Neurocomputing 526 (2023) 19–29

Contents lists available at ScienceDirect

Neurocomputing

journal homepage: www.elsevier.com/locate/neucom

Projection-preserving block-diagonal low-rank representation for subspace clustering

Zisen Kong [a,b], Dongxia Chang [a,b,*], Zhiqiang Fu [a,b], Jiapeng Wang [a,b], Yiming Wang [a,b], Yao Zhao [a,b]

[a] Institute of Information Science, Beijing Jiaotong University, Beijing, China
[b] Beijing Key Laboratory of Advanced Information Science and Network Technology, Beijing, China

ARTICLE INFO

Article history:
Received 8 September 2022
Revised 8 December 2022
Accepted 8 January 2023
Available online 13 January 2023

Keywords:
Subspace clustering
Projection preserve
Block diagonal regularity
Low-rank representation

ABSTRACT

In this paper, a novel model named projection-preserving block-diagonal low-rank representation (PBDIR) is proposed and can obtain a more distinguishable representation matrix for clustering. PBDIR acquires a more advantageous representation by extracting the essential features. Specifically, we introduce a projection matrix to our model to learn a new feature space that can capture more significant features. Therefore, our model learns a more robust representation, which can reduce noise interference. Meanwhile, we introduce a block-diagonal regularization to ensure that the obtained representation matrix involves a k-block diagonal, where k denotes the number of clusters. This term brings more benefits for clustering tasks. Experimental results on real datasets show that our model can significantly improve the clustering performance and the proposed approach is robust against Gaussian noise, Multiplicative noise, and Salt-and-Pepper noise.

1. Introduction

As an important unsupervised data mining method, clustering analysis has received more attention in recent years. Subspace clustering, as a significant clustering method, aims at exploring the subspace structure in the data spaces. It assumes that the samples come from the union of k subspaces, where k subspace corresponds to k clusters. Hence, the task of subspace clustering is to find the samples sampled from the same subspace [1]. Due to the outstanding performance of subspace, it has been widely used in a wide variety of fields such as motion segmentation [2,3], face image application [4,5], image recognition [6–9], sparse coding [10] and human face clustering [11,12].

In the field of subspace clustering, Sparse subspace clustering (SSC) [13] and low rank representation (LRR) [14] are two of the most representative algorithms. The main difference between SSC and LRR is how to effectively grasp the potential relationship of data points to obtain a good representation matrix. However, SSC only considers the sparse representation of data points in each single subspace, lacking guidance on the global structure of the entire data. Unlike SSC, LRR jointly seeks low-rank representations of all data points. In contrast, it can better capture the global information around each sample and show a good application prospect. Fei et al. [15] combined the adaptive distance penalty with subspace classification (LRRADP), which reveals the intrinsic structure between data by weighting nearby sample pairs. In addition, in order to make better use of local geometric information, [16] proposes to extract robust feature information by integrating manifold learning and low-rank representation into a framework, which can reduce the negative effects of corruption.

In practical applications, due to the interference of noise or outliers, these methods may be difficult to obtain a good block diagonal representation. In fact, as for data with clustering number k, exactly k-connected components are conducive to obtaining prominent clustering results [17]. Inspired by this, a block diagonal prior method [18] based on graph Laplace constraint is proposed, which forces the coefficient matrix to have exactly k connected components. However, due to the strict assumptions, the optimal solution method may not be stable enough, and the theoretical convergence conditions are weak [19]. To resolve the problem, Lu et al. [19] proposed a more general method, which directly obtains clustering result by imposing block diagonal rule constraints. The purpose of using block diagonal regularization is to properly group the data points in the union set of k subspaces into k clusters corresponding to their subspaces. This is conducive to improving clustering performance.

In addition, some clustering methods based on deep neural networks have been proposed. Ji et al. [20] combine the self-encoder with subspace clustering, and add a self expression layer in the

* Corresponding author at: Institute of Information Science, Beijing Jiaotong University, Beijing, China.
E-mail address: dxchang@bjtu.edu.cn (D. Chang).

https://doi.org/10.1016/j.neucom.2023.01.051
0925-2312/

encoding and decoding process to improve the clustering performance.

The effectiveness of the above method has been proved. However, the above method is to obtain the representation in the original space. Due to the influence of dimension and structure, the original space has feature redundancy [21], which may affect the correct clustering.

To improve this problem, we introduce projection into the low rank representation framework that encourages more discriminating representation for subspace clustering. Moreover, a block diagonal term is introduced to ensure that the obtained representation matrix involves k-block diagonal. Accordingly, the innovative projection-preserving block-diagonal low-rank representation (PBDIR) method is proposed for robust subspace clustering. Our contributions are summarized as follows:

- We propose a novel projection method to improve representation learning to make it more discriminative for clustering.
- We add a block diagonal regular term to the projection space, which makes the representation matrix more robust by directly pursuing the block diagonally.
- Our model obtains a better representation matrix by jointly optimizing the projection transform and block-diagonal regular terms, and extensive experimental results demonstrate the validity of our model.

The rest of this paper is organized as follows. Section 2 introduces the related work and explains the notations used in our paper. Section 3 presents the proposed projection-preserving block diagonal low-rank representation (PBDIR) and its solution. In Section 4, we analyze the convergence and complexity of the proposed algorithm. Section 5 reports experimental results on real data sets. Finally, Section 6 concludes our paper.

2. Related work

In this section, we first explain the symbols mentioned in our paper and then introduce the related work. The symbols are given in Table 1.

2.1. Low-rank representation

As a classical subspace clustering method, the basic idea of LRR [14] is to capture the lowest rank representation of data. Suppose that there is a dataset $X = \{x_1, x_2, x_3, \ldots, x_n\} \in \mathbb{R}^{d\times n}$ (d and n are the dimension and the number of input data, respectively). Considering the influence of noises, a robust LRR model can be written as

$$\min_{Z} \|Z\|_* + \lambda\|E\|_1, \ s.t. X = XZ + E \tag{1}$$

Table 1
Symbol description.

Symbols	Explanation
X	Original data sample
x_i	The i-th column of vector x
X_{ij}	the entry of i-th row and j-th column
$X_{i,*}$	the i-th row of X
$X_{*,i}$	the i-th column of X
$\|X\|_F$	The Frobenius norm of X
$\|X\|_*$	The nuclear norm of X
$\|X\|_1$	The ℓ_1 norm of X
X^T	The transpose of X
$tr(X)$	The trace of X
$\odot$	The Hadamand product
$\mathbf{1}$	The vector in which elements are 1
Δ	The matrix with all elements are 1

where $\|\cdot\|_*$ denotes the nuclear norm, $\lambda > 0$ is a weighting parameter to balance the intensity between E and Z. $\|E\|_1$ is the ℓ_1-norm to constraint noises and outliers.

2.2. LRRADP

Although LRR has been proved to be able to discover the global structure of the dataset, it is not sensitive to corrupted data, it may fail to capture the local structure of data. This problem can be solved by using the nearest neighbor information between samples to obtain the relationship of data [22], then a distance penalty method [15] is proposed. The core idea of LRRADP is to put low-rank representation with adaptive distance penalty because data points close to each other are more likely to belong to the same subspace and have the same label [23].

$$\min_{Z,E} \|Z\|_* + \lambda_1\|E\|_1 + \lambda_2 tr(\Delta(D \odot Z)), \ s.t. X = XZ + E, Z \geqslant 0 \tag{2}$$

where $D \in \mathbb{R}^{n\times n}$ is the distance matrix of X. $D_{ij} = \|X_{*,i} - X_{*,j}\|_2^2$ adopts the adaptive distance penalty parameter. $\odot$ is the Hadamand product, λ_1 and λ_2 are the regularization.

Generally, adjacent samples belong to the same subspace and thereby points that are close to each other should have large weights. Based on this, LRRADP can construct a good affinity graph. The graph identified can capture both the global structure and local structure to obtain a better representation. However, the discriminative power of Euclidean distance deteriorates due to the noise and redundant features in the original space, which may affect the performance.

3. Projection-preserving block-diagonal low-rank representation

In this section, the projection-preserving block-diagonal low-rank representation (PBDIR) algorithm based on LRR framework is introduced. At first, we introduce projection into the low-rank representation to reduce the interference of noise and outliers in the original samples. Then, the strategies of the distance penalty term and the block diagonal constraint are simultaneously taken into consideration. Finally, We give the implementation process and complexity analysis of PBDIR algorithm.

3.1. The model of PBDIR

In practical application, the original data may have noise and outliers due to the influence of illumination, shooting angle, and other factors [24]. In order to obtain better features, projection has proved to be an effective method [25]. Its basic idea is to map the data in the original space to a new subspace, which retains the structural features of the data while filtering part of the noise. In addition, When the dictionary is sufficient, the representation based on Frobenius norm is completely based on the nuclear norm [26]. So we employ the Frobenius norm as the primary metric to measure the reconstructive errors. Then, original data can be approximately reconstructed by a low-rank matrix Z and a projection transformation matrix P. In addition, to capture the local structure of data, we introduce the projection distance penalty to obtain the nearest neighbor relationship between data in the transformed space. Therefore, the PBDIR model can be represented as

$$\min_{Z,P} \underbrace{\sum_{i,j}\|P^TX_{*,i} - P^TX_{*,j}\|_F^2 Z_{i,j}}_{Projection\ \ Distance\ \ penalty} + \lambda_2 \underbrace{\|P^TX - P^TXZ\|_F^2}_{Projection\ \ representation\ \ error} + \lambda_1\|Z\|_F^2 \quad s.t. P^TP = I, Z \geqslant 0, Z_{i,i} = 0, Z_{*,i}\mathbf{1} = \mathbf{1} \tag{3}$$

where P^TX denotes the projection space of X. $\|\cdot\|_F$ is the Frobenius norm, which can be defined as $\|X\|_F^2 = \sum_{i,j} x_{i,j}^2$. The projection matrix $P \in \mathbb{R}^{h\times d}(h \ll n)$ maps the data to the projection space. To avoid a trivial solution, the projection matrix P is restricted to column orthogonal, i.e., $P^TP = I$. The constraint $Z \geqslant 0$ ensures that the representation matrix is non-negative. $Z_{i,i} = 0$ aims to avoid self-representation of data, which means that each point x_i can only be expressed as a linear combination of other points $x_j (i \neq j)$. The constraint $Z_{*,i}\mathbf{1} = 1$ avoids all elements in a row being 0.

From Eq. (3), for an ideal representation matrix, it should have a symmetric $k-block$ diagonal structure. However, in practice its symmetric structure is often difficult to guarantee when the data set is disturbed by noise or outliers. To solve this problem, we further introduce a block diagonal constraint to the representation matrix to improve its block symmetric performance. Therefore, our model can be further modified as

$$\min_{Z,P} \underbrace{\sum_{i,j}\|P^TX_{*,i} - P^TX_{*,j}\|_F^2 Z_{i,j}}_{\text{Projection Distance penalty}} + \lambda_1\|Z\|_F^2 + \lambda_2 \underbrace{\|P^TX - P^TXZ\|_F^2}_{\text{Projection representation error}} + \lambda_3 \underbrace{\|\frac{Z+Z^T}{2}\|_k}_{\text{Block Constraint}} \; s.t. P^TP = I, Z \geqslant 0, Z_{i,i} = 0, Z_{*,i}\mathbf{1} = \mathbf{1} \tag{4}$$

where $\left(Z + Z^T\right)/2$ guarantees the symmetry of the representation matrix. By introducing the blocking constraint, the obtained representation matrix satisfies the k-diagonal structure, which is beneficial to the clustering task. $\lambda_1, \lambda_2, \lambda_3$ are the regular parameters.

In our model, the projection matrix and block diagonal constraint will enhance each other as 1) the block diagonal constraint can guide the projection matrix P to learn more favorable features; 2) a better P can enhance the discrimination of the representation matrix Z. Finally, these two terms will help the model learn a better representation matrix to improve clustering performance.

3.2. Solution to PBDIR

The proposed model is solved using the alternating direction multiplier method (ADMM) [27]. An auxiliary variable S is introduced to make the objective function Eq. (4) separable. Using S, our model can be rewritten as

$$\min_{Z,S,P} \sum_{i,j}\|P^TX_{*,i} - P^TX_{*,j}\|_F^2 S_{i,j} + \lambda_1\|Z\|_F^2 + \lambda_2\|P^TX - P^TXZ\|_F^2 + \lambda_3\|\frac{S+S^T}{2}\|_k \; s.t. P^TP = I, S \geqslant 0, S_{i,i} = 0, S_{*,i}\mathbf{1} = \mathbf{1}, Z = S \tag{5}$$

Then, the augmented Lagrangian formula can be written as

$$\min_{Z,S,P,C_1} \sum_{i,j}\|P^TX_{*,i} - P^TX_{*,j}\|_F^2 S_{i,j} + \lambda_1\|Z\|_F^2 + \lambda_2\|P^TX - P^TXZ\|_F^2 + \lambda_3\|\frac{S+S^T}{2}\|_{\boxed{k}} + \frac{\mu}{2}\|Z - S + \frac{C_1}{\mu}\|_F^2 \tag{6}$$

where C_1 is the Lagrange multiplier, μ is the penalty parameter ($\mu > 0$). For the optimization problem Eq. (7), we can sequentially update the variables Z, S, and P by the alternating direction multiplier, i.e., update one variable, while other variables remain unchanged. The detailed iterative update steps are gvien as follow.1) **Update Z with P and S fixed.** Z can be obtained by minimizing the following objetive function

$$\min_Z \lambda_1\|Z\|_F^2 + \lambda_2\left\|P^TX - P^TXZ\right\|_F^2 + \frac{\mu}{2}\left\|Z - S + \frac{C_1}{\mu}\right\|_F^2 \tag{7}$$

By setting the derivative of Eq. (7) to zero, we can get the closed form solution directly, i.e.,

$$Z = \left(2\lambda_1 I + \mu I + 2\lambda_2 X^TPP^TX\right)^{-1}\left(2\lambda_2 X^TPP^TX + \mu S - C_1\right) \tag{8}$$

where I is the identity matrix.2) **Update S with Z and P fixed.** S can be obtained by minimizing the following problem

$$\min_S \sum_{i,j}\left||P^TX_{*,i} - P^TX_{*,j}|\right|_F^2 S_{i,j} + \lambda_3\left|\frac{S+S^T}{2}\right|_k + \frac{\mu}{2}\left||Z - S + \frac{C_1}{\mu}|\right|_F^2 s.t. S \geqslant 0, S_{i,i} = 0, S_{*,i}\mathbf{1} = \mathbf{1} \tag{9}$$

From Eq. (9), the $k-block$ diagonal regularizer equivalent to the sum of the k smallest eigenvalues of L_B [28], then $\|\cdot\|_{\boxed{k}}$ can be obtained by

$$\|\frac{S+S^T}{2}\|_{\boxed{k}} = \sum_{i=n-k+i}^{n}\lambda_i(L_B) \tag{10}$$

where L_B is the Laplacian matrix, which can be defined as $Diag\left(\left(S\mathbf{1} + S^T\mathbf{1}\right)/2\right) - \left(S + S^T\right)/2), \lambda_i(L_B)$ is the $i-th$ eigenvalue of L_B. Inspired by [29], we reformulate $\|\cdot\|_{\boxed{k}}$ as

$$\sum_{i=n-k+1}^{n}\lambda_i(L_B) = \min_W\langle L_B, W\rangle, \; s.t. 0 \preceq W \preceq I, Tr(W) = k \tag{11}$$

Substituting Eq. (12) into Eq. (11) yields

$$\|\frac{S+S^T}{2}\|_{\boxed{k}} = \min_W\langle L_B, W\rangle, \; s.t. 0 \preceq W \preceq I, Tr(W) = k \tag{12}$$

Therefore, Eq. (9) can be rewritten as

$$\min_S \sum_{i,j}\left||P^TX_{*,i} - P^TX_{*,j}|\right|_F^2 S_{i,j} + \lambda_3\langle L_B, W\rangle + \frac{\mu}{2}\left||Z - S + \frac{C_1}{\mu}|\right|_F^2 s.t. S \geqslant 0, S_{i,i} = 0, S_{*,i}\mathbf{1} = \mathbf{1} \tag{13}$$

Further, Eq. (13) can be modified as

$$\min_S\left\|P^TX_{*,i} - P^TX_{*,j}\right\|_F^2 S_{i,j} + \lambda_3\langle Diag(B\mathbf{1}) - B, W\rangle + \frac{\mu}{2}\left\|Z - S + \frac{C_1}{\mu}\right\|_F^2 \tag{14}$$

where $B = \left(S + S^T\right)/2, W = FF^T, F$ is the eigenvector of the k smallest eigenvalues of L_B.

This problem is equivalent to

$$\min_S \; tr\left(D_{A_{i,j}}^T S\right) + \frac{\mu}{2}\|Z - S + \frac{C_1}{\mu} + \frac{\lambda_3}{\mu}\langle Diag(B\mathbf{1}) - B, W\rangle\|_F^2 \; s.t. \; S \geqslant 0, S_{i,i} = 0 \tag{15}$$

where $D_{A_{i,j}} = \sum_{i,j}\|PX_{*,i} - PX_{*,j}\|_2^F S_{i,j}$.

For (15), it is obvious that

Table 2
The experiment setting on seven databases.

Database	n	d	k
COIL20	1440	1024	20
Jaffe	213	676	10
ORL	400	1024	40
USPS	1854	256	10
Digit	1797	64	10
Yale	165	1024	15
YaleB	2414	1024	38

$$\min_{S} \; tr\left(D_{A_{ij}}^{T} S\right) + \frac{\mu}{2}\|Z - S + \frac{C_1}{\mu} + \frac{\lambda_3}{2\mu}\left(D_w + D_w^T\right)\|_F^2 s.t.\ S \geqslant 0, S_{i,i} = 0 \tag{16}$$

where $D_w = diag(W)\mathbf{1}^T - W$.

Deriving S in Eq. (16)we have

$$S = \frac{2\mu Z + 2C_1 - 2D_{A_{ij}} + \lambda_3\left(D_w + D_w^T\right)}{2\mu} \; s.t.\ S \geqslant 0, S_{i,i} = 0 \tag{17}$$

We can obtain the unconstrained solution of variable S in the following way. First, We set the diagonal element to 0 to satisfy $S_{i,i}$ = 0, we have

$$S_{ij} = \begin{cases} 0 & j = i \\ \frac{2\mu Z_{ij} + 2(C_1)_{ij} - 2D_{A_{ij}} + \lambda_3\left(D_w + D_w^T\right)}{2\mu} & j \neq i \end{cases} \tag{18}$$

Then, in order to ensure all elements of the matrix S are not less than 0, we have

$$S_{ij} = \begin{cases} 0 & S_{ij} \leqslant 0 \\ S_{ij} & S_{ij} > 0 \end{cases} \tag{19}$$

3) **Update P with Z and S fixed.** P can be gotten by minimizing the following problem

$$\min_{P} \sum_{i,j} \left\|P^T X_{*,i} - P^T X_{*,j}\right\|_F^2 S_{ij} + \frac{\lambda_2}{2}\left\|P^T X - P^T X Z\right\|_F^2 \; s.t. \quad P^T P = I \tag{20}$$

Table 3
The clustering performance of compared methods on seven datasets.

Data	Metrics	COIL20	Jaffe	ORL	USPS	Digit	Yale	YaleB
SSC [13]	ACC	72.57	93.43	63.25	62.42	79.19	49.09	31.77
	NMI	83.20	91.15	78.62	57.07	76.16	48.05	47.95
	ARI	73.05	86.42	53.33	50.05	69.12	22.43	20.88
LRR [14]	ACC	65.49	100	71.25	68.34	79.86	45.45	70.46
	NMI	75.41	100	49.59	68.22	78.14	49.59	72.99
	ARI	60.42	100	69.28	57.24	72.18	22.69	46.27
LSR [34]	ACC	61.53	98.59	74.25	67.64	79.41	47.88	67.40
	NMI	72.52	97.80	83.74	67.54	76.49	56.47	70.46
	ARI	59.18	97.23	64.94	55.48	70.67	31.49	55.27
LatLRR [35]	ACC	61.39	100	60.25	50.49	78.19	59.39	72.00
	NMI	72.52	100	76.96	57.75	73.35	56.45	72.86
	ARI	53.65	100	45.35	38.81	67.31	31.89	38.10
LSLRSSC [36]	ACC	65.90	100	74.75	72.11	78.58	49.70	50.04
	NMI	77.19	100	84.83	69.33	77.19	49.96	57.37
	ARI	55.78	100	63.16	62.11	70.15	29.26	23.79
RKLRR [37]	ACC	64.86	100	68.00	62.14	79.13	47.27	66.21
	NMI	75.46	100	79.37	66.83	76.32	50.34	61.18
	ARI	57.42	100	54.32	55.12	69.51	24.76	47.06
LRRADP [15]	ACC	78.89	97.65	64.75	70.66	81.41	54.55	45.65
	NMI	84.47	96.54	78.49	75.63	77.53	53.67	46.92
	ARI	71.64	95.00	48.95	65.78	70.21	29.63	17.35
IBDIR [38]	ACC	67.29	88.73	64.75	69.74	81.58	52.73	68.54
	NMI	73.80	86.44	75.34	68.10	80.86	50.78	80.43
	ARI	55.71	82.04	43.62	59.77	70.08	24.81	55.10
BDR [19]	ACC	84.31	100	71.00	67.31	76.04	56.36	70.75
	NMI	93.24	100	84.75	68.05	76.56	54.38	72.18
	ARI	81.28	100	58.00	55.56	71.06	29.48	36.55
FULRR [39]	ACC	66.46	95.31	68.00	69.31	79.86	51.52	59.32
	NMI	78.33	93.75	83.44	67.33	75.44	54.07	75.63
	ARI	50.48	90.38	48.96	57.47	67.37	24.20	27.87
PBDIR	ACC	99.31	100	76.25	77.83	81.36	59.39	77.15
	NMI	99.19	100	85.63	78.23	81.85	57.84	80.66
	ARI	98.65	100	63.94	69.50	71.68	33.57	61.11

In fact, for matrix X, the equivalence relationship can be obtained in the following ways:

$$\sum_{i,j=1}^{n}\|P^TX_*,i-P^TX_*,j\|_F^2S_{i,j}=2Tr\left(P^TX^TL_ZXP\right) \tag{21}$$

Substituting Eq. (21) into Eq. (20) gives

$$\min_P 2tr\left(P^TXL_ZX^TP\right)+\frac{\lambda_2}{2}\left\|P^TX-P^TXZ\right\|_F^2 s.t.\quad P^TP=I \tag{22}$$

where L_Z is the Laplacian matrix of Z obtained by $L_Z=D_Z-W_Z, D_Z=diag(sum(W_Z))$ and $W_Z=\left(Z+Z^T\right)/2$. Then Eq. (20) can be further rewritten as

$$\min_P 2tr\left(P^TXL_ZX^TP\right)+\frac{\lambda_2}{2}tr\left(P^T(X-XZ)(X-XZ)^TP\right) s.t, P^TP=I \tag{23}$$

The solution of Eq. (23) can be gotten by solving the following minimum eigenvalues problem

$$\left(P^TXL_sX^TP+\lambda_2\left(P^T(X-XZ)(X-XZ)^TP\right)=\gamma P\right. \tag{24}$$

where $P=[p_1,p_2,\ldots,p_h]\in\mathbb{R}^{d\times h}$, The column vectors p_i $(1\leqslant i\leqslant h)$ in P are the eigenvectors corresponding to the first h-dimensional $(h\leqslant n)$ eigenvalue of Eq. (23).4) **Update Lagrangian multiplier** C_1 **and penalty parameter** ρ

$$\mu=min(\rho\mu,\mu_{max}) \tag{25}$$

$$C_1=C_1+\mu(Z-S) \tag{26}$$

The detailed process of our algorithm is summarized in Algorithm 1.

Algorithm 1 Solving Problem (6) by ADMM

Input: Data matrix X, the maximum number of iterations *maxiter*, and parameter $\lambda_1,\lambda_2,\lambda_3$;
Output: Z,S,P;
1: **Initialization:** Initializing Z by constructing the k-nearest neighbor representation,
$S=Z, C_1=0, \mu=0.1, \rho=1.1, \mu_{max}=10^8$;
2: **while** not converged **do:**
3: Update variable Z as (8)
4: Update variable S as (19)
5: Update variable P as (24)
6: Update variable μ as (25)
7: Update Lagrange multipliers C_1 as (26)
8: **end while**;

4. Algorithm analysis

In this section, the convergence and the complexity of our proposed PBDIR are discussed.

4.1. Convergence analysis

The objective function of PBDIR is a nonconvex optimization problem. We can use ADMM [30] algorithm to obtain a local optimal solution, but it is still difficult to directly prove the convergence of variables Z,S,P. Fortunately, inspired by [31–33], the convergence of our method can be given as follows:

Theorem 1. *Let* $\{\Psi(t)\}_{t=1}^{\infty}$ *be a sequence satisfying the condition* $\lim_{t\to\infty}(\Psi(t+1)-\Psi(t))=0$ *generated by ADMM algorithm (6). We can conclude that for any cumulative point of* $\{\Psi(t)\}_{t=1}^{\infty}$, *it is the KKT point of problem (5). And when* $\{\Psi(t)\}_{t=1}^{\infty}$ *converges to the KKT point, objective function (6) converges.*

Proof. The detailed proof of Theorem 1 is shown in the Appendix A. □

4.2. Computational complexity analysis

In this part, the computational complexity of Algorithm 1 is given. Solving Algorithm 1 is divided into three steps, e.g. step 3 to step 5. Step 3 needs to calculate $\left(2\lambda_1I+\mu I+2\lambda_2X^TP^TPX\right)$, and its computational complexity is $\mathcal{O}(n^3)$. It needs to compute a generalized eigenvalue problem in step 5, and the time complexity is $\mathcal{O}(n^3)$. The computational complexity of step 4 is $\mathcal{O}(n^2)$. Considering the number of iterations, the complexity of Algorithm 1 is $\mathcal{O}(\phi(2n^3+n^2))$, where ϕ is the number of iterations.

5. Experiments

In this section, we evaluated the performance of PBDIR on seven widely used benchmark datasets. The performance of PBDIR is compared with ten traditional subspace clustering algorithms in terms of three clustering evaluation metrics, namely ACC (Clustering accuracy), NMI (Normalized mutual information), and ARI (Adjusted Rand index). For these indicators, the higher the value, the better the performance of the clustering.

We compare our proposed PBDIR with the following methods, including SSC [13], LRR [14], LSR [34], LatLRR [35], LSLRSSC [36], RKLRR [37], LRRADP [15], IBDLR [38], BDR [19], FULRR [39].

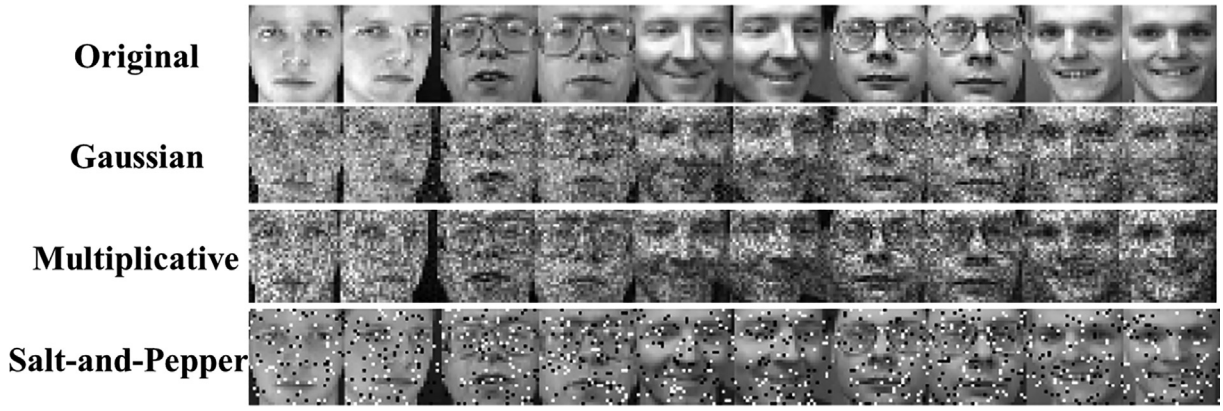


Fig. 1. Some images after original and added noise.

5.1. Datasets

For convenience, we summarize the seven sets in Table 2 with the characteristics of the data sets. The three columns show the number of data points n, the dimension of the feature space d, and the number of classes k for each data set.

5.2. Clustering results and analysis

The clustering results on the seven real datasets are shown in Table 3, where the ACC, NMI, and ARI are given and the best result is highlighted in red, while the second-best is reported in blue. From these three tables we can draw the following conclusions:

- Based on the overall results of seven data sets, compared with most of the compared algorithms, PBDIR outperforms the comparison algorithms on most datasets. For example, for the object image dataset COIL20, compared to the second highest algorithm BDR, PBDIR is 15% higher on the clustering index ACC.
- The selected comparison algorithm obtains the representation in the original data space, and the redundant data in the original space may affect the clustering performance. Compared with LSLRSSC, PBDIR obtains the representation in the learned projection space. The experimental results show the effectiveness of our algorithm.
- Compared with LRR, LSR, LatLRR, and FULRR, our algorithm has better performance in most cases. Since these algorithms capture the global structure by obtaining the lowest rank representation of the data, LRRADP acquires the local structure by adding distance penalty constraints at the same time. Thereby, this shows that learning the global and local structure of data is helpful to obtain correct clustering results.
- By comparing SSC, LSLRSSC, LRRADP, and PBDIR, PBDIR achieves better clustering performance. Because SSC only captures the local structure of data by obtaining the sparse representation of data, LRRADP introduces a distance penalty to get a more geometric structure, but PBDIR obtains a better discriminative representation through projection transformation.
- Both BDR and IBDIR use block constraints to obtain k diagonal blocks, which are robust to noise and outliers. Further, PBDIR integrates projection transformation, projection penalty, and block diagonal constraint into a joint framework, which makes our algorithm obtain higher accuracy.

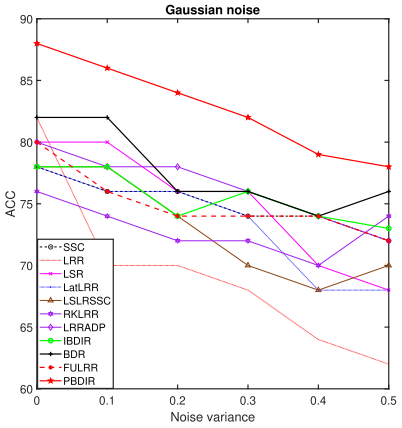
Overall, the experimental results prove the effectiveness of projection transformation and block diagonal constraint. By fusing it with the projection distance penalty constraint, the proposed PBDIR is superior to the comparison algorithm in performance.

5.3. Robustness against random noise interference

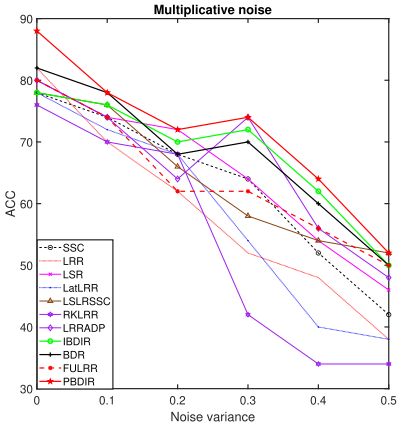
In this subsection, the ORL dataset is used to evaluate performance of PBDIR and other algorithms. To improve computing efficiency, we select a subset of ORL, which includes 50 images with 32×32 pixels. The visualization results of the original image and the image after adding noise are shown in Fig. 1 (The variance of Gaussian noise is 0.01, and the variance of Multiplicative noise and Salt-and-Pepper noise are both 0.1). Formally, three kinds of noise, i.e., Gaussian noise, Multiplicative noise, and Salt-and-Pepper noise are chosen as a comparison. The mean value of these three kinds of noise is zero. The Gaussian noise variance is set in the range of [0.01, 0.02, 0.03, 0.04, 0.05], and the other two noise variances range from [0.1, 0.2, 0.3, 0.4, 0.5].

As shown in Fig. 2, we find that PBDIR outperforms other algorithms in most cases. In particular, when the noise variance becomes larger, the rate of decline of PBDIR is slower than other algorithms, which means that PBDIR is better than the comparison algorithm in robustness. Further, we can see that PBDIR is more robust to Gaussian noise, which is approximately 10% higher than the comparison algorithm.

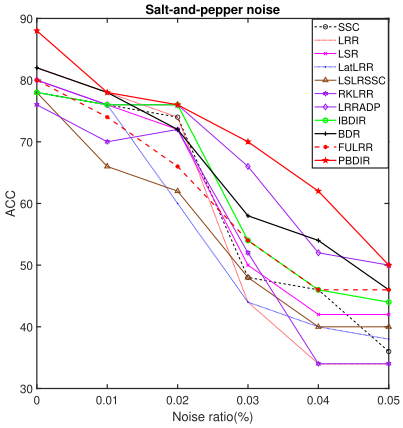
Moreover, we conducted face denoising and restoration experiments on the Yale dataset, adding Gaussian noise, multiplicative noise, and salt and pepper noise with a mean value of 0 and variance of 0.01 to the original image respectively. Fig. 3 shows original



(a) Gaussian noise



(b) Multiplicative noise



(c) Salt-and-Pepper noise

Fig. 2. Experimental results on (a) Gaussian noise, (b) Multiplicative noise, and (c) Salt-and-Pepper noise.

faces, noised faces, and recovered faces. It can be seen that PBDIR is superior in restoring damaged face images than others.

5.4. Robustness against random block

To verify the robustness of our algorithm to random block noise. We randomly add some occlusion patches at different positions of the image, and other positions of the image remain unchanged. The clustering results of the occlusion data of COIL20 and ORL datasets are shown in Fig. 4, where the size of the occlusion block is set to $[0 \times 0, 3 \times 3, 6 \times 6, 9 \times 9, 12 \times 12, 15 \times 15]$, and the images corresponding to different occlusion blocks are also displayed.

From the results shown in Fig. 4, we can find that: (1) Our algorithm PBDIR outperforms existing algorithms in terms of clustering performance. (2) Although the ACC value of the algorithm keeps decreasing when the block size becomes larger, the performance degradation of our algorithm is smaller than other algorithms. The results present its denoising ability for occlusions.

5.5. Performance evaluation of projection transformation

To show the performance of projection, we conduct experiments on the Yale and Jaffe datasets. We perform spectral clustering on the original space and the projected space respectively. The experimental results are given in Fig. 5.

As can be seen from Fig. 5, mapping data to the learned projection space helps to improve clustering performance. (reflected in X and P^TX) similarly, we compare the data representation of the original space with that of the projection space (reflected in XZ and P^TXZ), and we also get a similar conclusion. This is because compared with the original space, the projection space can eliminate some redundant data and obtain a more differentiated representation structure.

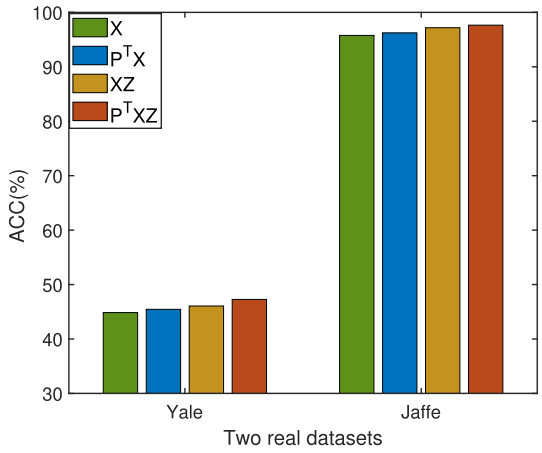


Fig. 5. Comparison of clustering accuracy of X, P^TX, XZ and P^TXZ on Yale and Jaffe datasets.

5.6. Performance evaluation of block diagonal representation structure

In this subsection, the representation matrix obtained by the algorithms SSC, LRR, LSR, LatLRR, LSLRSSC, RKLRR, LRRADP, IBDIR, BDR, FULRR, and PBDIR on the Jaffe dataset is shown in Fig. 6. Here, all the obtained representation matrices are treated symmetrically by $\left(|Z| + |Z|^T\right)$.

For LRR, SSC, LatLRR, IBDIR, RKLRR, FULRR, the obtained block diagonal structure contains more noise, which may affect the clustering performance. Compared with LSR, LSLRSSC, LRRADP, BDR and PBDIR obtains 10 exact block diagonals, and the structure obtained is closest to the block diagonal shape, which proves the effectiveness of block diagonal constraint.

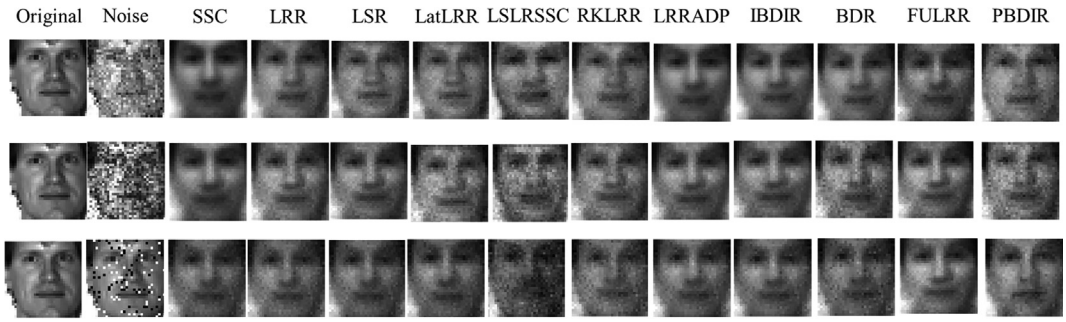


Fig. 3. Reconstruction images with noises on Yale dataset.

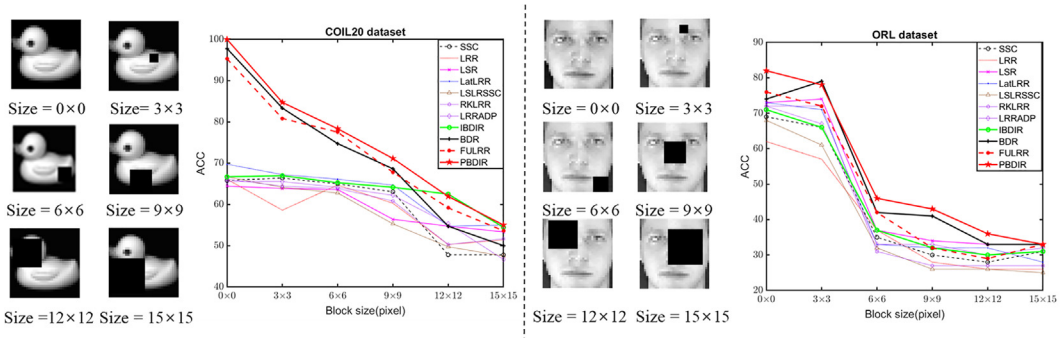


Fig. 4. Clustering performance on COIL20 and ORL datasets.

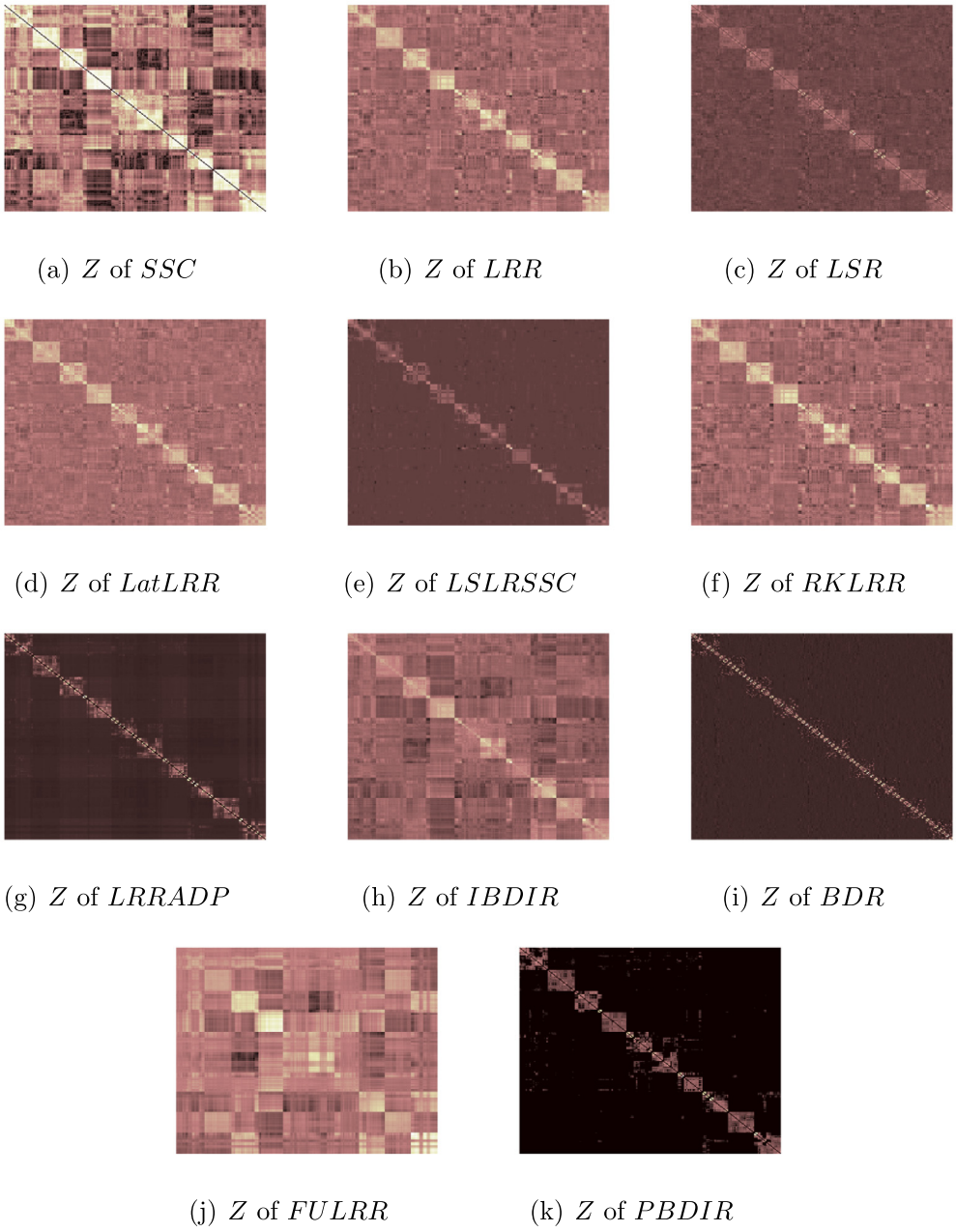


Fig. 6. Visualization of representation matrices of different algorithms.

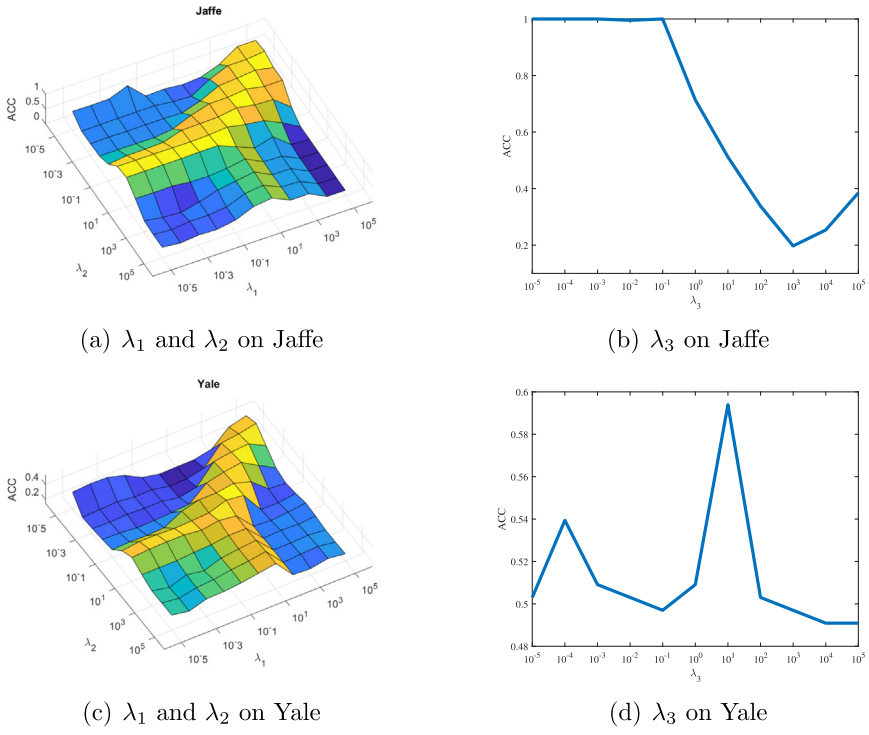
5.7. Parametric analysis

From the objective function of PBDIR in Eq. (5), we can see that there are three parameters λ_1, λ_2, and λ_3 in the proposed algorithm, which constrained control of low-rank constraint, projection representation constraint and block diagonal constraint separately. In this section, We visualize the sensitivity of three parameters on the Jaffe and Yale datasets, in which the three parameters are selected from $\left\{10^{-5}, 10^{-4}, 10^{-3}, \ldots, 10^{3}, 10^{4}, 10^{5}\right\}$.

In our experiments, we first fix λ_3 and then adjust λ_1 and λ_2. The change of sensitivity is shown in Fig. 7(a), then we fix λ_1, λ_2 and perform sensitivity analysis on λ_3. The results are shown in Fig. 7 (b). We use the same method on the Yale dataset and the results obtained for the three parameters are shown in Fig. 7(c,d). For the Jaffe dataset, when $10^{-3} \leqslant \lambda_1 \leq 10^{1}$, $10^{-1} \leqslant \lambda_2 \leq 10^{1}$ and $\lambda_3 \leq 10^{0}$, PBDIR can get better and accurate results. For the Yale dataset, when $\lambda_1 \leq 10^{-1}, 10^{-1} \leqslant \lambda_2 \leq 10^{1}$ and $10^{0} \leqslant \lambda_3 \leq 10^{2}$ PBDIR can get better accurate results. Through the experiment, we can get a rough parameter selection range and further get the best parameter combination based on the rough parameter range to find the optimal solution faster.

6. Conclusion

In this paper, we propose a novel subspace clustering method named projection-preserving block-diagonal low-rank representation (PBDIR). In our algorithm, a projection matrix is introduced to



(a) λ_1 and λ_2 on Jaffe (b) λ_3 on Jaffe

(c) λ_1 and λ_2 on Yale (d) λ_3 on Yale

Fig. 7. Parametric analysis of PBDIR on the Jaffe and Yale dataset.

transform the original space into a better projection space which can effectively improve the discriminative of learned representation and weaken the interference caused by noise. In addition, low-rank and distance penalty constraint is using to capture the global and local structures of the data simultaneously. Moreover, the block-diagonal structure of the representation matrix is directly pursued by using the block-diagonal regular term to obtain a more optimal representation of the coefficient matrix. Experimental results show that our algorithm is superior to most existing algorithms.

CRediT authorship contribution statement

Zisen Kong: Conceptualization, Methodology, Writing-original-draft. **Dongxia Chang:** Writing-review-editing, Supervision. **Zhiqiang Fu:** Project-administration, Investigation. **Jiapeng Wang:** Formal-analysis, Data-curation. **Yiming Wang:** Investigation. **Yao Zhao:** Validation.

Data availability

Data will be made available on request.

Declaration of Competing Interest

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.

Acknowledgment

The authors would like to thank the editors and the anonymous referees for their helpful comments to improve the quality of the paper. This work was supported in part by the National Natural Science Foundation of China under Grant 62272035, in part by the National Key Research and Development of China (No. 2018AAA0102100), and in part by the National Natural Science Foundation of China (No. U1936212).

Appendix A. Proof of Theorem 1

First, we can find in Eq. (6) that its KKT point should satisfy the following KKT conditions:

$$Z^{\dagger} = (2\lambda_1 I + \mu I + 2\lambda_2 H)^{-1}(2\lambda_2 H + \mu S - C_1) \tag{27}$$

$$S^{\dagger} = Z + \frac{2C_1 - 2D_{A_{ij}} + \lambda_3\left(D_w + D_w^T\right)}{2\mu} \tag{28}$$

$$C_1^{\dagger} = C_1 + \mu(Z - S) \tag{29}$$

where $H = X^T P P^T X$. The subscript "†" is used to denote iterative values at the new iteration. Note that the variable P to be proved is not given because P does not involve the calculation of the Lagrange multiplier.

Next, we give some results of the convergence of the proposed ADMM algorithm. For convenience, we define the triplet as

$$\Psi \triangleq (Z, S, C_1) \tag{30}$$

Then, we discuss the relationship between $\Psi$and other variables within the triplet. In fact, if the point Ψ is the KKT condition of the Eq. (5), we have

$$Z - S = 0 \tag{31}$$

$$2\lambda_1 Z + 2\lambda_2 H(Z - I) + C_1 = 0 \tag{32}$$

$$\mu C_1 + D_A + \lambda_3\left(D_w + D_w^T\right) = 0 \tag{33}$$

We can update the ADMM formula in (6) to

$$C_1^{\dagger} - C_1 = \mu(Z - S) \tag{34}$$

$$(2\lambda_1 I + \mu I + 2\lambda_2 H)(Z^{\dagger} - Z) = (2\lambda_2 H + \mu S - C_1) - (2\lambda_1 I + \mu I + 2\lambda_2 H)Z \tag{35}$$

$$S^{\dagger} - S = Z - S + \frac{\mu C_1 + D_A + \lambda_3\left(D_w + D_w^T\right)}{\mu} \tag{36}$$

For the assumption $\Psi^{\dagger} - \Psi \to 0$, both sides of the above equation tend to 0. We rewrite $\Psi^{\dagger}$ in the original equation as $\Psi(t+1)$ and replacing $Z^{\dagger}$ by $Z(t+1), S^{\dagger}$ by $S(t+1)$ and $C_1^{\dagger}$ by $C_1(t+1)$.

If we assume that t tends to infinity, noted that $\Psi(t+1) = \Psi(t) + (\Psi(t+1) - \Psi(t))$. where $(\Psi(t+1) - \Psi(t))$ gradually disappears. We can obtain

$$Z(t) - S(t) \to 0 \tag{37}$$

$$\lambda_1 Z(t) + \lambda_2 H(Z(t) - I) + C_1(t) \to 0 \tag{38}$$

$$\mu C_1(t) + D_A + \lambda_3\left(D_w + D_w^T\right) \to 0 \tag{39}$$

If the variable C_1 can converge to a fixed point, e.g, $(C_1(t+1) - C_1(t)) \to 0$, we can get $(Z(t) - S(t)) \to 0$ as a result from Eq. (34). Therefore, the KKT condition and the proof of Eq. (37) are complete.

Next, for the variable Z, similar to Eq. (35). We can obtain the following equation as

$$\begin{aligned}&(2\lambda_1 I + \mu I + 2\lambda_2 H)(Z(t+1) - Z(t))\\ &= (2\lambda_2 H + \mu S - C_1) - (2\lambda_1 I + \mu I + 2\lambda_2 H)Z\\ &= \mu(S(t) - Z(t)) - (2\lambda_1 Z(t) + 2\lambda_2 H(Z(t) - I) + C_1) = 0\end{aligned} \tag{40}$$

If the variable Z also converge to a fixed point, e.g., $(Z(t+1) - Z(t)) \to 0$, then we can get $(2\lambda_1 Z + 2\lambda_2 H(Z - I) + C_1) \to 0$. since $(Z - S) \to 0$. Thus the KKT condition proof of Eq. (32) are complete.

Similarly, for the variable S, the following equation can be obtained

$$S(t+1) - S(t) = Z(t) + \frac{\mu(C_1(t)) + D_{A_{ij}} + \lambda_3\left(D_w + D_w^T\right)}{\mu} \tag{41}$$

From (37), we have shown that $(Z - S) \to 0$. Therefore, it is easy to obtain that $\left(\mu(C_1(t))_{ij} + D_{A_{ij}} + \lambda_3\left(D_w + D_w^T\right)\right) \to 0$, the proof of Eq. (33) are complete.

Finally, all the KKT conditions have been proved. Besides, we can observe that if there is a bounded sequence solution$\{\Psi(t)\}_{t=1}^{\infty}$ present in the proposed method, then under constraint $\lim_{t\to\infty}(\Psi(t+1) - \Psi(t)) = 0$, Eqs. (34)–(36) are approximately equal to 0. Therefore, $\{\Psi(t)\}_{t=1}^{\infty}$ meet KKT conditions, the proof of Theorem 1 is complete. □

References

[1] F. Nie, W. Chang, Z. Hu, X. Li, Robust subspace clustering with low-rank structure constraint, IEEE Trans. Knowl. Data Eng. 34 (2022) 1404–1415.
[2] W. Chen, E. Zhang, Z. Zhang, A laplacian structured representation model in subspace clustering for enhanced motion segmentation, Neurocomputing 208 (2016) 174–182.
[3] G. Xia, H. Sun, L. Feng, G. Zhang, Y. Liu, Human motion segmentation via robust kernel sparse subspace clustering, IEEE Trans. Image Process. 27 (2018) 135–150.
[4] C. Kotropoulos, K. Pitas, Subspace clustering applied to face images, in: 2nd International Workshop on Biometrics and Forensics, 2014, pp. 1–6.
[5] X. Zhang, D. Phung, S. Venkatesh, D.-S. Pham, W. Liu, Multi-view subspace clustering for face images, in: International Conference on Digital Image Computing: Techniques and Applications, 2015, pp. 1–7.
[6] J. Francis, M. Baburaj, S.N. George, An $l_{\frac{1}{2}}$ and graph regularized subspace clustering method for robust image segmentation, ACM Trans. Multim. Comput. Commun. Appl. 18 (2022) 53:1–53:24.
[7] H. Liu, J. Cheng, F. Wang, Sequential subspace clustering via temporal smoothness for sequential data segmentation, IEEE Trans. Image Process. 27 (2018) 866–878.
[8] J. Cheng, M. Qiao, W. Bian, D. Tao, 3d human posture segmentation by spectral clustering with surface normal constraint, Signal Process. 91 (2011) 2204–2212.
[9] J. Yu, D. Tao, M. Wang, Adaptive hypergraph learning and its application in image classification, IEEE Trans. Image Process. 21 (2012) 3262–3272.
[10] J. Yu, Y. Rui, D. Tao, Click prediction for web image reranking using multimodal sparse coding, IEEE Trans. Image Process. 23 (2014) 2019–2032.
[11] H. Chen, W. Wang, X. Feng, R. He, Discriminative and coherent subspace clustering, Neurocomputing 284 (2018) 177–186.
[12] Y. Qiu, P. Hao, Self-supervised deep subspace clustering network for faces in videos, Vis. Comput. 37 (2021) 2253–2261.
[13] E. Elhamifar, R. Vidal, Sparse subspace clustering, in: IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2009, pp. 2790–2797.
[14] G. Liu, Z. Lin, S. Yan, J. Sun, Y. Yu, Y. Ma, Robust recovery of subspace structures by low-rank representation, IEEE Trans. Pattern Anal. Mach. Intell. 35 (2013) 171–184.
[15] L. Fei, Y. Xu, X. Fang, J. Yang, Low rank representation with adaptive distance penalty for semi-supervised subspace classification, Pattern Recognit. 67 (2017) 252–262.
[16] L. Xie, M. Yin, X. Yin, Y. Liu, G. Yin, Low-rank sparse preserving projections for dimensionality reduction, IEEE Trans. Image Process. 27 (2018) 5261–5274.
[17] F. Nie, X. Wang, M.I. Jordan, H. Huang, The constrained laplacian rank algorithm for graph-based clustering, in: Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence, 2016, pp. 1969–1976.
[18] J. Feng, Z. Lin, H. Xu, S. Yan, Robust subspace segmentation with block-diagonal prior, in: IEEE Conference on Computer Vision and Pattern Recognition, 2014, pp. 3818–3825.
[19] C. Lu, J. Feng, Z. Lin, T. Mei, S. Yan, Subspace clustering by block diagonal representation, IEEE Trans. Pattern Anal. Mach. Intell. 41 (2019) 487–501.
[20] P. Ji, T. Zhang, H. Li, M. Salzmann, I.D. Reid, Deep subspace clustering networks, in: Annual Conference on Neural Information Processing Systems, 2017, pp. 24–33.
[21] L. Zhang, Y. Li, J. Zhang, P. Li, J. Li, Nonlinear sparse feature selection algorithm via low matrix rank constraint, Multim. Tools Appl. 78 (2019) 33319–33337.
[22] J. Zhang, J. Yu, D. Tao, Local deep-feature alignment for unsupervised dimension reduction, IEEE Trans. Image Process. 27 (2018) 2420–2432.
[23] C. Wang, J. Zhang, F. Du, G. Shi, Symmetric low-rank representation with adaptive distance penalty for semi-supervised learning, Neurocomputing 316 (2018) 376–385.
[24] W.K. Wong, Z. Lai, J. Wen, X. Fang, Y. Lu, Low-rank embedding for robust image feature extraction, IEEE Trans. Image Process. 26 (2017) 2905–2917.
[25] W. Yu, R. Wang, F. Nie, F. Wang, Q. Yu, X. Yang, An improved locality preserving projection with ℓ_1-norm minimization for dimensionality reduction, Neurocomputing 316 (2018) 322–331.
[26] X. Peng, C. Lu, Z. Yi, H. Tang, Connections between nuclear-norm and frobenius-norm-based representations, IEEE Trans. Neural Networks Learn. Syst. 29 (2018) 218–224.
[27] S.P. Boyd, N. Parikh, E. Chu, B. Peleato, J. Eckstein, Distributed optimization and statistical learning via the alternating direction method of multipliers, Found. Trends Mach. Learn. 3 (2011) 1–122.
[28] J. Cai, E.J. Candès, Z. Shen, A singular value thresholding algorithm for matrix completion, SIAM J. Optim. 20 (2010) 1956–1982.
[29] M. Belkin, P. Niyogi, Laplacian eigenmaps and spectral techniques for embedding and clustering, in: Proceedings of the 14th International Conference on Neural Information Processing Systems: Natural and Synthetic, 2001, pp. 585–591.
[30] Z. Lin, M. Chen, Y. Ma, The augmented lagrange multiplier method for exact recovery of corrupted low-rank matrices, arXiv preprint arXiv:1009.5055 (2010).
[31] J. Wen, N. Han, X. Fang, L. Fei, K. Yan, S. Zhan, Low-rank preserving projection via graph regularized reconstruction, IEEE Trans. Cybern. 49 (2019) 1279–1291.
[32] G. Gordon, R. Tibshirani, Karush-kuhn-tucker conditions, Optimization 10 (2012) 725.
[33] Y. Zhang, An alternating direction algorithm for nonnegative matrix factorization, Technical Report, 2010.
[34] C. Lu, H. Min, Z. Zhao, L. Zhu, D. Huang, S. Yan, Robust and efficient subspace segmentation via least squares regression, in: European Conference on Computer Vision, 2012, pp. 347–360.
[35] G. Liu, S. Yan, Latent low-rank representation for subspace segmentation and feature extraction, in: IEEE International Conference on Computer Vision, 2011, pp. 1615–1622.
[36] V.M. Patel, H.V. Nguyen, R. Vidal, Latent space sparse and low-rank subspace clustering, IEEE J. Sel. Top. Signal Process. 9 (2015) 691–701.
[37] S. Xiao, M. Tan, D. Xu, Z.Y. Dong, Robust kernel low-rank representation, IEEE Trans. Neural Networks Learn. Syst. 27 (2016) 2268–2281.
[38] X. Xie, X. Guo, G. Liu, J. Wang, Implicit block diagonal low-rank representation, IEEE Trans. Image Process. 27 (2018) 477–489.

[39] Q. Shen, Y. Liang, S. Yi, J. Zhao, Fast universal low rank representation, IEEE Trans. Circuits Syst. Video Technol. 32 (2022) 1262–1272.

Zisen Kong received the B.S. degree in mathematics from Tianjin University of Commerce, Tianjin, China, in 2021. He is currently pursuing the Ph.D. degree in Institute of Information Science, Beijing Jiaotong University. His current research interests include pattern recognition and subspace clustering.

Dongxia Chang received the M.S. degree in mathematics from Xidian University and the Ph.D. degree in control science and engineering from Tsinghua University in 2003 and 2009, respectively. She is currently an associate professor of the Institute of Information Science of Beijing Jiaotong University. Her research interests include clustering, pattern recognition, and image segmentation.

Zhiqiang Fu received the B.S. degree in electronic engineering from Beijing Jiaotong University, Beijing, China, in 2017. He is currently pursuing the Ph.D. degree in Institute of Information Science, Beijing Jiaotong University. His current research interests include pattern recognition and clustering.

Jiapeng Wang received his bachelor's degree in Internet of Things Engineering from Hebei University of Technology, Tianjin, China in 2021. He is currently studying for a master's degree in the School of Computer and Information Technology of Beijing Jiaotong University. His current research interests include pattern recognition and multimodal fusion.

Yiming Wang received the B.S. degree in computer science and technology from the Shandong Normal University, Jinan, China, in 2017, and the M.E. degree in electronics and communications engineering from Beijing Jiaotong University, Beijing, China, in 2019. He is currently working toward the Ph.D. degree in the school of computer and information technology at Beijing Jiaotong University. His current research interests include unsupervised learning and graph neural networks.

Yao Zhao received the Ph.D. degree from the Institute of Information Science, Beijing Jiaotong University (BJTU), Beijing, China, in 1996. He became an Associate Professor with BJTU in 1998 and became a Professor in 2001. From 2001 to 2002, he was a Senior Research Fellow with the Information and Communication Theory Group, Faculty of Information Technology and Systems, Delft University of Technology, Delft, The Netherlands. In 2015, he visited the Swiss Federal Institute of Technology, Lausanne, Switzerland (EPFL). From 2017 to 2018, he visited the University of Southern California. He is currently the Director of the Institute of Information Science, BJTU. His current research interests include image/video coding, digital watermarking and forensics, video analysis and understanding, and artificial intelligence. Dr. Zhao serves or served on the Editorial Boards of several international journals, including as an Associate Editor of the IEEE TRANSACTIONS ON CYBERNETICS, a Senior Associate Editor of the IEEE SIGNAL PROCESSING LETTERS, and an Area Editor of the Signal Processing: Image Communication. He was named as a Distinguished Young Scholar by the National Science Foundation of China in 2010, and was elected as a Chang Jiang Scholar of the Ministry of Education of China in 2013.