, the studies in [17, 21, 27] proposed methods, which learn the similarity matrix cooperatively to improve clustering performance. Even so, these methods obtain similarity between data samples by assuming that similar samples

✉ XiangJun Shen
xjshen@ujs.edu.cn

Stanley Ebhohimhen Abhadiomhen
5103190343@stmail.ujs.edu.cn;
stanley.abhadiomhen@unn.edu.ng

1   School of Computer Science and Communication
    Engineering, JiangSu University, JiangSu, 212013 China

2   Department of Computer Science, University of Nigeria,
    Nsukka, Nigeria

reside near each. Therefore, they often do not perform well in practical cases when a given data has an arbitrary distribution. This drawback exists because two data samples may seem similar by their spatial location yet belong to different subspaces [39], which means two unrelated data samples can then reconstruct one another. Although the methods presented in [23, 40, 41, 49] utilize nonnegative constraints to encourage the similarity matrix to be block diagonal, they also follow a two-phase procedure to obtain the clustering structure. So, these methods' discrimination ability would degrade when the similarity matrix produced in the first phase fails to hold an accurate block diagonal shape

To this end, we propose a novel method, which guarantees an ideal clustering structure through coupled low-rank representation to enhance clustering performance. First, our method obtains a manifold recovery structure, which corrects inadequacy in the low-rank representation of data using the k-block-diagonal regularizer [23]. This manifold recovery structure is directly useful for the subspace clustering task because our method can instantly find a clustering projection matrix that obeys the k-block diagonal property by utilizing it. Unlike the common block diagonal property learned by spectral-based methods, the k-block diagonal property acquired through the clustering projection matrix guarantees the within-cluster consistency and independence between clusters. Hence, our method finds an ideal similarity matrix that approximates the clustering projection matrix with a rank constraint on the normalized Laplacian matrix to ensure that it denotes our clustering structure with an accurate k number of connected components. This approach avoids k-means post-processing of the low dimensional embedding matrix, which is inherent in spectral clustering algorithms. Additionally, we couple our method to allow the clustering structure to approximate the low-rank representation adaptively in order to capture a more accurate manifold structure of data. Therefore, the proposed method can achieve better performance in practice since the cleaner low-rank matrix guarantees a better clustering structure.

The main contributions are as follows:

1) We propose a new method, which learns an ideal similarity matrix through a clustering projection matrix. This approach guarantees more discrimination ability than existing subspace methods, which learn the similarity matrix in advance by utilizing a low dimensional representation obtained directly from the original data. Besides, our method avoids k-means spectral post-processing of the low-dimensional embedding matrix by imposing a rank constraint on the normalized Laplacian matrix to allow the similarity matrix to denote our clustering structure explicitly with k connected components.

2) Furthermore, we couple our method to allow mutual capturing of the data's manifold structure by the clustering structure and the low-rank representation.

3) To evaluate our method, we perform a wide range of experiments on several benchmark datasets. The results show that our method outperforms similar state-of-the-art methods in six evaluation metrics.

The rest of this paper is structured as follows. Section 2 provides the background and a review of the related works. In Section 3, we present the proposed method. Section 4 demonstrates the effectiveness of the proposed method. Then, Section 5 gives the conclusion.

2 Background and related work

In this section, we provide the main notations and a review of the related works.

2.1 Notation

In this paper, we denote matrices with an uppercase letter. $\| .\|_*$ denotes nuclear norm, $\| .\|_1$ denotes L_1-norm, $\| .\|_F$ denotes Frobenius norm, $\| .\|_k$ denotes k-block diagonal inducing regularizer, $\| .\|_{2,1}$ denotes $L_{2,1}$-norm, and $\text{Tr}(.)$ denotes the trace of a square matrix. Furthermore, we denote the $(i, j)th$ entry of a matrix Z by $Z_{i,j}$, where i and j represent the ith row and jth column. $Z \geq 0$ implies that all entries of Z are nonnegative. $\text{Diag}(Z)$ denotes a diagonal matrix whose element is the ith diagonal entry of Z. Then I denote an identity matrix.

2.2 Related work

Given an unlabeled dataset $X = \{x_1, \ldots, x_n\} \in R^{d \times n}$ drawn from a union of subspaces, the SSC approach uses L_1-norm regularization to find a sparse representation of the data samples. However, the L_1-norm does not capture a global data structure since it obtains a sparse representation independently for each data sample [2]. LRR approach, on the other hand, captures a global data structure using the nuclear norm regularization as follows:

$$\min_{Z,E} \| Z\|_* + \lambda \| E\|_{2,1},$$
$$s.t., \ X = XZ + E,$$

(1)

where Z denotes the low-rank representation, X denotes the self-dictionary, E denotes the error matrix. λ is a parameter to balance the first and second terms, and the $\| .\|_{2,1}$ norm used in (1) is very efficient when only a part of the samples is contaminated by noise. $Z_{i,j}$ can then correctly capture the data's local manifold structure when only similar data samples x_i and x_j belonging to the same cluster reconstruct

one another. Therefore, considering its robustness, several studies inspired by LRR, such as [5, 6, 9, 18, 28, 34, 49], have been conducted over the years. In [34], Vidal and Favaro proposed a method, which attempts to find a clean low-rank matrix by decomposing a data matrix into three matrices: a noise-free matrix, a corrupted matrix, and an error matrix. In [28], Shen and Li proposed a method, which factorizes a nuclear norm regularized matrix to capture the exact similarity between data samples. In [49], Zheng et al. proposed a method named SinNLRR, which imposes a non-negative constraint on the low-rank matrix to ensure that only similar data samples are connected while unrelated data samples are not connected. Also presented is a low-rank local embedding representation (LRLER) method [9]. LRLER uses local manifold embedding constraint to enforce a global low-rank representation since most data often do not hold a linear structure. So, a method named LRLTSER was introduced together with LRLER to handle neighborhood aliasing distortion in data. Besides, Lu et al. [23] introduced the Block Diagonal Representation (BDR) method, which uses the k-block diagonal regularizer to encourage a proper low dimensional matrix.

Upon learning the low-dimensional representation of data, the above methods can directly construct a similarity matrix. Afterward, a spectral clustering algorithm such as [26] is applied to obtain the clustering structure. Thus, it is easily noticeable that the accuracy of the clustering structure learned in the two-phase process will depend entirely on the quality of the similarity matrix obtained in the first phase. Towards this end, many other studies, such as [21, 27, 40], proposed methods, which learn a similarity matrix cooperatively in search of an optimum solution. For example, given an initial similarity matrix A, the Constrained Laplacian Rank (CLR) [27] can obtain an ideal similarity matrix S by learning S and A simultaneously. Similarly, the Extreme Learning Machine (ELM-CLR) [21] and the Implicit Block Diagonal Low-Rank Representation (IBDLR) [40] can find an ideal similarity matrix by adapting it on dual data representations where one data representation relies on the actual data, and another on its affine space. Meanwhile, Yang et al. [44] proposed a model based on a low-rank variation dictionary (LRVD) to improve performance accuracy. This approach constructs the dictionary independently using variations of data. In [47], Zhang et al. presented a method named CMC by applying nonnegative matrix factorization. CMC mainly focuses on capturing actual similarity between data samples even with limited samples.

The proposed method is comparable to IBDLR since they both apply the same k-block diagonal regularizer [23] and LRR approaches to solve the subspace segmentation problem. However, our approach can guarantee a more discriminating ability with explicit block diagonal

clustering structure through coupled LRR without spectral post-processing procedure. Furthermore, a coupling strategy was used in a previous study in [48], but it relied on additional k-means post-processing to obtain the clustering results.

3 The proposed method

In this section, we present our proposed method based on coupled low-rank representation. First, we formulate the model in Section 3.1. Then propose an efficient optimization method to solve it in Section 3.2.

3.1 Model formulation

3.1.1 Relationship between data

LRR based methods or any other subspace method would efficiently find a low dimensional matrix that preserves the relationship between data samples when underlying data is clean. Because only similar data samples will reconstruct one another using the data's self-expressiveness property. On the other hand, the self-expressiveness property alone will not guarantee an accurate low dimensional matrix when the data is noisy. The reason is that two unrelated data samples, x_i and x_j, may be selected to reconstruct themselves where the singular value Z_{ij} is non-zero instead of 0. Assuming this to be the case, it means the similarity matrix learned after using the low-dimensional matrix would be fallible— and the goal of subspace clustering to partition the data samples into k appropriate clusters will then be defeated. The ideal is for only similar samples from the same class to be connected, while the samples from different classes should not stay together. To this end, we consider a block diagonal compliant model directly such that a manifold recovery structure obtained can correct a possible inadequacy in the low-rank representation of data as follows:

$$\min_{Z,E,L} \| Z \|_* + \lambda_1 \| E \|_{2,1} + \lambda_2 \| L \|_k,$$
$$s.t., \ X = XZ + E, L = (|Z| + |Z^T|)/2, Diag(L) = 0, L \geq 0, \quad (2)$$

where $X \in R^{d \times n}$ represents the self-dictionary. $Z \in R^{n \times n}$ denotes the low-rank representation. $L \in R^{n \times n}$ is the manifold recovery structure. $Diag(L) = 0$ and $L \geq 0$ are non-negative constraints to enforce a diagonal block structure on L. $E \in R^{d \times n}$ is the error matrix whose columns are encouraged to be zero by $\| . \|_{2,1}$, assuming that the data corruption is precise. As such, this error matrix is incredibly efficient since noise, or noise-free element, is generally not known in advance. Thus, most subspace methods, such as LRR in (1), usually estimate Z using X. However, when X is noisy, Z's estimation becomes biased, which means Z

cannot effectively capture the similarity between the data samples. The core idea in (2) attempts to recover the actual manifold structure from Z through the k-block diagonal regularizer defined below so that L_{ij} is only 0 for two unrelated data samples, i and j.

Definition 1 [23] (k-Block Diagonal Regularizer): For any similarity matrix $L \in R^{n \times n}$ the k-block diagonal regularizer is defined as the sum of the smallest k eigenvalues of L_L, i.e.

$$\| L \|_k = \sum_{i=n-k+1}^{n} \lambda_i(L_L), \tag{3}$$

where $L_L = Diag(L1) - L$ is the Laplacian matrix of L. $\lambda_i(L_L)$ is ith eigenvalue of L_L in the descending order. This k-block diagonal regularizer performs two functions: First, it encourages L to be block-diagonal with nonnegative constraints mentioned above. Secondly, it ensures that there is k exact number of connected blocks in L.

3.1.2 Clustering Structure with constrained Laplacian rank

Ideally, an optimal L by (2) denotes the similarity matrix of existing methods, made possible by $L = (|Z| + |Z^T|)/2$. On the contrary, our method obtains the similarity matrix through a clustering projection matrix to guarantee more discrimination capability. Thus, considering that L has an inherent Laplacian structure L_L by definition in (3), the clustering projection matrix is obtained directly from it in the following way:

$$Tr(H^T L H),$$
$$s.t., \quad H^T H = I \tag{4}$$

Apparently, H is similar to the primary output of a spectral clustering algorithm in the two-phase methods. Nonetheless, let us consider that the k-block diagonal regularization on L may not guarantee an exactly k-block diagonal structure as suggested in [40]. Then it is easy to see that the constraint $H^T H = I$ only enforces the common block diagonal property on H. That also means $H^T H = I$ can only guarantee within-cluster consistency without considering overlapping between clusters. To mitigate against this problem, we observe that the binary matrix HH^T obeys the k-block-diagonal property by itself, which hints that HH^T can guarantee both the within-cluster consistency and independence between clusters. As a result, an ideal similarity matrix $S = \{s_i^T, \ldots, s_n^T\} \in R^{n \times n}$ is redefined, which is not only approximate to HH^T but denotes our clustering structure with $rank(L_s) = n - k$ through the following theorem.

Theorem 1: if S is non-negative, the multiplicity k of the zero eigenvalue of the graph Laplacian L_s corresponds to the number of connected components in the graph associated with S.

Theorem 1 implies that if $rank(L_s) = n - k$, then S becomes an ideal similarity matrix from which the data samples can be partitioned into k clusters explicitly without spectral post-processing of S. The proof of Theorem 1 can be found in [7, 36].

Hence, we have the following formulation:

$$\min_{L,H,S} \| S - HH^T \|_F^2 + \lambda_3 Tr(H^T L H),$$

$$s.t., \ H^T H = I, S1 = 1, S \geq 0, rank(L_s) = n - k, \tag{5}$$

where $S1 = 1$ ensures that each data sample can only belong to one cluster, then $L_S = D_S - S$ denotes the normalized Laplacian matrix because $D_S \in R^{n \times n}$ represents the diagonal matrix whose ith entry $\sum_j S_{i,j} = 1$. Besides, HH^T denotes the clustering projection matrix, where $(HH^T)_{ij}$ is 1 if $h_i = h_j$ and 0 otherwise. In other words, $(HH^T)_{ij} = 1$ indicates that x_i and x_j belongs to the same cluster.

3.1.3 Coupled low rank representation

Achieving the objective of (5) may still be challenging since the manifold structure is pre-captured analogously to the two-phase methods such as SinNLRR [49], BDR [23], and IBDLR [40]. Thus, the idea of dynamic approximation is introduced into our model using the coupling term $\| S - Z \|_F^2$ to allow for joint capturing of the data's manifold structure by both S and Z. I.e., $\| S - Z \|_F^2 \to 0$ as much as possible, leading to the problem of coupled low-rank representation and subspace clustering:

$$\min_{Z,H,L,S,E} \| Z \|_* + \lambda_2 \| L \|_k + \lambda_4 \| S - HH^T \|_F^2$$
$$+ \lambda_3 Tr(H^T L H) + \lambda_5 \| S - Z \|_F^2 + \lambda_1 \| E \|_{2,1},$$
$$s.t., \ X = XZ + E, L = (|Z| + |Z^T|)/2, Diag(L) = 0, L \geq 0,$$
$$H^T H = I, S1 = 1, S \geq 0, rank(L_s) = n - k, \tag{6}$$

where parameters $\lambda_1, \lambda_2, \lambda_3, \lambda_4,$ and λ_5 controls the tradeoff between the terms, respectively.

The goal of the above objective function is to make a perfect subspace segmentation more possible by simultaneously learning S and Z. Therefore, our method uses the coupling term $\| S - Z \|_F^2$ to measure the effect of the k-block diagonal representation on S. Because once the obtained Z varies largely to S, it indicates that Z is not clean enough. In that case, our method finds a cleaner Z that captures a more accurate manifold structure of the data so that S can then also capture a better manifold structure correspondingly. Besides, this term allows S and Z subproblems to be strongly convex, making convergence analysis very easy.

3.2 Optimization

We propose an efficient optimization method based on Augmented Lagrange Multiplier (ALM) [16] to solve (6).

First, (6) is not easily solvable because of the non-convex nature of $\| L \|_k$. So, following Ky Fan's theorem [12], $\| L \|_k$ is reformulated as a convex problem in the following way:

$$
\| L \|_k = \min_W (L_L, W), \\
s.t., \ 0 \leq W \leq I, Tr(W) = k, \tag{7}
$$

By substituting (7) in place of $\| L \|_k$, we rewrite (6) as:

$$
\min_{Z,W,H,L,S,E} \| Z \|_* + \lambda_2 (Diag(L1) - L, W) + \lambda_4 \| S - HH^T \|_F^2 \\
+ \lambda_3 Tr(H^T LH) + \lambda_5 \| S - Z \|_F^2 + \lambda_1 \| E \|_{2,1}, \\
s.t., \ X = XZ + E, L = (|Z| + |Z^T|)/2, Diag(L) = 0, L \geq 0, \\
H^T H = I, S1 = 1, S \geq 0, rank(L_s) = n - k, \\
0 \leq W \leq I, Tr(W) = k, \\
. \tag{8}
$$

Equation 8 may still be hard to solve because $rank(L_s) = n - k$ is not linear. Resolving this problem, we denote the i-th smallest eigenvalue of L_s as $\theta_i(L_s)$. Then given a large enough λ_6, (8) is the same as:

$$
\min_{Z,W,H,L,S,E} \| Z \|_* + \lambda_2 (Diag(L1) - L, W) + \lambda_4 \| S - HH^T \|_F^2 \\
+ \lambda_3 Tr(H^T LH) + \lambda_5 \| S - Z \|_F^2 + \lambda_6 \sum_{i=1}^{k} \theta_i(L_s) + \lambda_1 \| E \|_{2,1}, \\
s.t., \ X = XZ + E, L = (|Z| + |Z^T|)/2, Diag(L) = 0, L \geq 0, \\
H^T H = I, S1 = 1, S \geq 0, \\
0 \leq W \leq I, Tr(W) = k, \\
. \tag{9}
$$

where $\theta_i(L_s) \geq 0$ since L_s is positive semidefinite matrix. Thus, when λ_6 is large enough, $\sum_{i=1}^{k} \theta_i(L_s) = 0$ to satisfy the constraint $rank(L_s) = n - k$. According to Ky Fan's theorem, we know that $\sum_{i=1}^{k} \theta_i(L_s)$ is equivalent to minimizing $Tr(F^T L_s F)$ subject to $F^T F = I$, so we get:

$$
\min_{Z,W,H,L,S,E} \| Z \|_* + \lambda_2 (Diag(L1) - L, W) + \lambda_4 \| S - HH^T \|_F^2 \\
+ \lambda_3 Tr(H^T LH) + \lambda_5 \| S - Z \|_F^2 + \lambda_6 Tr(F^T L_s F) + \lambda_1 \| E \|_{2,1}, \\
s.t., \ X = XZ + E, L = (|Z| + |Z^T|)/2, Diag(L) = 0, L \geq 0, \\
H^T H = I, S1 = 1, S \geq 0, F^T F = I, \\
0 \leq W \leq I, Tr(W) = k, \\
. \tag{10}
$$

where F is a low dimensional embedding matrix the same as H. Except that, S has an inherent k-block diagonal structure already by approximating HH^T. As such, λ_6 doesn't need tuning like the other parameters [27]. It may only turn low or high when the number of connected components is larger or smaller than k.

Moreover, we take a clue from [49] by introducing an intermediate term J to relax (10) further as follows:

$$
\min_{Z,W,H,L,S,E} \| J \|_* + \lambda_2 (Diag(L1) - L, W) + \lambda_4 \| S - HH^T \|_F^2 \\
+ \lambda_3 Tr(H^T LH) + \lambda_5 \| S - Z \|_F^2 + \lambda_6 Tr(F^T L_s F) + \lambda_1 \| E \|_{2,1}, \\
s.t., \ X = XZ + E, L = (|Z| + |Z^T|)/2, Diag(L) = 0, L \geq 0, \\
J = Z, H^T H = I, S1 = 1, S \geq 0, F^T F = I, \\
0 \leq W \leq I, Tr(W) = k. \\
. \tag{11}
$$

The Lagrangian function of (11) is given below:

$$
\mathcal{L} = \| J \|_* + \lambda_2 (Diag(L1) - L, W) + \lambda_4 \| S - HH^T \|_F^2 \\
+ \lambda_3 Tr(H^T LH) + \lambda_5 \| S - Z \|_F^2 + \lambda_6 Tr(F^T L_s F) + \lambda_1 \| E \|_{2,1} \\
+ Tr[M_1^T(X - XZ - E)] + Tr[M_2^T(J - Z)] \\
+ \frac{\mu}{2}(\| X - XZ - E \|_F^2 + \| J - Z \|_F^2), \tag{12}
$$

M_1, M_2 are Lagrange multipliers. Now, we divide (12) into several subproblems by denoting each variable as a subproblem. By doing so, we remove the terms not connected to each subproblem to obtain an optimal solution by fixing others as follows:

J subproblem:

$$
\min_J \frac{1}{\mu} \| J \|_* + \frac{1}{2} \| J - (Z + \frac{M_2}{\mu}) \|_F^2. \tag{13}
$$

Then, denoting the singular value thresholding [3] of $S_\mu[M]$ as $U \mathcal{H}_\mu[\Sigma] V^T$ the optimal solution of (13) is given as:

$$
J = S_{1/\mu}[Z + \frac{M_2}{\mu}]. \\
\mathcal{H}_\mu[\Sigma] = \min(0, \Sigma + \mu) + \max(0, \Sigma - \mu). \tag{14}
$$

W subproblem:

$$
\min_w \lambda_2 (Diag(L1) - L, W), \\
s.t., \ 0 \leq W \leq I, Tr(W) = k. \tag{15}
$$

Similar to [40] we can obtain a close form solution for W as follows:

$$
W = UU^T, \tag{16}
$$

where $U \in R^{n \times k}$ holds the k connected components corresponding to the k eigen vectors associated with k smallest eigen values of L_L.

H subproblem:

$$
\min_H \lambda_4 \| S - HH^T \|_F^2 + \lambda_3 Tr(H^T LH), \\
s.t., \ H^T H = I. \tag{17}
$$

Since (17) is unique for different i, we rewrite it as:

$$
\min_H Tr(H^T(L + 2\lambda_4 S)H), \\
s.t., \ H^T H = I. \tag{18}
$$

Therefore, the optimal H in (18) is derived by the k eigenvectors of the corresponding topmost k highest eigenvalues of $(L + 2\lambda_4 S)$.

Z subproblem:

$$
\min_Z \lambda_5 \| S - Z \|_F^2 + Tr[M_1^T(X - XZ - E)] + Tr[M_2^T(Z - J)] \\
+ \frac{\mu}{2}(\| X - XZ - E \|_F^2 + \| J - Z \|_F^2). \tag{19}
$$

By setting the derivative $\frac{\partial}{\partial Z} = 0$, we obtain:

$$
Z = [(\frac{2}{\mu} + 1)I + X^T X]^{-1}((X^T X - X' * E + J) \\
+ (2\lambda_5 S + X^T M_1 - M_2)/\mu). \tag{20}
$$

S subproblem:

$$\min_S \lambda_4 \parallel S - HH^T \parallel_F^2 + \lambda_5 \parallel S - Z \parallel_F^2 + \lambda_6 Tr(F^T L_s F),$$
$$s.t., \ S1 = 1, S \geq 0, F^T F = I.$$
$$(21)$$

Equation (21) can be rewriting it as:

$$\min_S \parallel S - (\lambda_4 HH^T + \lambda_5 Z) \parallel_F^2 + \lambda_6 Tr(F^T L_s F),$$
$$\min_S \parallel S \parallel_F^2 - 2\langle \lambda_4 HH^T + \lambda_5 Z, S \rangle + \lambda_6 Tr(F^T L_s F),$$
$$s.t., \ S1 = 1, S \geq 0, F^T F = I.$$
$$(22)$$

Then, denoting $2(\lambda_4 HH^T + \lambda_5 Z)$ by G, we obtain the following:

$$\min_S \parallel S \parallel_F^2 - Tr(GS^T) + \lambda_6 Tr(F^T L_s F),$$
$$s.t., \ S1 = 1, S \geq 0, F^T F = I.$$
$$(23)$$

Same as (4), the optimal solution of F in (23) contains eigenvectors of the corresponding k smallest eigenvalue of L_S. Upon that, we rewrite (23) as:

$$\min_{S_j} (\lambda_6 \parallel f_i - f_j \parallel_2^2 - g_{ij}) s_{ij} + s_j^T s_j,$$
$$s.t., \ s_j \geq 0, 1^T s_j = 1.$$
$$(24)$$

Then, denoting $\lambda_6 \parallel f_i - f_j \parallel_2^2 - g_{ij}$ as q_{ij}, (24) is simplified as:

$$\min_{S_j} \frac{1}{2} \parallel s_j + \frac{q_j}{2} \parallel_2^2,$$
$$s.t., \ s_j \geq 0, 1^T s_j = 1.$$
$$(25)$$

Similar to Eq. 30 in [27], we obtain s_j as follows:

$$s_j = \left(-\frac{q_j}{2} + \eta 1\right)_+.$$
$$(26)$$

L subproblem:

$$\min_L \lambda_3 Tr(H^T LH) + \lambda_2 (Diag(L1) - L, W),$$
$$s.t., \ L = (|Z| + |Z^T|)/2, Diag(L) = 0, L \geq 0.$$
$$(27)$$

The optimal solution L is:

$$L = \frac{1}{2} \parallel \lambda_3 (HH^T) + \lambda_2 [Diag(W) * 1^T - W] \parallel^2 \quad (28)$$

Furthermore, by denoting $A = \lambda_3 (HH^T) + \lambda_2 [Diag(W) * 1^T - W]$, $\hat{A} = A - diag(diag(A))$, we rewrite (28) as follows:

$$L = [(\hat{A} + \hat{A}^T)/2]_+. \quad (29)$$

E subproblem:

$$E = \min_E \frac{\lambda_1}{\mu} \parallel E \parallel_{2,1} + \frac{1}{2} \parallel E - \left(X - XZ + \frac{Y_1}{\mu}\right) \parallel_F^2.$$
$$(30)$$

Algorithm 1 outlines the complete solution of our method.

Algorithm 1 Proposed algorithm of our method.

1: Input: Data X, clusters size k
2: Initialize $M_1 = 0$, $M_2 = 0$, $\varepsilon = 10^{-4}$, Z and S matrices are based on k nearest neighbor graph; $\rho = 1.01$; $\mu = 0.01$; $\mu_{\max} = 10^8$
3: **while** not converged **do**
4:      Update J through (14).
5:      Update W through (16).
6:      Update H through (18).
7:      Update Z through (20).
8:      Update S through (26).
9:      Update L through (29).
10:      Update E through (30).
11:      Update the multipliers.
         $M_1 = M_1 + \mu(X - XZ - E)$
         $M_2 = M_2 + \mu(S - Z)$
12:      Update μ by $\mu = \min(\rho\mu, \mu_{max})$.
13:      Check the convergence conditions:
         $\parallel X - XZ - E \parallel_\infty < \varepsilon$ and $\parallel S - Z \parallel_\infty < \varepsilon$.
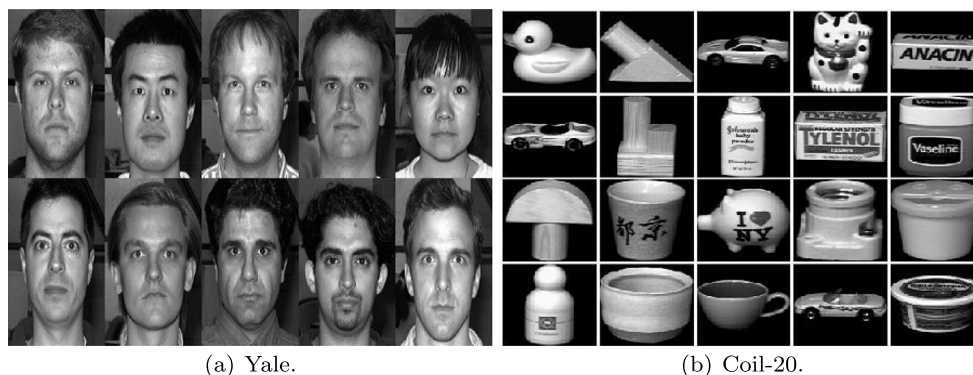14: **end while**
15: Output: S.

3.3 Complexity analysis

Our method's complexity depends on three key components: matrix multiplication required to solve Z and E subproblems, matrix inversion required to solve Z subproblem, and eigendecomposition required to solve W, H, and S subproblems. For each matrix multiplication, the inverse of a matrix, and eigendecomposition, the time complexity is $O(n^3)$. But notice that the Z subproblem needs more than one multiplication. So, it takes $(k + 1)O(n^3)$, which may be problematic in large-scale datasets. In that case, one can perform inverse and multiplication operations using a graphics processing unit (GPU) device or apply large-scale techniques such as Accelerated LRR [11] to speed up our method. Besides, (25) takes $O(n)$ since it requires Euclidean projection onto simplex space. Overall, each iteration's time complexity is $O(3n^3 + (k + 1)n^3 + n)$.

4 Experiments

In this section, we demonstrate the effectiveness of our method. To begin with experiment settings, we describe six benchmark datasets, nine state-of-the-art compared methods, and six standard metrics utilized to evaluate our proposed method. We then present the experimental results with analysis in Section 4.2.

(a) Yale.   (b) Coil-20.

4.1 Experiment settings

4.1.1 Data set

To assess our method's performance, we conduct widespread experiments on six benchmark datasets, namely UCI Digits, USPS, ORL, Yale, COIL-20, and Caltech101-07 datasets.

- UCI digits[1]: dataset has 2000 data samples of handwritten digits from 0 to 9. Each data sample has 240 features or pixels that describe each digit.
- USPS[2]: dataset has 1854 data samples of handwritten digits from 0 to 9, where 256 features describe each digit.
- ORL[3]: dataset comprises 400 images of 40 different subjects. Each subject contains 10 distinct images taken under various conditions such as lights, facial expressions (smiling or not smiling), and facial details (with glasses or without glasses). Besides, each image has 1024 features to describe it.
- Yale[4]: dataset contains 165 gray-scale images of 15 individuals with 6750 Gabor features for each image. I.e., every individual contributes 11 images taken with different facial settings such as glasses, without glasses, happy mood, sad mood, and sleepy mood.
- COIL-20[5]: dataset contains 20 classes with a total of 1440 images. For each class, there are 72 images, which corresponds to 1024 features per image.
- Caltech101-07[6]: dataset covers 101 classes of images. However, we selected the commonly used seven

classes and obtained 1474 images with 1984 HOG features.

Table 1 gives a summary of the datasets, while Fig. 1 provides a pictorial view of Yale and COIL-20 datasets.

4.1.2 Compared methods

We compare our method with nine state-of-the-art subspace clustering methods: LRR [19], SSC [10], CASS [22], S^3C [17], ELM-CLR [21], SinNLRR [49], BDR [23], IBDLR [40], and FGNSC [43]. Most of these methods were selected for comparison because they use a similar approach as our proposed method. For example, IBDLR, SinNLRR, CASS, and our proposed method uses LRR's nuclear norm regularization approach to find a low dimensional representation of data.

Therefore, each method's parameters are tuned following the corresponding literature to achieve the best clustering performance. Accordingly, we give a brief description of the different methods below.

- LRR [19]: This method uses the nuclear norm to pursue a low-rank matrix of data from which it obtains a similarity matrix.
- SSC [10]: This method finds a sparse representation of data independently using the L_1 norm.
- CASS [22]: This method encodes the grouping effect of LRR and the sparsity nature of SSC.
- S^3C [17]: This method extends SSC by learning a similarity matrix and data segmentation simultaneously.
- ELM-CLR [21]: This method obtains a similarity matrix adaptively through dual data representations. One representation is based on the actual data, while another relies on the data's affine space.
- SinNLRR [49]: This method uses nonnegative constraints to ensure that similar data samples are connected while unrelated data samples are not connected.
- BDR [23]: This method employs the k-block diagonal regularizer to encourage a block diagonal similarity matrix.

Table 1 Summary of the datasets

Dataset	Samples	Features	Clusters
UCI-digits	2000	240	10
USPS	1854	256	10
ORL	400	1024	40
Yale	165	6750	15
COIL-20	1440	1024	20
Caltech101-07	1474	1984	7

- IBDLR [40]: This method also uses the k-block diagonal regularizer by BDR to find a proper similarity matrix. Besides, it adapts the similarity on dual data representations similar to ELM-CLR.
- FGNSC [43]: This method makes use of post-processing technique to optimize both sparsity and connectivity of self-representation simultaneously to find good neighbors.

Note that all the compared methods except ELM-CLR apply a spectral clustering algorithm on the learned similarity matrix to obtain a clustering structure. Meanwhile, ELM-CLR can obtain a cluster structure explicitly with a rank constraint on the similarity matrix's Laplacian matrix, the same as our proposed method. Furthermore, we adopt the best value from [0.001, 0.01, 0.1, ..., 1000] using a grid search strategy for our tunable parameters. Hence, to ensure

no generalization loss, we perform each experiment ten times and report the mean and the standard deviation of the clustering performance.

4.1.3 Evaluation metrics

Six conventional metrics, namely Accuracy (ACC), Normalized Mutual Information (NMI), F-score, Recall, Precision, and Adjusted Rand index (AR), are used to evaluate the clustering performance of each method. These metrics capture different aspects whose values are directly proportional to the clustering performance. For example, ACC measures the percentage of correctly clustered data samples in the learned clustering structure compared with the ground truth labels. NMI refers to the amount of information that one can extract from shared random variables. F-score is a measure of an experiment's accuracy calculated from the experiment's Precision and Recall values. AR computes a similarity measure between two clusterings by considering all data samples' pairs and counting pairs assigned in the same or different clusters of the learned clustering structure and ground truth labels.

4.2 Result and analysis

We present the clustering performance of different algorithms on six benchmark datasets in Tables 1, 2, 3 and 4.

Table 2 Clustering performance of different algorithms on handwritten digits

Dataset	Methods	ACC	NMI	F-score	Recall	Precision	AR
UCI Digits	SSC	66.40±0.00	68.37±0.01	58.82±0.03	71.37±0.02	50.02±0.03	53.36±0.03
	LRR	55.10±0.00	51.48±0.00	52.80±0.00	68.67±0.00	41.55±0.00	40.80±0.00
	CASS	72.14±3.95	66.70±1.45	63.20±1.95	66.22±1.73	60.46±2.49	58.92±2.21
	S^3C	66.15±0.00	67.67±0.00	58.24±0.00	70.92±0.00	49.41±0.00	52.69±0.00
	ELM-CLR	87.92±6.02	82.87±3.06	81.03±4.89	82.15±3.22	80.02±6.41	78.89±5.52
	SinNLRR	87.00±0.00	79.40±0.00	77.80±0.00	78.22±0.00	77.38±0.00	75.33±0.00
	BDR	81.73±0.02	83.99±0.04	80.93±0.10	86.38±0.04	76.12±0.09	78.67±0.11
	IBDLR	77.85±0.05	76.59±0.14	73.63±0.12	76.67±0.12	70.82±0.12	70.58±0.13
	FGNSC	89.51±0.97	93.37±1.08	86.24±1.68	84.46±2.06	89.00±1.81	81.57±1.74
	Ours	**97.25±0.00**	**93.86±0.00**	**94.52±0.00**	**94.64±0.00**	**94.40±0.00**	**93.91±0.00**
USPS	SSC	53.16±1.18	56.28±0.21	47.28±1.22	60.25±2.19	38.91±0.72	39.40±1.32
	LRR	66.61±0.00	64.08±0.00	52.93±0.00	60.56±0.00	47.01±0.00	46.47±0.00
	CASS	53.88±0.04	64.19±0.08	54.57±0.05	58.99±0.08	50.77±0.03	48.66±0.06
	S^3C	53.46±0.15	56.23±0.05	47.68±0.08	60.96±0.09	39.15±0.07	39.83±0.09
	ELM-CLR	65.20±3.38	69.78±1.73	62.68±4.64	67.31±1.50	58.81±4.64	57.84±3.35
	SinNLRR	71.95±0.00	69.77±0.00	66.21±0.00	66.08±0.00	66.34±0.00	62.16±0.00
	BDR	69.45±003	64.96±0.00	62.81±0.00	61.14±0.00	64.59±0.00	58.49±0.00
	IBDLR	71.31±0.16	67.07±0.17	64.33±0.18	63.94±0.19	64.73±0.17	60.08±0.20
	FGNSC	73.53±0.46	68.80±0.36	68.50±0.22	67.61±0.24	69.41±0.13	63.60±0.09
	Ours	**81.39±0.00**	**80.14±0.00**	**77.65±0.00**	**80.66±0.00**	**74.86±0.00**	**74.86±0.00**

Table 3 Clustering performance of different algorithms on face images

Dataset	Methods	ACC	NMI	F-score	Recall	Precision	AR
ORL	SSC	60.15±1.72	78.82±1.15	48.25±2.63	57.47±2.06	41.67±3.23	46.86±2.73
	LRR	48.00±0.00	58.68±0.00	33.99±0.00	38.28±0.00	30.89±0.00	31.14±0.00
	CASS	61.25±2.60	77.39±1.39	48.70±2.78	52.06±2.77	45.75±2.83	47.43±2.85
	S^3C	62.30±1.75	79.63±0.75	50.71±1.44	59.36±1.87	44.28±1.50	49.40±1.48
	ELM-CLR	65.35±0.16	79.60±0.19	44.61±0.39	**64.94±0.23**	25.45±0.33	35.46±0.41
	SinNLRR	53.65±1.91	72.94±0.80	40.01±1.69	42.37±1.67	37.90±1.76	38.54±1.73
	BDR	68.30±0.18	82.31±1.32	58.07±0.69	61.83±1.21	54.75±0.82	57.05±1.06
	IBDLR	65.83±0.66	79.82±0.87	52.40±1.92	55.37±1.94	49.74±2.05	51.24±1.98
	FGNSC	60.52±2.02	78.55±0.90	49.16±2.22	55.28±1.57	44.36±3.17	47.85±2.32
	Ours	**69.25±0.00**	**84.25±0.00**	**59.16±0.00**	62.17±0.00	**56.43±0.00**	**58.17±0.00**
Yale	SSC	65.58±0.80	69.12±0.91	49.68±1.45	51.84±1.47	47.69±1.46	46.27±1.55
	LRR	31.52±0.00	39.95±0.00	18.98±0.00	28.48±0.00	14.23±0.00	11.81±0.00
	CASS	50.06±1.68	53.65±1.43	33.37±1.82	34.24±1.83	32.54±1.84	28.92±1.95
	S^3C	66.67±2.07	70.50±1.33	50.51±3.06	55.03±1.60	46.66±4.62	47.00±3.50
	ELM-CLR	66.03±1.02	67.23±1.22	48.87±1.83	53.88±1.21	44.75±2.43	45.21±2.03
	SinNLRR	44.24±1.33	51.47±1.16	29.16±1.25	30.33±1.08	28.09±1.25	24.37±1.26
	BDR	65.36±1.31	63.15±0.99	48.41±1.67	50.36±1.63	46.63±1.70	44.20±1.80
	IBDLR	69.90±0.57	69.83±0.57	54.02±0.64	54.99±0.70	53.08±0.58	50.98±0.68
	FGNSC	63.82±3.71	64.90±2.19	47.92±2.68	49.31±2.75	46.61±2.63	44.44±2.85
	Ours	**72.12±0.00**	**73.61±0.00**	**58.38±0.00**	**62.91±0.00**	54.46±0.00	**55.47±0.00**

In each table, the values in bold symbolize the best performance.

Table 2 shows the clustering performance of different methods on UCI Digits and USPS handwritten datasets. It is easy to see that our method achieves the best clustering performance in all evaluation metrics. Also, FGNSC achieves better performance than others, followed closely by ELM-CLR, which may be due to the adaptive nature of learning the similarity matrix. On the contrary, the S^3C method has the worst performance than SinNLRR and CASS, which obtains a similarity matrix uncooperatively. Arguably, SinNLRR may benefit from nonnegative

Table 4 Clustering performance of different algorithms on object images

Dataset	Methods	ACC	NMI	F-score	Recall	Precision	AR
COIL-20	SSC	74.67±0.03	87.55±0.03	71.53±0.05	85.83±0.08	61.31±0.06	69.79±0.06
	LRR	65.21±0.00	74.41±0.00	52.06±0.00	68.80±0.00	41.87±0.00	48.93±0.00
	CASS	60.17±1.63	70.66±0.73	53.05±1.48	57.37±1.83	49.34±1.30	50.42±1.55
	S^3C	75.47±1.68	87.89±0.64	72.54±2.01	85.72±0.53	62.96±3.39	70.88±2.16
	ELM-CLR	85.67±0.16	93.00±0.23	83.23±0.39	92.38±0.99	75.10±0.41	82.27±0.42
	SinNLRR	68.75±0.00	79.81±0.02	63.55±0.03	70.67±0.02	57.73±0.04	61.45±0.03
	BDR	85.75±0.45	93.06±0.38	83.14±0.27	91.09±0.08	75.44±0.42	82.31±0.29
	IBDLR	84.58±0.59	91.84±0.56	80.86±0.69	88.96±1.30	73.10±0.13	76.56±0.70
	FGNSC	85.91±1.06	93.71±1.51	83.57±1.40	94.31±1.20	73.54±1.11	82.87±1.02
	Ours	**87.22±0.00**	**94.07±0.00**	**84.97±0.00**	**96.86±0.00**	**75.68±0.00**	**84.09±0.00**
Caltech101-07	SSC	60.70±0.03	45.44±0.05	56.22±0.05	46.72±0.02	70.56±0.13	36.81±0.09
	LRR	77.00±0.00	60.94±0.00	76.84±0.00	70.12±0.00	84.99±0.00	59.46±0.00
	CASS	66.42±0.00	47.88±0.00	61.84±0.00	55.54±0.00	69.76±0.00	42.03±0.00
	S^3C	61.13±0.32	49.31±0.72	58.31±0.38	46.35±0.47	78.69±3.03	41.60±1.18
	ELM-CLR	62.75±5.99	59.79±2.95	71.19±4.96	58.67±5.19	**92.11±3.78**	59.49±5.87
	SinNLRR	44.30±0.00	43.22±0.00	49.55±0.00	36.93±0.00	75.28±0.00	32.40±0.00
	BDR	71.06±0.06	55.18±0.12	75.28±0.23	67.13±0.04	83.79±0.03	53.46±0.04
	IBDLR	61.94±0.03	47.52±0.00	59.71±0.01	44.01±0.01	77.20±0.01	40.68±0.05
	FGNSC	64.76±0.27	49.23±0.30	60.38±0.28	56.50±0.20	63.82±0.43	42.86±0.33
	Ours	**82.77±0.00**	**61.46±0.00**	**79.13±0.00**	**77.73±0.00**	88.35±0.00	**60.39±0.00**

(a) S of SSC.

(b) S of LRR.

(c) S of CASS.

(d) S of S^3C.

(e) S of ELM-CLR.

(f) S of SinNLRR.

(g) S of BDR.

(h) S of IBDLR.

(i) S of FGNSC.
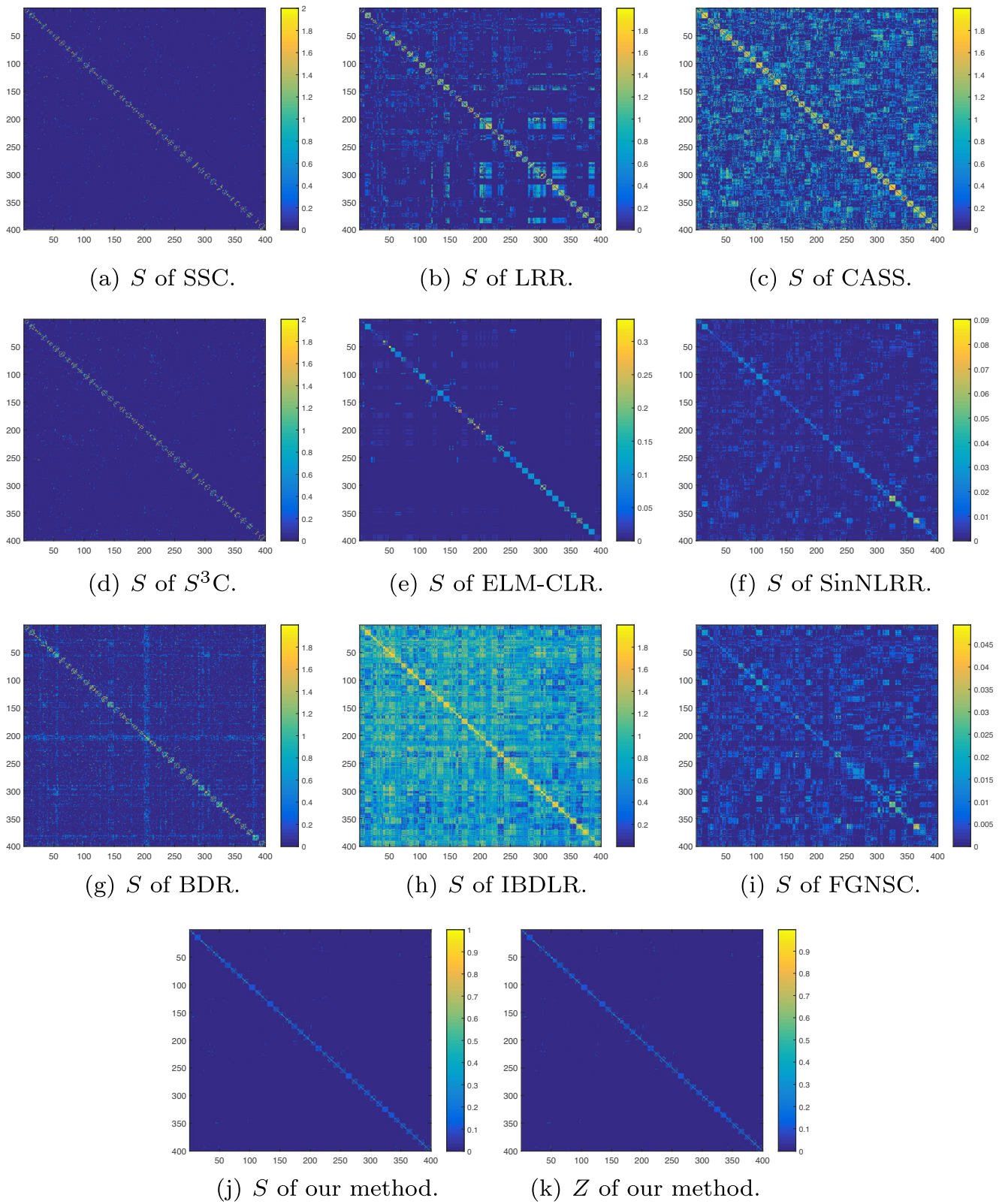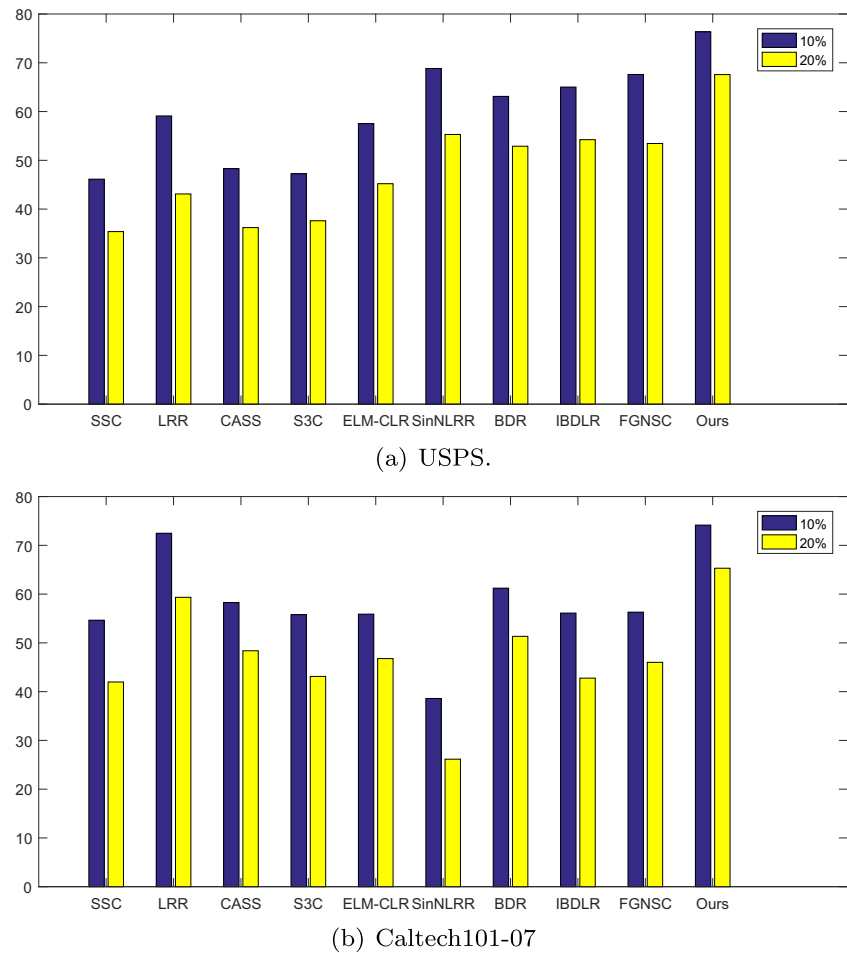
(j) S of our method.

(k) Z of our method.

**Fig. 2** Visualization of similarity matrices of different methods

Fig. 3 Performance comparison of different methods w.r.t. ACC on USPS and Caltech-07 datasets with 10% and 20% levels of noise


(a) USPS.

(b) Caltech101-07

constraints imposed on the low-rank matrix, which seems useful for relatively low dimensional UCI Digits and USPS datasets. Moreover, CASS combines the grouping effect of LRR and SSC's sparsity nature to enhance its performance. S^3C, on the other hand, appears limited by the L_1-norm low dimensional matrix, which does not capture the data's global structure.

Notwithstanding, UCI Digits' results clearly show that our method outperforms FGNSC by over 9.44%, 0.49%, 8.28%, 10.18%, 5.4%, and 12.34% in ACC, NMI, F-score, Recall, Precision, and AR evaluation metrics, respectively. Furthermore, FGNSC, SinNLRR, and IBDLR have comparable performances on the USPS dataset, better than the rest methods. Still, our method consistently outperforms these methods in all evaluation metrics on this dataset. Table 3 tabulates the clustering performance on ORL and Yale face image datasets. Thus, our method maintains the best performance on Yale dataset with 72.12%, 73.61%, 58.38%, 62.91%, 54.46%, and 55.47%, in ACC, NMI, F-score, Recall, Precision, and AR, respectively. Similarly, on ORL dataset, our method has

better performance in all metrics except Recall. This performance is only slightly higher than that of BDR by 0.95%, 1.94%, 1.09%, 0.34%, 1.68%, 1.12%, in ACC, NMI, F-score, Recall, Precision, and AR, respectively, perhaps because BDR also applies the k block diagonal regularizer. Additionally, ELM-CLR outperforms S^3C by 2.62% and 13.58% just in ACC and Recall evaluation metrics, respectively. In contrast, S^3C outperforms ELM-CLR in all evaluation metrics on the Yale dataset. Nevertheless, the standard deviation indicates that ELM-CLR is more stable than S^3C on face image clustering. Moreover, ELM-CLR and S^3C have better performances than SSC, LRR, CASS, SinNLRR, and FGNSC in most evaluation metrics to demonstrate adaptive similarity matrix's effectiveness. Furthermore, our method outperforms ELM-CLR in ACC by over 3% on ORL datasets even as the nonlinear data structure of ORL favors ELM-CLR since it learns the similarity matrix on the actual data and the affine space of the data simultaneously. That being the case, Fig. 2 provides a visual comparison of similarity matrices of the different methods to demonstrate more superiority
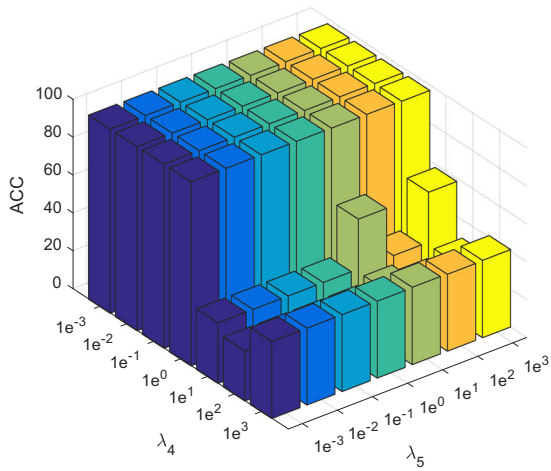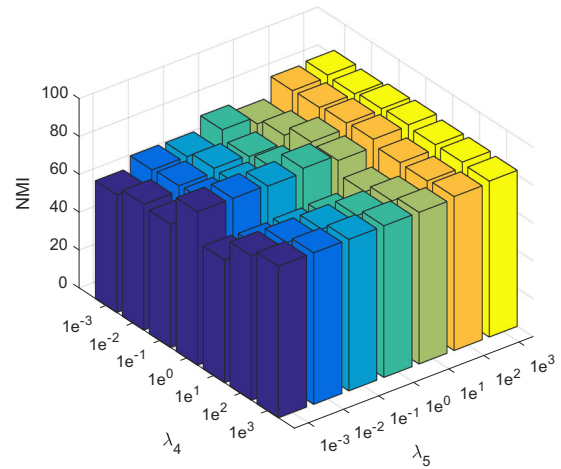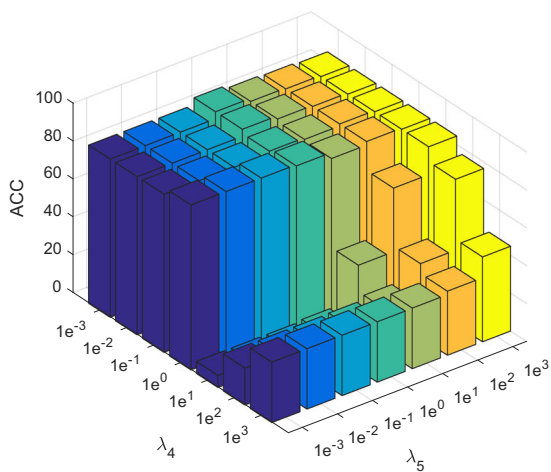
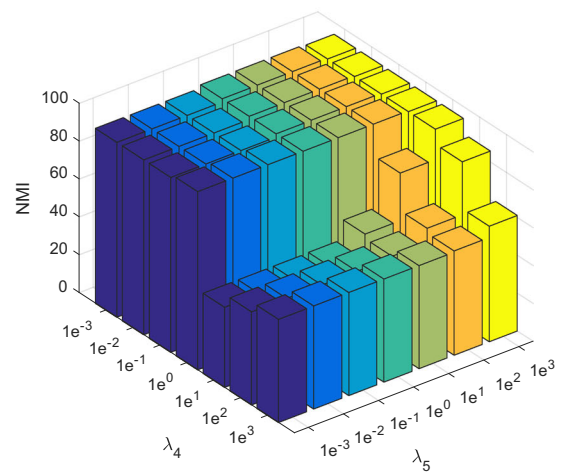(a) ACC UCI digits
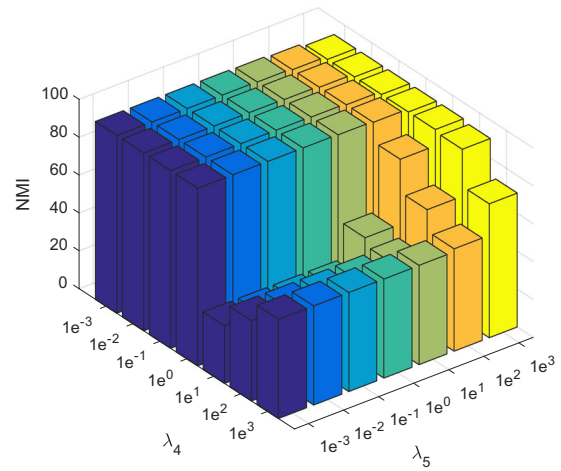
(b) NMI UCI digits
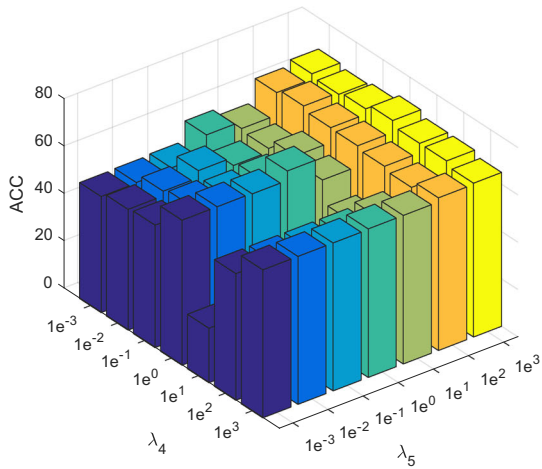
(c) ACC ORL

(d) NMI ORL

(e) ACC COIL-20

(f) NMI COIL-20

Fig. 4 Visualization of the clustering performance of our method w.r.t. ACC and NMI evaluation metrics by fixing other parameters and varying-parameters λ_4 and λ_5 on UCI Digits dataset, ORL dataset, and COIL-20 dataset

of our proposed method on ORL dataset. Thus, we can see that most methods, including CASS, IBDLR, and FGNSC, obtained similarity matrices with much density, which means the clusters in these matrices are not independent.

On the other hand, Fig. 2 shows that the similarity matrix obtained by our proposed method has exactly k connected components, which means that our method can capture an accurate block diagonal property. We can also observe that the similarity matrix S and the low-rank matrix Z of our proposed method are very close, confirming our coupled learning effectiveness.

Table 4 compares our method with six state-of-the-art methods using COIL-20 and Caltech101-07 object datasets. On COIL-20 dataset, our method outperforms FGNSC, BDR, and ELM-CLR only slightly in NMI, F-score, and Precision, and AR, respectively. Whereas it performs better than its closest competitor, LRR, by over 5%, 2%, 7%, and 3% in ACC, F-score, Recall, and Precision, respectively, on Caltech 101-07 dataset. We note that ELM-CLR achieves better performance than our method by 3.76% in Precision, which perhaps is due to significant class imbalance in this dataset. However, ELM-CLR fails to maintain the same performance in other metrics. For example, our method outperforms ELM-CLR by over 20% and 19% in ACC and Recall, respectively. Besides, excluding SinNLRR and SSC, it is easily noticeable that the other methods also perform better than ELM-CLR and S^3C methods.

Additionally, Fig. 3 shows the robustness of our method against noise on USPS and Caltech-07 datasets. We use random noise following uniform distribution between -0.5 and 0.5, similar to ELM-CLR [21], to perform this experiment. In particular, we adopt two experimental settings in each dataset. Firstly, 10% of the original pixels of sample images are replaced with random values in the range above. Then, 20% in the second case. Thus, one may observe that all methods have degrading performance, increasing the noise level. However, our proposed method is generally more robust against noise when compared to the other methods.

4.2.1 Parameter analysis

Figure 4 shows our proposed method's clustering performance by varying λ_4 and λ_5 while fixing other on UCI digits, ORL, and COIL-20 datasets. Note that λ_6 does not need tuning like the other parameters because a similar approach in [21, 27] is followed to obtain its best value in a heuristic way to accelerate the process. Initially, we set $\lambda_6 = 1$ and automatically multiply it by two or divide it by two in each iteration when the number of connected components is smaller or larger than k. It is easy to see from Fig. 4 that our proposed method is relatively stable on all datasets by varying λ_4 and λ_5 in a range of [0.001,0.01,0.1,1]. Therefore, our proposed method can ensure good clustering performance by finding a suitable combination of both parameters.

4.2.2 Computational runtime analysis

Here we investigate the computational time of different methods on all six datasets. Specifically, we ran each method using MATLAB 2016b installed on Windows 10 CORE i5 computer system to perform this experiment. Accordingly, Table 5 displays the computational time obtained by different methods. Notice that the computation time is excessively high for most methods, which performs k-means spectral post-processing of the low dimensional embedding matrix, such as CASS, S^3C, and SinNLRR. However, it is not surprising to see that CASS has a much higher computational time than LRR and SSC since it combines both approaches to learn the similarity matrix. Besides, IBDLR has three times more computation time than the similar BDR method in most datasets— the reason

Table 5 The computational runtime (in seconds) of different methods on six benchmark datasets

Dataset	UCI digits	USPS	ORL	Yale	COIL-20	Caltech101-07
SSC	131.71	115.87	35.68	360.04	170.07	381.12
LRR	50.90	75.45	19.68	15.33	638.21	851.61
CASS	$2.24e^{+03}$	$2.05e^{+03}$	282.71	32.28	$1.27e^{+04}$	$2.75e^{+04}$
S^3C	446.72	$1.24e^{+03}$	75.73	908.52	233.12	$1.02e^{+03}$
ELM-CLR	45.15	44.09	1.58	0.74	13.98	23.77
SinNLRR	$2.24e^{+04}$	$6.26e^{+04}$	3.89	4.81	62.69	$3.30e^{+03}$
BDR	125.23	44.65	8.76	1.12	238.83	199.26
IBDLR	511.80	425.11	27.26	6.61	401.64	338.39
FGNSC	87.27	83.67	1.99	0.42	130.94	206.51
Ours	61.65	56.31	1.75	1.09	26.45	34.26

is not far-fetched because IBDLR learns a similarity matrix on dual data representations. We can also observe that our method has comparable computational time with ELM-CLR since both methods obtain the clustering structure explicitly without k-means spectral post-processing step. Thus, both methods converge very quickly, making their computational time much lesser than others. Nonetheless, when one considers that our method has six subproblems to update in each iteration as against a few by ELM-CLR, it may then be seen that our proposed method has a relatively good computational time in all six datasets

4.2.3 Convergence analysis

According to Liu et al. [19] and Luo et al. [24], the convergence of the inexact Augmented Lagrange Multiplier (ALM) method with more than three sub-problems is still not easy to theoretically prove. As a result, the relative error of the term $\| S - Z \|_F^2$ is computed to illustrate our method's convergence behavior using Figs. 5 and 6. As expected, one may see from Fig. 6 that the difference between the clustering structure and the low-rank representation reduces by each iteration step before convergence because our

Fig. 5 Visualization of the clustering performance of our method at different stages of iterations

(a) UCI digits

(b) USPS

(c) ORL
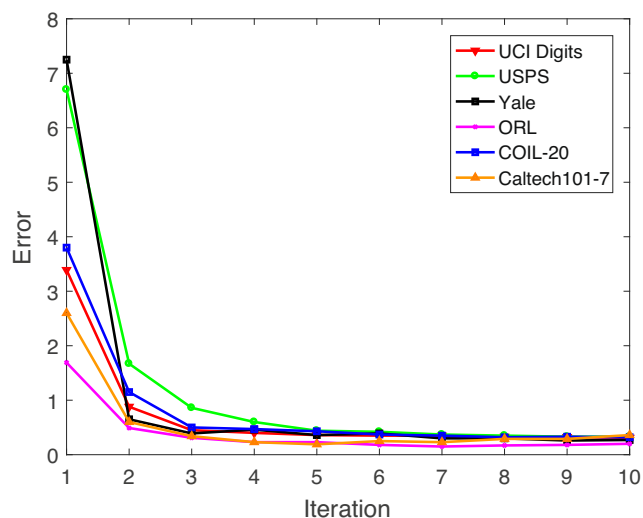
(d) Yale

(e) COIL-20

(f) Caltech101-07
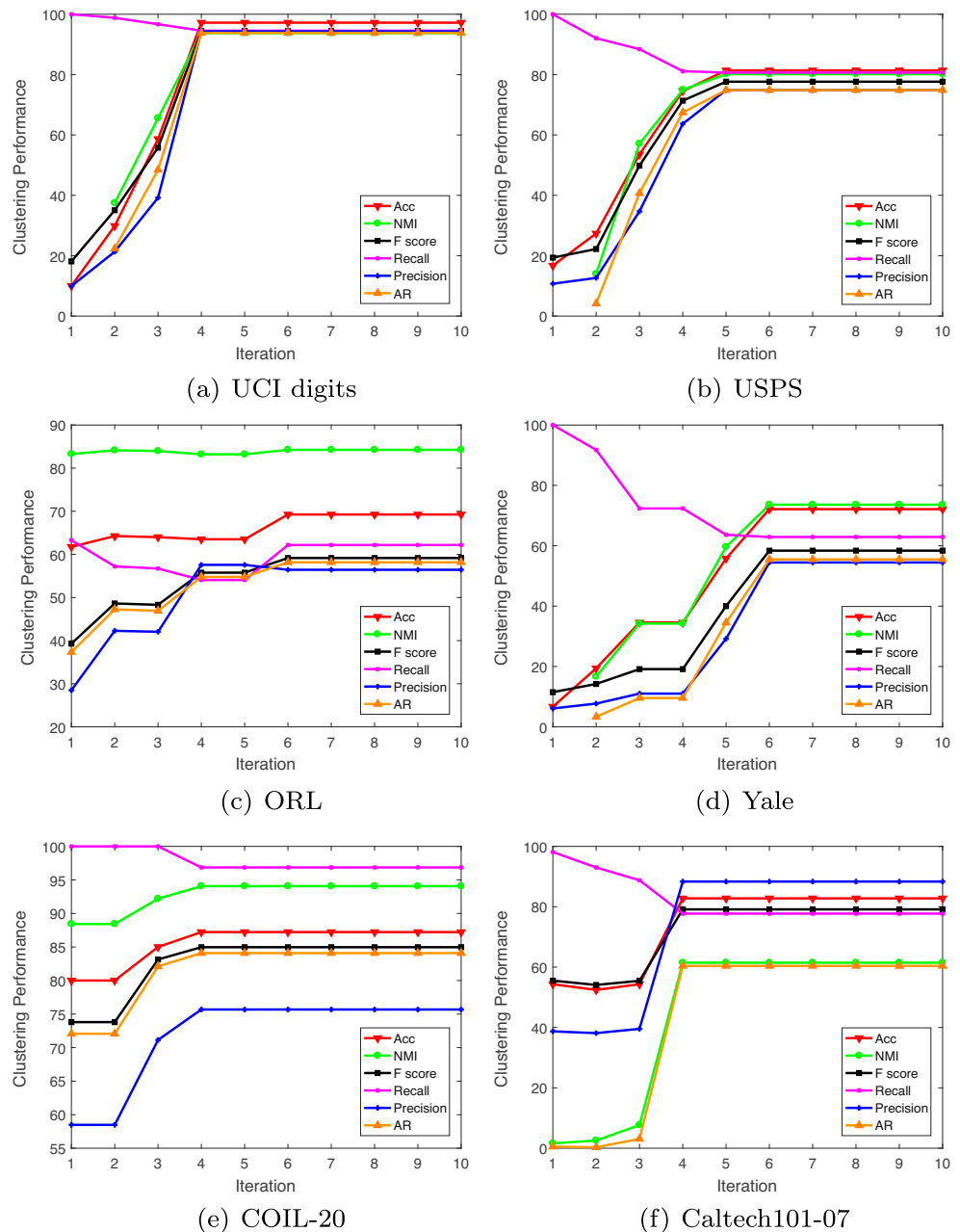
Fig. 6 Visualization of the convergence behavior of our coupled method ($\| S - Z \|_F^2$) at different stages of iterations

method learns cleaner low-rank representation in successive iteration circles. Therefore, Fig. 5 shows that our method converges within ten iterations on all six benchmark datasets, which means it is very stable.

5 Conclusion

This paper proposed a novel method that guarantees better clustering performance through coupled low-rank representation. To achieve that, we considered the practical case where a given data may have arbitrarily distribution, in which two unrelated samples are likely to reconstruct themselves. Therefore, we learned a manifold recovery structure to correct any incorrectness in the data's low-rank representation. Then we obtained an ideal similarity matrix through this manifold structure differently from existing subspace clustering methods, which construct the similarity matrix by utilizing the low-dimensional representation acquired directly from original data. Furthermore, we ignored k-means spectral post-processing of the low-dimensional embedding matrix by imposing a rank constraint on the similarity matrix's Laplacian matrix to obtain the clustering structure explicitly. Besides, we found that our proposed method can learn a more accurate clustering structure by incorporating the clustering structure into the low-rank representation. Comprehensive experiments on six benchmark datasets show that our method outperforms similar state-of-the-art methods in Accuracy, Normalized Mutual Information, F-score, Recall, Precision, and Adjusted Rand Index evaluation metrics. In future work, we will extend our proposed method to Multi-view learning.

Author Contributions Stanley Ebhohimhen Abhadiomhen: Conceptualization, Data Curation, Methodology, Software, Investigation, Validation, Formal analysis, Visualization, Writing - Original Draft, Writing - Review & Editing

ZhiYang Wang: Software, Validation, Formal analysis.

XiangJun Shen: Conceptualization, Funding acquisition, Project administration, Resources, Supervision, Validation, Review & Editing.

Availability of data and material https://www.kaggle.com/bistaumanga/usps-dataset https://archive.ics.uci.edu/ml/datasets/Optical+Recognition+of+Handwritten+Digits http://cam-orl.co.uk/facedatabase.html http://vision.ucsd.edu/content/yale-face-database https://www.cs.columbia.edu/CAVE/software/softlib/coil-20.php http://www.vision.caltech.edu/archive.html

Code Availability Custom code implemented using MATLAB 2016b installed on Windows 10 CORE i5 computer system

Declarations

Conflicts of interest/Competing interests The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

1. Brbić M, Kopriva I (2018) Multi-view low-rank sparse subspace clustering. Pattern Recogn 73:247–258
2. Brbić M, Kopriva I (2020) ℓ_0-motivated low-rank sparse subspace clustering. IEEE Trans Cybern 50(4):1711–1725. https://doi.org/10.1109/TCYB.2018.2883566
3. Cai JF, Candès EJ, Shen Z (2010) A singular value thresholding algorithm for matrix completion. SIAM J Optim 20(4):1956–1982
4. Cai X, Huang D, Wang CD, Kwoh CK (2020) Spectral clustering by subspace randomization and graph fusion for high-dimensional data. In: Pacific-Asia Conference on Knowledge Discovery and Data Mining. Springer, pp 330–342
5. Chen J, Zhang H, Mao H, Sang Y, Yi Z (2016a) Symmetric low-rank representation for subspace clustering. Neurocomputing 173:1192–1202
6. Chen Y, Zhang L, Yi Z (2016b) A novel low rank representation algorithm for subspace clustering. Int J Pattern Recogn Artif Intell 30(04):1650007
7. Chung FR, Graham FC (1997) Spectral graph theory. 92, American Mathematical Society
8. Cui G, Li X, Dong Y (2018) Subspace clustering guided convex nonnegative matrix factorization. Neurocomputing 292:38–48
9. Deng T, Ye D, Ma R, Fujita H, Xiong L (2020) Low-rank local tangent space embedding for subspace clustering. Inf Sci 508:1–21
10. Elhamifar E, Vidal R (2013) Sparse subspace clustering: algorithm, theory, and applications. IEEE Trans Pattern Anal Mach Intell 35(11):2765–2781
11. Fan J, Tian Z, Zhao M, Chow TW (2018) Accelerated low-rank representation for subspace clustering and semi-supervised classification on large-scale data. Neural Netw 100:39–48
12. Fan K (1950) On a theorem of weyl concerning eigenvalues of linear transformations i. Proc Natl Acad Sci U S A 35(11):652

13. Gruber A, Weiss Y (2004) Multibody factorization with uncertainty and missing data using the em algorithm. In: Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, CVPR 2004, vol 1. IEEE, pp I–I

14. Hinojosa C, Bacca J, Arguello H (2018) Coded aperture design for compressive spectral subspace clustering. IEEE J Sel Top Signal Process 12(6):1589–1600

15. Li A, Qin A, Shang Z, Tang YY (2019) Spectral-spatial sparse subspace clustering based on three-dimensional edge-preserving filtering for hyperspectral image. Int J Pattern Recogn Artif Intell 33(03):1955003

16. Li C, Wang CL, Wang J (2016) Convergence analysis of the augmented lagrange multiplier algorithm for a class of matrix compressive recovery. Appl Math Lett 59:12–17

17. Li CG, You C, Vidal R (2017) Structured sparse subspace clustering: a joint affinity learning and subspace clustering framework. IEEE Trans Image Process 26(6):2988–3001

18. Liu G, Xu H, Yan S (2012) Exact subspace segmentation and outlier detection by low-rank representation. In: Artificial Intelligence and Statistics, pp 703–711

19. Liu G, Lin Z, Yan S, Sun J, Yu Y, Ma Y (2013) Robust recovery of subspace structures by low-rank representation. IEEE Trans Pattern Anal Mach Intell 35(1):171–184

20. Liu M, Wang Y, Sun J, Ji Z (2020) Structured block diagonal representation for subspace clustering. Appl Intell:1–14

21. Liu T, Lekamalage CKL, Huang GB, Lin Z (2018) An adaptive graph learning method based on dual data representations for clustering. Pattern Recogn 77:126–139

22. Lu C, Feng J, Lin Z, Yan S (2013) Correlation adaptive subspace segmentation by trace lasso. In: Proceedings of the IEEE international conference on computer vision, pp 1345–1352

23. Lu C, Feng J, Lin Z, Mei T, Yan S (2019) Subspace clustering by block diagonal representation. IEEE Trans Pattern Anal Mach Intell 41(2):487–501

24. Luo S, Zhang C, Zhang W, Cao X (2018) Consistent and specific multi-view subspace clustering. In: Proceedings of the AAAI Conference on Artificial Intelligence, vol 32

25. Muthu S, Tennakoon R, Rathnayake T, Hoseinnezhad R, Suter D, Bab-Hadiashar A (2020) Motion segmentation of rgb-d sequences: Combining semantic and motion information using statistical inference. IEEE Trans Image Process 29:5557–5570

26. Ng AY, Jordan MI, Weiss Y (2002) On spectral clustering: Analysis and an algorithm. In: Advances in neural information processing systems, pp 849–856

27. Nie F, Wang X, Jordan MI, Huang H (2016) The constrained laplacian rank algorithm for graph-based clustering. In: AAAI. Citeseer, pp 1969–1976

28. Shen J, Li P (2016) Learning structured low-rank representation via matrix factorization. In: Artificial Intelligence and Statistics, pp 500–509

29. Sun W, Ma J, Yang G, Du B, Zhang L (2017) A poisson nonnegative matrix factorization method with parameter subspace clustering constraint for endmember extraction in hyperspectral imagery. ISPRS J Photogramm Remote Sens 128:27–39

30. Tipping ME, Bishop CM (1999) Mixtures of probabilistic principal component analyzers. Neural Comput 11(2):443–482

31. Tolić D, Antulov-Fantulin N, Kopriva I (2018) A nonlinear orthogonal non-negative matrix factorization approach to subspace clustering. Pattern Recogn 82:40–55

32. Tsakiris MC, Vidal R (2017a) Algebraic clustering of affine subspaces. IEEE Trans Pattern Anal Mach Intell 40(2):482–489

33. Tsakiris MC, Vidal R (2017b) Filtrated algebraic subspace clustering. SIAM J Imaging Sci 10(1):372–415

34. Vidal R, Favaro P (2014) Low rank subspace clustering (lrsc). Pattern Recogn Lett 43:47–61

35. Vidal R, Ma Y, Sastry S (2005) Generalized principal component analysis (gpca). IEEE Trans Pattern Anal Mach Intell 27(12):1945–1959

36. Von Luxburg U (2007) A tutorial on spectral clustering. Stat Comput 17(4):395–416

37. Wang J, Shi D, Cheng D, Zhang Y, Gao J (2016) Lrsr: low-rank-sparse representation for subspace clustering. Neurocomputing 214:1026–1037

38. Wu Z, Yin M, Zhou Y, Fang X, Xie S (2018) Robust spectral subspace clustering based on least square regression. Neural Process Lett 48(3):1359–1372

39. Xiao Y, Wei J, Wang J, Ma Q, Zhe S, Tasdizen T (2020) Graph constraint-based robust latent space low-rank and sparse subspace clustering. Neural Comput Appl 32(12):8187–8204

40. Xie X, Guo X, Liu G, Wang J (2018) Implicit block diagonal low-rank representation. IEEE Trans Image Process 27(1):477–489

41. Xu J, Yu M, Shao L, Zuo W, Meng D, Zhang L, Zhang D (2019) Scaled simplex representation for subspace clustering. IEEE Transactions on Cybernetics

42. Xu X, Cheong LF, Li Z (2018) Motion segmentation by exploiting complementary geometric models. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp 2859–2867

43. Yang J, Liang J, Wang K, Rosin PL, Yang M (2020) Subspace clustering via good neighbors. IEEE Trans Pattern Anal Mach Intell 42(6):1537–1544, https://doi.org/10.1109/TPAMI.2019.2913863

44. Yang X, Jiang X, Tian C, Wang P, Zhou F, Fujita H (2020a) Inverse projection group sparse representation for tumor classification: A low rank variation dictionary approach. Knowl-Based Syst 196:105768

45. Yang Z, Liang N, Yan W, Li Z, Xie S (2020b) Uniform distribution non-negative matrix factorization for multiview clustering. IEEE Transactions on Cybernetics

46. Zhang H, Zhai H, Zhang L, Li P (2016) Spectral–spatial sparse subspace clustering for hyperspectral remote sensing images. IEEE Trans Geosci Remote Sens 54(6):3672–3684

47. Zhang X, Yang Y, Li T, Zhang Y, Wang H, Fujita H (2021) Cmc: a consensus multi-view clustering model for predicting alzheimer's disease progression. Comput Methods Prog Biomed 199:105895

48. Zhang Y, Yang Y, Li T, Fujita H (2019) A multitask multiview clustering algorithm in heterogeneous situations based on lle and le. Knowl-Based Syst 163:776–786

49. Zheng R, Li M, Liang Z, Wu FX, Pan Y, Wang J (2019) Sinnlrr: a robust subspace clustering method for cell type detection by non-negative and low-rank representation. Bioinformatics 35(19):3642–3650

50. Zhu P, Zhu W, Hu Q, Zhang C, Zuo W (2017) Subspace clustering guided unsupervised feature selection. Pattern Recogn 66:364–374

Stanley Ebhohimhen Abhadiomhen received his B.Eng. degree in Electrical/Electronics Engineering from Ambrose Alli University Edo state in 2011. Then, he received his MSc in Computing from Oxford Brookes University, Oxford United Kingdom in 2013. He is currently a lecturer with the Department of Computer Science, University of Nigeria, Nsukka, and studying for his Ph.D. at Jiangsu University. His research interest includes Machine learning, Computer vision, Pattern Recognition and Deep Learning.

Dr. XiangJun Shen received the M.S. degree in Computer Science in 2003 from the Jiangsu University, China and the Ph.D. degree in Automation in 2006 from the University of Science and Technology, China. He is currently a Professor with the Department of Software Engineering, School of Computer Science and Communication Engineering, Jiangsu University. His research interests in the areas of cross multimedia analysis, computer vision, pattern recognition, and statistical machine learning.

ZhiYang Wang received his B.S. degree in computer science and technology from Southeast University Chengxian College in 2019 and is currently pursuing master's degree in JiangSu University. His research interests are in the areas of machine learning, image processing and pattern recognition.