

Subspace clustering based on latent low rank representation with Frobenius norm minimization

Song Yu^{a,*}, Wu Yiquan^{a,b,c,d}

^aSchool of Electronic and Information Engineering, Nanjing University of Aeronautics and Astronautics, Jiangjun Avenue No. 29, Nanjing 211106, China

^bKey Laboratory of Port, Waterway and Sedimentation Engineering of the Ministry of Transport, Nanjing Hydraulic Research Institute, Guangzhou Street No. 223, Nanjing 210029 China

^cState Key Laboratory of Urban Water Resource and Environment, Nangang District, Huanghe Street No. 73, Harbin 150090, China

^dThe Key Laboratory of Rivers and Lakes Governance and Flood Protection of Yangtse River Water Conservancy Committee, Huangpu Street No. 23, Wuhan 430010, China

ARTICLE INFO

Article history:

Received 13 December 2016

Revised 9 November 2017

Accepted 10 November 2017

Available online 16 November 2017

Communicated by Zhao Zhang

Keywords:

Subspace clustering

Low rank representation

Latent low rank representation

Frobenius norm minimization

ABSTRACT

The problem of subspace clustering which refers to segmenting a collection of data samples approximately drawn from a union of linear subspaces is considered in this paper. Among existing subspace clustering algorithms, low rank representation (LRR) based subspace clustering is a very powerful method and has demonstrated that its performance is good. Latent low rank representation (LLRR) subspace clustering algorithm is an improvement of the original LRR algorithm when the observed data samples are insufficient. The clustering accuracy of LLRR is higher than that of LRR. Recently, Frobenius norm minimization based LRR algorithm has been proposed and its clustering accuracy is higher than that of LRR demonstrating the effectiveness of Frobenius norm as another convex surrogate of the rank function. Combining LLRR and Frobenius norm, a new low rank representation subspace clustering algorithm is proposed in this paper. The nuclear norm in the LLRR algorithm is replaced by the Frobenius norm. The resulting optimization problem is solved via alternating direction method of multipliers (ADMM). Experimental results show that compared with LRR, LLRR and several other state-of-the-art subspace clustering algorithms, the proposed algorithm can get higher clustering accuracy. Compared with LLRR, the running time of the proposed algorithm is reduced significantly.

© 2017 Elsevier B.V. All rights reserved.

1. Introduction

In real world applications, data samples are always high-dimensional. In order to deal with these high-dimensional data samples more efficiently, the underlying structure of these data samples should be detected and used. A widely accepted assumption is that these high-dimensional data samples lie in some low-dimensional manifold. To model such low-dimensional manifold, the most natural choice is to use linear subspace. If the data samples are drawn from one class, they can be considered as drawn from a single linear subspace and single linear subspace models [1–3] can be used. When the data samples are drawn from multiple classes, they can be considered as drawn from a union of linear subspaces and subspace clustering models can be used. Subspace clustering refers to the problem of segmenting a collection of data samples approximately drawn from a union of linear subspaces. Subspace clustering has been widely applied in

computer vision [4–6], image processing [7–9], pattern recognition and system identification [10]. The existing subspace clustering approaches can be divided into four main categories: iterative [11–16], algebraic [17–20], statistical [21,22] and spectral clustering-based approaches [23–27]. Compared to the first three kind of subspace clustering approaches, spectral clustering-based approaches are more precise and efficient and these approaches have drawn more research attention. Among spectral clustering-based approaches, there are two well-known approaches which are sparse subspace clustering (SSC) [28] and low rank representation (LRR) [29]. These two approaches share some similarity with each other. In SSC and LRR, data samples are represented as linear combinations of a dictionary composed of the data samples themselves. A convex optimization problem is formed through which the representation matrix can be solved. The concept from sparse representation [30–32] and low rank recovery [33–35] are used to regularize the representation matrix. LRR has demonstrated very promising clustering accuracy and efficiency and many improvements based on LRR have been proposed in order to get higher clustering accuracy and efficiency.

* Corresponding author.

E-mail address: 519559374@qq.com (S. Yu).

1.1. Related works

In order to improve the robustness and representation ability of the original SSC or LRR based methods, many improvements have been proposed in recent years. These methods try to improve the original models from different aspects. Here we divide these methods into five categories, where methods in each category focus on one particular aspect of the original models.

- (1) To improve the representation by incorporating manifold information. Methods in this category try to improve the representation ability by incorporating manifold information of the data samples. In the original SSC or LRR based methods, the similarity of the representation coefficients or features is not directly related to the similarity of the data samples. In order to preserve the similarity of the representation coefficients or features, the manifold information of the data samples can be incorporated into the objective function. A Laplacian regularized term can be added in order to make the representation coefficients or features resemble the relationship of data samples. In [36], a Laplacian regularized term containing the representation coefficients of the LRR model is added. In [37,38], two Laplacian regularized terms are added. The first term regularizes the right representation coefficients while the second term regularizes the projected data samples using the left representation matrix. Similarly, in [39], two Laplacian regularized terms are added. The first term regularizes the right representation coefficients while the second term regularizes the left representation coefficients. By considering the manifold information, the representation ability of the original model is improved and higher subspace segmentation accuracy can be obtained.
- (2) To improve the representation by simultaneously considering the low rankness and sparsity. Methods in this category try to improve the representation ability by jointly exploring the low rankness and sparsity of the representation. In the original model, only one aspect of the representation is considered which is either low rankness or sparsity. In fact, these two properties can be simultaneously considered when discovering the subspace structure. In [40], a low rank and sparse principal feature coding formulation is proposed that decomposes given data into a component part that encodes low rank sparse principal features and a noise fitting error part. In [41], besides learning a low rank component encoding principal features, the method also learns a similarity-preserving sparse projection for extracting salient features. In the above models, the low rankness and sparsity of the underlying subspaces is jointly considered when processing the data samples.
- (3) To improve the representation by using nonconvex nonsmooth approximation to the rank function. Methods in this category try to improve the representation ability by using nonconvex nonsmooth approximation to the rank function of a matrix. In the original model, the rank function is approximated by the convex nuclear norm. In order to make more accurate approximation of the rank function, nonconvex nonsmooth functions can also be used. In [42], the authors generalize the nonconvex surrogates of l_0 -norm of vectors to matrix rank functions. The corresponding objective function can be solved by iteratively reweighted nuclear norm algorithm. Since nuclear norm is a loose approximation of the rank function, using nonconvex approximations can achieve better accuracy of the representation coefficients. In [43], the authors use matrix γ -norm and min-max concave penalty norm to approximate the l_0 -norm and rank of a matrix. The method proposed in [43] also incorpo-

rates graph construction with low rank learning, which can simultaneously learn the low rank coding coefficients and the similarity graph.

- (4) To improve the representation by considering hidden effects of the data samples. In the original model, the data samples are assumed to be sufficiently sampled. When the data samples are insufficient, the hidden effects of the unobserved data samples should be considered. In [44], the authors proposed latent low rank representation to represent the observed data samples as linear combinations of both observed data samples and unobserved data samples. In [38–40], besides considering the hidden effects of the data samples, the manifold information is also incorporated so better representation ability can be obtained. In this paper, we also consider the hidden effects of the data samples. In [45], the latent low rank representation model is applied to the transfer learning scenario, where the hidden effects brought by missing modality in the training data are considered.
- (5) To improve the representation by learning a dictionary. In the original model, the data samples are directly used as the dictionary to represent the data samples. In practical, data samples could be contaminated by noises or outliers. Learning a dictionary rather than directly using the data samples themselves could improve the accuracy of the representation. In [46], a non-negative dictionary is learned during the optimization and better clustering results are obtained. Similarly, in [47], a dictionary using samples from source domain and target domain is learned to realize transfer learning. However, the objective function will become nonconvex when the dictionary is not fixed and need to be optimized so the effect of dictionary learning needs to be further investigated in the subspace clustering framework.

1.2. The motivation of the paper

1.2.1. Using latent low rank representation when the data samples are insufficient

One improvement of LRR is called latent low rank representation (LLRR) [44]. Latent low rank representation aims at solving the problem when the data samples are insufficient. The observed data samples are represented by both the observed and unobserved data. The effects of the hidden data can be approximately recovered by solving a nuclear norm minimization problem. The objective function of LLRR has two representation matrices which need to be solved. Nuclear norm minimization of the two matrices is to be achieved and the objective function is solved via alternating direction method of multipliers. Besides its capability of dealing with the situation when the data samples are insufficient, LLRR can also be used as a feature extraction method. One of the matrices obtained can be used a transformation matrix to extract salient features from new data samples. The clustering accuracy of LLRR is higher than that of the original LRR. Two singular value decompositions (SVD) have to be performed in each iteration during the optimization procedure, thus causing the computational complexity of LLRR being very high.

1.2.2. Using Frobenius norm instead of nuclear norm

In the original LRR model, the nuclear norm of the representation matrix is to be minimized in the objective function. Nuclear norm has been proven to be an effective convex surrogate of the rank function. During the optimization procedure of LRR model, a singular value decomposition must be performed in every iteration. Doing singular value decomposition is time consuming and limits the ability of the algorithm to process large scale data sets. In order to solve this problem, Frobenius norm is proposed to replace the nuclear norm [48]. Frobenius norm minimization based

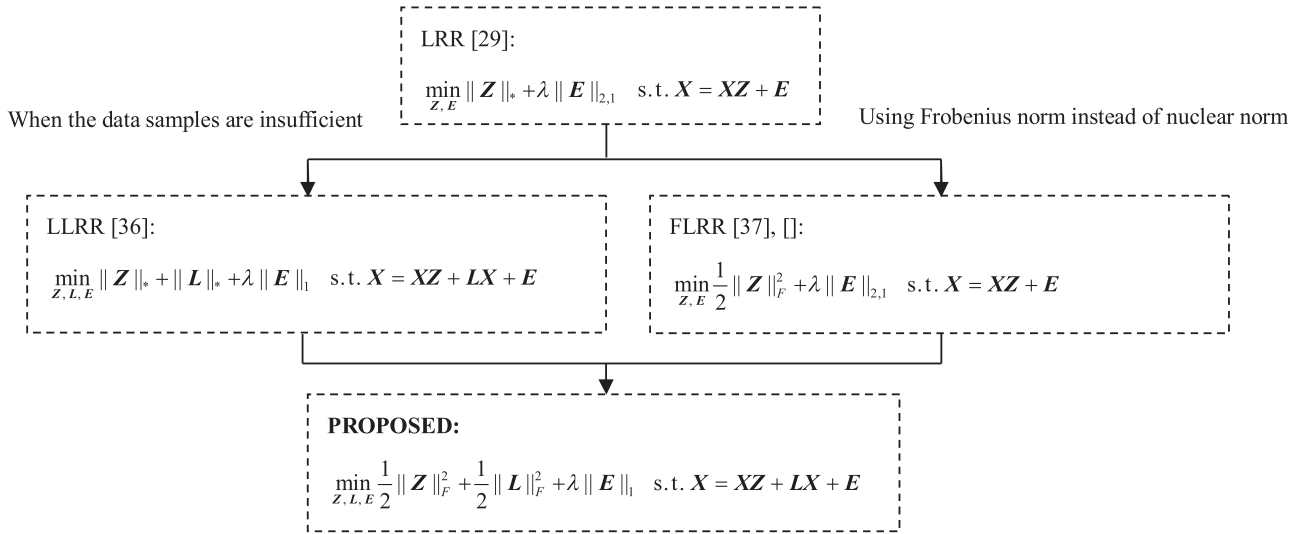


Fig. 1. The relationship between the proposed model and previously proposed subspace clustering models.

low rank representation has several advantages compared with nuclear norm minimization based low rank representation. Not only the computational complexity is reduced significantly but the clustering accuracy is also improved. Theoretical and experimental evidence has been presented to demonstrate the effectiveness of the Frobenius norm as another convex surrogate of the rank function. In [49], the least squares regression is proposed and Frobenius norm is used to regularize the representation matrix. Theoretical analyses are given in [49] which state that the Frobenius norm satisfies the enforced block diagonal conditions. Using Frobenius norm has another advantage that the coefficients of a group of correlated data are approximately equal. Some nuclear norm minimization based subspace clustering algorithms have been improved by using Frobenius norm instead of nuclear norm. One such example is the improvement of low rank subspace clustering [50]. Replacing nuclear norm by Frobenius norm in low rank subspace clustering leads to a new algorithm called efficient dense subspace clustering [51] and its clustering performance is better than that of the low rank subspace clustering. In [52,53], the authors proposed a robust subspace clustering algorithm via thresholding ridge regression. l_2 -norm is used to regularize the representation coefficients in order to have closed form solutions. In [54], theoretical analysis is provided to explain the working mechanism of the Frobenius norm based subspace clustering methods. In that paper, two situations are analyzed. When the dictionary can provide enough representative capacity, Frobenius norm based representation (FNR) is exactly the nuclear norm based representation (NNR). Otherwise, FNR and NNR are two solutions on the column space of the dictionary. In the algorithm proposed in this paper, the latent low rank representation and Frobenius norm representation is combined.

Based on the above two improvements of the original LRR model, in this paper a new low rank representation model is proposed. The proposed model combines the original LLRR model and Frobenius norm minimization based low rank representation into a unified framework. The nuclear norm in the original LLRR model is replaced by the Frobenius norm and the resulting model is called Frobenius norm minimization based LLRR (FLLRR). The proposed model can be solved by the ADMM method. The relationship between the proposed model and the previous proposed models is shown in Fig. 1.

The organization of the paper is given as follows. Section 2 introduces the Frobenius norm minimization based low rank representation.

Section 3 introduces the LLRR model and its optimization procedure. In Section 4, the proposed FLLRR model together with its optimization algorithm is presented. Section 5 gives experimental results and analysis followed by some concluding remarks in Section 6.

2. Frobenius norm minimization based low rank representation

In the original LRR model, nuclear norm has been used a convex surrogate of the rank function. Through theoretical and experimental studies, it has been shown that nuclear norm is not the only convex surrogate of the rank function. Frobenius norm can also be used as a convex surrogate of the rank function. By replacing the nuclear norm in the original LRR model, the Frobenius norm minimization based low rank representation (FLRR) is obtained. The objective function of the FLRR model is given as follows

$$\min_{Z, E} \frac{1}{2} \|Z\|_F^2 + \lambda \|E\|_{2,1} \quad \text{s.t. } X = XZ + E \quad (1)$$

where $X \in \mathbb{R}^{p \times n}$ is the data matrix, $Z \in \mathbb{R}^{n \times n}$ is the representation matrix and $E \in \mathbb{R}^{p \times n}$ is the error matrix. Notice the similarity and difference between the objective function of the FLRR model and the original LRR model. The term $\|Z\|_*$ containing nuclear norm has been replaced by the term $\frac{1}{2} \|Z\|_F^2$ containing Frobenius norm. The objective function of FLRR can also be solved by using the ADMM method and there is no need to add another equality constraint containing an auxiliary matrix. The objective function in Eq. (1) can be rewritten as

$$\min_{Z, E} \frac{1}{2} \|Z\|_F^2 + \lambda \|E\|_{2,1} + \text{tr}(Y^T (X - XZ - E)) + \frac{\mu}{2} \|X - XZ - E\|_F^2 \quad (2)$$

where Y is the Lagrange multiplier. This optimization problem can be solved by fixing the other variables and optimize it with only one variable.

3. Latent low rank representation subspace clustering algorithm

Latent low rank representation is proposed to solve the problem when the data samples are insufficient. In the LLRR model, the observed data samples are represented as linear combinations

of both the observed data samples and the unobserved data samples. The objective function of the LLRR model is given as follows

$$\min_{\mathbf{Z}, \mathbf{L}, \mathbf{E}} \|\mathbf{Z}\|_* + \|\mathbf{L}\|_* + \lambda \|\mathbf{E}\|_1 \text{ s.t. } \mathbf{X} = \mathbf{XZ} + \mathbf{LX} + \mathbf{E} \quad (3)$$

where $\mathbf{X} \in \mathbb{R}^{p \times n}$ is the data matrix, $\mathbf{Z} \in \mathbb{R}^{n \times n}$ and $\mathbf{L} \in \mathbb{R}^{p \times p}$ are the representation matrices and $\mathbf{E} \in \mathbb{R}^{p \times n}$ is the error matrix. The term $\mathbf{XZ}$ can be seen as the linear combinations of the observed data samples while the term $\mathbf{LX}$ can be seen as the linear combinations of the unobserved data samples. The effect of unobserved data samples is considered in LLRR so that the representation model is better than the original LRR model. The objective function of the LLRR model can be solved by the ADMM method. The objective function in Eq. (3) can be rewritten as

$$\min_{\mathbf{Z}, \mathbf{L}, \mathbf{J}, \mathbf{S}, \mathbf{E}} \|\mathbf{J}\|_* + \|\mathbf{S}\|_* + \lambda \|\mathbf{E}\|_1 \text{ s.t. } \mathbf{X} = \mathbf{XZ} + \mathbf{LX} + \mathbf{E}, \mathbf{Z} = \mathbf{J}, \mathbf{L} = \mathbf{S} \quad (4)$$

where the matrices $\mathbf{J}$ and $\mathbf{S}$ are auxiliary variables. The above constraint optimization problem is then transformed into an unconstrained optimization problem by adding Lagrange multipliers. The resulting optimization problem is given as follows

$$\begin{aligned} \min_{\mathbf{Z}, \mathbf{L}, \mathbf{J}, \mathbf{S}, \mathbf{E}} & \|\mathbf{J}\|_* + \|\mathbf{S}\|_* + \lambda \|\mathbf{E}\|_1 + \text{tr}(\mathbf{Y}_1^T (\mathbf{X} - \mathbf{XZ} - \mathbf{LX} - \mathbf{E})) \\ & + \text{tr}(\mathbf{Y}_2^T (\mathbf{Z} - \mathbf{J})) + \text{tr}(\mathbf{Y}_3^T (\mathbf{L} - \mathbf{S})) + \frac{\mu}{2} (\|\mathbf{X} - \mathbf{XZ} - \mathbf{LX} - \mathbf{E}\|_F^2 \\ & + \|\mathbf{Z} - \mathbf{J}\|_F^2 + \|\mathbf{L} - \mathbf{S}\|_F^2) \end{aligned} \quad (5)$$

where $\mathbf{Y}_1$, $\mathbf{Y}_2$ and $\mathbf{Y}_3$ are Lagrange multipliers. The above objective function can be solved by fixing the other variables and optimize it with only one variable.

4. Frobenius norm minimization based latent low rank representation

Comparing the optimization algorithm of the FLRR model and that of the original LRR model, we can find that there are two major differences. The first is that no auxiliary variable is introduced in the objective function of the FLRR model. In the objective function of the original LRR model, an auxiliary variable has to be introduced and a corresponding equality constraint also has to be

introduced. The reason for introducing auxiliary variable is to obtain closed form solution in the nuclear norm minimization procedure. Without introducing any auxiliary variable, the optimization algorithm of the FLRR model is simpler. The second is that a SVD has to be performed in each iteration of the optimization algorithm of the original LRR model while there is no such operation in the optimization algorithm of the FLRR model. Doing SVD is very time consuming and wastes a lot of computational resources. These two differences make the FLRR model be a better subspace clustering model than the original LRR model. If the nuclear norm in the original LLRR model is replaced by the Frobenius norm, the SVD procedure in the optimization algorithm of the LLRR model will no longer be needed. Besides, the auxiliary variables will not be needed either. The objective function of Frobenius norm minimization based low rank representation is given as follows

$$\min_{\mathbf{Z}, \mathbf{L}, \mathbf{E}} \frac{1}{2} \|\mathbf{Z}\|_F^2 + \frac{1}{2} \|\mathbf{L}\|_F^2 + \lambda \|\mathbf{E}\|_1 \text{ s.t. } \mathbf{X} = \mathbf{XZ} + \mathbf{LX} + \mathbf{E} \quad (6)$$

The above constraint optimization problem can be transformed to an unconstrained optimization problem by adding Lagrange multipliers. The resulting objective function is given as follows

$$\begin{aligned} \min_{\mathbf{Z}, \mathbf{L}, \mathbf{E}} & \frac{1}{2} \|\mathbf{Z}\|_F^2 + \frac{1}{2} \|\mathbf{L}\|_F^2 + \lambda \|\mathbf{E}\|_1 + \text{tr}(\mathbf{Y}^T (\mathbf{X} - \mathbf{XZ} - \mathbf{LX} - \mathbf{E})) \\ & + \frac{\mu}{2} \|\mathbf{X} - \mathbf{XZ} - \mathbf{LX} - \mathbf{E}\|_F^2 \end{aligned} \quad (7)$$

where $\mathbf{Y}$ is the Lagrange multiplier. Note that here no auxiliary variables are introduced into the objective function. The above objective function can be solved by fixing the other variables and optimize it with only one variable. The optimization algorithm for solving the FLLRR model is presented in Algorithm 1. In the optimization procedure, the SVD operations are omitted thus reducing the computational complexity of the original LLRR model in a great extent. The time complexity of the LLRR model and FLLRR model is analyzed as follows. First we want to analyze the time complexity of the LLRR model. In each iteration of the LLRR model, two SVDs have to be performed which has the time complexity $O(p^3 + n^3)$. Besides, two matrix inversions have to be performed which has the time complexity $O(p^3 + n^3)$. In each iteration of the FLLRR model, no SVD has to be performed. The two matrix inversions have the time complexity $O(p^3 + n^3)$. Compared to the LLRR model, the time complexity of the FLLRR model is reduced to 1/2 of the LLRR model.

Algorithm 1

The optimization algorithm for solving FLLRR.

Input: data matrix $\mathbf{X}$, parameter λ
Initialize: $\mathbf{Z}_{(0)} = \mathbf{0}$, $\mathbf{L}_{(0)} = \mathbf{0}$, $\mathbf{E}_{(0)} = \mathbf{0}$, $\mathbf{Y}_{(0)} = \mathbf{0}$, $\mu = 10^{-6}$, $\mu_{\max} = 10^6$, $\rho = 1.1$, $\varepsilon = 10^{-8}$ and $k = 0$
while not converged **do**
 1. fix the others and update $\mathbf{Z}$ by

$$\mathbf{Z}_{(k+1)} = (\mathbf{I} + \mu \mathbf{X}^T \mathbf{X})^{-1} \mu \mathbf{X}^T \left(\mathbf{X} - \mathbf{L}_{(k)} \mathbf{X} - \mathbf{E}_{(k)} + \frac{\mathbf{Y}_{(k)}}{\mu} \right)$$

 2. fix the others and update $\mathbf{L}$ by

$$\mathbf{L}_{(k+1)} = \mu \left(\mathbf{X} - \mathbf{XZ}_{(k+1)} - \mathbf{E}_{(k)} + \frac{\mathbf{Y}_{(k)}}{\mu} \right) \mathbf{X}^T (\mathbf{I} + \mu \mathbf{X}^T \mathbf{X})^{-1}$$

 3. fix the others and update $\mathbf{E}$ by

$$\mathbf{E}_{(k+1)} = \arg \min_{\mathbf{E}} \frac{\lambda}{\mu} \|\mathbf{E}\|_1 + \frac{1}{2} \left\| \mathbf{E} - \left(\mathbf{X} - \mathbf{XZ}_{(k+1)} - \mathbf{L}_{(k+1)} \mathbf{X} + \frac{\mathbf{Y}_{(k)}}{\mu} \right) \right\|_F^2$$

$$\mathbf{E}_{(k+1)} = \mathcal{S}_{\frac{\lambda}{\mu}} \left(\mathbf{X} - \mathbf{XZ}_{(k+1)} - \mathbf{L}_{(k+1)} \mathbf{X} + \frac{\mathbf{Y}_{(k)}}{\mu} \right)$$

 4. update the multipliers

$$\mathbf{Y}_{(k+1)} = \mathbf{Y}_{(k)} + \mu (\mathbf{X} - \mathbf{XZ}_{(k+1)} - \mathbf{L}_{(k+1)} \mathbf{X} - \mathbf{E}_{(k+1)})$$

 5. update the parameter μ by $\mu = \min(\rho \mu, \mu_{\max})$
 6. check the convergence conditions:

$$\|\mathbf{X} - \mathbf{XZ}_{(k+1)} - \mathbf{L}_{(k+1)} \mathbf{X} - \mathbf{E}_{(k+1)}\|_{\infty} < \varepsilon$$

end while
Output: Matrix $\mathbf{Z} = \mathbf{Z}_{(k+1)}$, $\mathbf{L} = \mathbf{L}_{(k+1)}$ and $\mathbf{E} = \mathbf{E}_{(k+1)}$

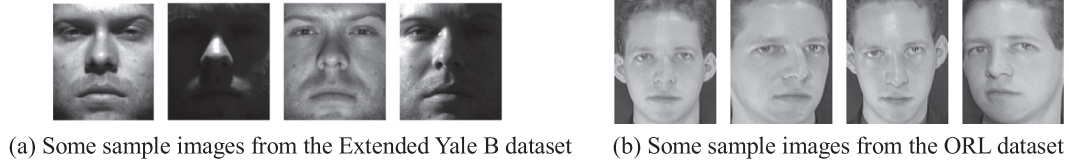


Fig. 2. Some sample images from the face datasets.

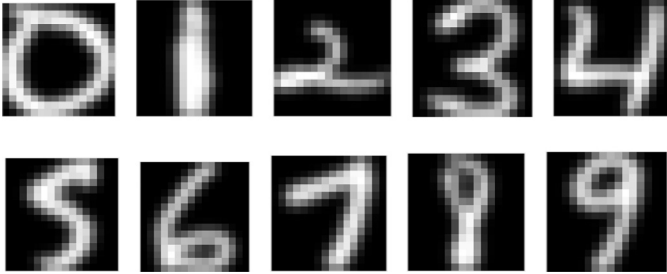


Fig. 3. Some data samples from the USPS dataset.



Fig. 4. Some data samples from the MNIST dataset.

Here we want to give a brief theoretical analysis about the relationship between matrix norms. Matrix norms are a natural generalization of the vector norms. The three most important vector norms used in the compressed sensing framework are the l_1 , l_2 , and l_∞ norms. Generalizing these norms to the matrix case, we get the nuclear norm, Frobenius norm and operator norm. The nuclear norm of a matrix can be seen as the l_1 norm of the singular values of the matrix. Its definition is given as follows:

$$\|\mathbf{X}\|_* := \sum_{i=1}^r \sigma_i(\mathbf{X}) \quad (8)$$

where r is the rank of the matrix $\mathbf{X}$ and $\sigma_i(\mathbf{X})$ is the i th singular value of the matrix.

The Frobenius norm of a matrix can be seen as the l_2 norm of the singular values of the matrix. Its definition can be given as follows:

$$\|\mathbf{X}\|_F := \left(\sum_{i=1}^r \sigma_i^2(\mathbf{X}) \right)^{\frac{1}{2}} \quad (9)$$

The operator norm of a matrix can be seen as the l_∞ norm of the singular values of the matrix. Its definition can be written as follows:

$$\|\mathbf{X}\| := \sigma_1(\mathbf{X}) \quad (10)$$

Actually, these three matrix norms are closely related to each other. The relationship between these matrix norms is as follows [55]:

$$\|\mathbf{X}\| \leq \|\mathbf{X}\|_F \leq \|\mathbf{X}\|_* \leq \sqrt{r} \|\mathbf{X}\|_F \leq r \|\mathbf{X}\| \quad (11)$$

The proof of the above inequality is given in the Appendix.

It can be proved that the convex envelop of $\text{rank}(\mathbf{X})$ on the set $\{\mathbf{X} \in \mathbb{R}^{m \times n} : \|\mathbf{X}\| \leq 1\}$ is the nuclear norm $\|\mathbf{X}\|_*$ [55]. Using nuclear norm as a convex surrogate of the rank function is widely used in the low rank based subspace clustering methods. By studying the relationship between the nuclear norm and the Frobenius norm, we can see that Frobenius norm can also be regarded as a convex surrogate of the rank function. With the above inequality, it can be said that these three matrix norms are equivalent to each other to some extent. Minimizing one of the matrix norms will also minimize the other two matrix norms. From inequality (4), we can see that the Frobenius norm is an upper bound and as well as a lower bound of the nuclear norm. So using the Frobenius norm instead of the nuclear norm in the algorithm LatLRR is reasonable. After the optimization procedure, we get a low rank matrix $\mathbf{Z}$ and $\mathbf{L}$. Actually, in the proposed algorithm, we use the square of the Frobenius norm for computational convenience.

5. Experimental results and analysis

In this section the performance of the proposed model as well as several other subspace clustering models are evaluated using four real-world datasets. The performance of the proposed model is compared with the best state-of-the-art subspace clustering models including LSA [23], SLBF [24], LLMC [26], SCC [27], MSL [14], SSC [28], LRR [29], LLRR [44] and rLRR [37]. LSA, SLBF, LLMC can be considered as local spectral clustering-based approaches and SCC, SSC, LRR, LLRR, rLRR can be considered as global spectral clustering-based approaches. MSL can be considered as a kind of iterative statistical approach.

Datasets. Four real-world datasets are considered here. The first dataset is the Extended Yale B dataset [56], which consists of face images of 38 human subjects, where images of each subject lie in a low-dimensional subspace. There are 64 frontal face images for each subject acquired under various lighting conditions. The second dataset is the ORL dataset, which consists of face images of 40 human subjects. There are 10 frontal face images for each subject acquired under different situations. Some sample images from the face datasets are shown in Fig. 2. The third dataset is the handwritten digit dataset which includes the USPS dataset and MNIST dataset. The USPS dataset consists of a training set with 7291 images and a test set with 2007 images. The digits have been size-normalized and centered in a fixed-size image. The MNIST dataset of handwritten digits has a training set of 60,000 examples and a test set of 10,000 examples. The digits have also been size-normalized. The images of each digit also lie in a low-dimensional subspace. Some examples from the USPS and MNIST dataset are shown in Figs. 3 and Fig. 4 respectively. The fourth dataset contains many river segments extracted from a set of remote sensing images. Each river segment is of size 60×60 . After the river segments are extracted, the rotation operation is done to some of the river segments to make the river segments more aligned. The river segments do not have ground truth classification but they can be considered as drawn from a union of linear subspaces and thus subspace clustering models can be applied. The remote sensing image containing the rivers as well as some of the river segments extracted from that image are shown in Fig. 5. The details of all datasets used in the experiment are given in Table 1.

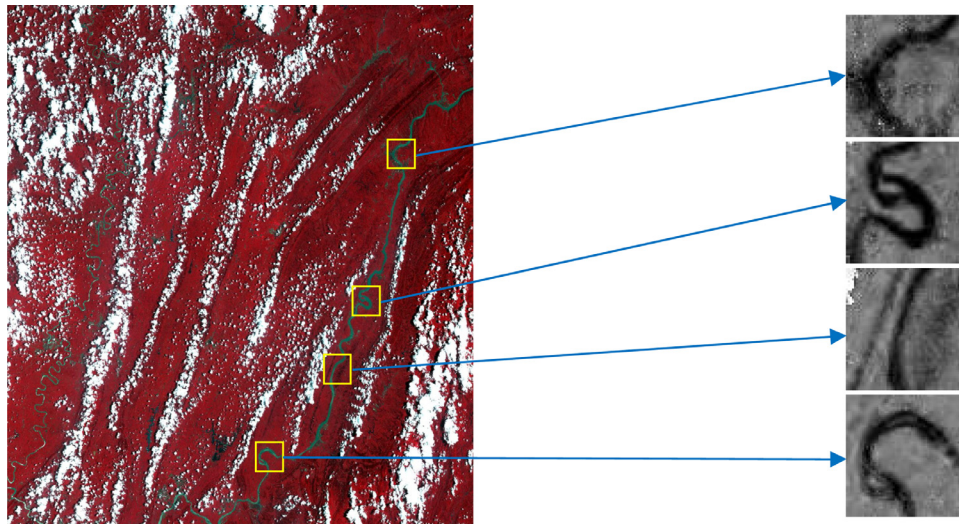


Fig. 5. The remote sensing image and some of the river segments extracted from that image.

Table 1
Details of the datasets used for clustering experiments in this work.

Dataset	Training samples	Test samples	Dimension	Classes
Extended Yale B	/	2432	192×168	38
ORL	/	400	92×112	40
USPS	7291	2007	16×16	10
MNIST	60,000	10,000	28×28	10
River	/	128	60×60	/

Parameter determination by different approaches. The parameters determined by different approaches for face clustering and handwritten digit clustering are shown in Table 2 for clarity. For the approach LSA, K represents the nearest neighbors for a data sample to calculate the local linear subspace and d represents the subspace dimension of the local linear subspace. For the approach SLBF, K_{start} and K_{step} represent the start size and step size of the local neighborhood selection to calculate the local linear subspace and d represents the subspace dimension of the local linear subspace. For the approach LLMC, K represents the number of nearest neighbors for representing a data sample. For the approach SCC, d represents the subspace dimension of the local linear subspace. For the approach LRR, λ is the weighting parameter for the sparse outlier term, its value is chosen following the strategies provided by [57]. For the approach SSC, λ_z is the weighting parameter for the noise term and λ_e is the weighting parameter for the sparse outlier term. For the approach LLRR, λ is the weighting parameter for the sparse outlier term. For the approach rLRR, γ_l is the weighting parameter for the Laplacian regularized term and γ_E is the weighting parameter

for the sparse outlier term. For the approach FLLRR, λ is the weighting parameter for the sparse outlier term. For river segments clustering, we only compare the approach LLRR and FLLRR.

5.1. Face clustering

The face images in the Extended Yale B dataset are acquired under a fixed pose with varying lighting conditions. It has been shown that, under the Lambertian assumption, images of a subject with a fixed pose and varying lighting conditions lie close to a linear subspace of dimension 9 [28]. Thus, the collection of face images of multiple subjects lies close to a union of 9D subspaces [28,29].

5.1.1. Clustering accuracy comparison

The clustering performance of FLLRR as well as the state-of-the-art approaches on the Extended Yale B dataset are evaluated and compared in this section. As done in [28], we also downsample the images to 48×42 pixels to reduce the computational cost and the memory requirements of all algorithms. Thus each data point is a 2016D vector, hence $p=2016$. Similarly, we downsample the images from the ORL dataset to 23×28 .

The 38 subjects are divided into 4 groups as done in [28], where the first 3 groups correspond to subjects 1 to 10, 11 to 20, 21 to 30, and the fourth group corresponds to subjects 31 to 38. The four groups are referred to as groups 1–4 respectively. The authors in [28] consider all choices of $n \in \{2, 3, 5, 8, 10\}$ subjects for each of the first 3 groups and $n \in \{2, 3, 5, 8\}$ for the last group. In this experiment, only the case $n=10$ is considered for each of the

Table 2
The parameters selected for different approaches.

Approach	Face clustering	Handwritten digit clustering	
		USPS	MNIST
LSA [23]	$K=15, d=9$	$K=5, d=9$	$K=5, d=9$
SLBF [24]	$K_{start}=10, K_{step}=5, d=9$	$K_{start}=2, K_{step}=2, d=9$	$K_{start}=2, K_{step}=2, d=9$
LLMC [26]	$K=10$	$K=5$	$K=5$
SCC [27]	$d=9$, affine version	$d=9$, affine version	$d=9$, affine version
MSL [14]	$d=9$	$d=9$	$d=9$
LRR [29]	$\lambda = 1/\sqrt{\log(n)}$ [57]	$\lambda = 1/\sqrt{\log(n)}$ [57]	$\lambda = 1/\sqrt{\log(n)}$ [57]
SSC [28]	$\lambda_z = 800/\mu_z, \lambda_e = 20/\mu_e$	$\lambda_z = 800/\mu_z, \lambda_e = 20/\mu_e$	$\lambda_z = 800/\mu_z, \lambda_e = 20/\mu_e$
LLRR [44]	$\lambda = 5 \times 10^{-4}$	$\lambda = 10^{-3}$	$\lambda = 10^{-3}$
rLRR [37]	$\gamma_l = 5 \times 10^{-4}, \gamma_E = 5 \times 10^{-4}$	$\gamma_l = 10^{-3}, \gamma_E = 10^{-3}$	$\gamma_l = 10^{-3}, \gamma_E = 10^{-3}$
FLLRR	$\lambda = 5 \times 10^{-4}$	$\lambda = 10^{-3}$	$\lambda = 10^{-3}$

Table 3

Clustering accuracy (%) of different approaches on the Extended Yale B dataset without preprocessing the data.

Approach	Group 1	Group 2	Group 3	Group 4
LSA [23]	25.00	25.68	25.31	30.44
SLBF [24]	37.26	38.36	34.06	40.12
LLMC [26]	40.09	43.18	32.34	38.54
SCC [27]	29.25	20.39	30.31	30.63
MSL [14]	64.94	59.55	71.41	42.69
SSC [28]	83.18	89.89	94.22	78.06
LRR [29]	95.44	85.23	96.72	90.32
LLRR [36]	94.18	97.75	98.13	93.68
rLRR [37]	94.65	97.59	98.13	94.27
FLLRR	97.48	88.93	99.06	97.63

Table 4

Clustering accuracy (%) of different approaches on the ORL dataset.

Approach	Group 1	Group 2	Group 3	Group 4
SSC [28]	99.00	71.00	88.00	70.00
LRR [29]	77.00	79.00	64.00	66.00
LLRR [36]	97.00	84.00	70.00	90.00
rLRR [37]	97.00	81.00	70.00	81.00
FLLRR	97.00	86.00	84.00	88.00

first 3 groups and $n=8$ for the last group since this is the most challenging case for face clustering. In [28], the authors reported clustering results under 3 different settings: the first is applying RPCA [2] separately on each subject and then cluster the data samples, the second is applying RPCA simultaneously on all subjects and cluster the data samples and the third is directly clustering the original data samples. Here the first and second settings are not considered and all the clustering approaches are directly applied to the original data samples. The clustering accuracy results using the Extended Yale B dataset are presented in Table 3. The 40 subjects in the ORL dataset are also divided into 4 groups. The clustering accuracy results using the ORL dataset are presented in Table 4. We only report the results of global spectral clustering methods in Table 4.

From the table we can make the following conclusions. Local spectral clustering-based approaches such as LSA, SLBF and LLMC obtain a relatively low clustering accuracy on the dataset. Local spectral clustering-based approaches rely on estimating the local linear subspace of a data sample using the K nearest neighbors of that data sample. As for the Extended Yale B dataset, there are a relatively large number of data samples near the intersection of subspaces; the local neighborhood of a sample must contain a lot

of samples from the other subspaces. The local subspace estimated is not accurate and the similarity built based on this local subspace is not accurate either.

Global spectral clustering-based approach such as SCC also obtains relatively low clustering accuracy. This kind of method does not need to estimate the local linear subspace. However, the noises and outliers existed in the dataset will greatly affect the estimation of the polar curvature and thus the similarity built based on this polar curvature is not accurate.

Iterative statistical approach such as MSL obtains relatively low clustering accuracy. This method needs an initial clustering of the dataset. This method can easily get stuck in local minimum around the initial segmentation and the clustering accuracy is thus highly dependent on the initial clustering.

Global spectral clustering-based approaches such as LRR and SSC obtain more accurate clustering results while LRR gets higher clustering accuracy than SSC. LRR and SSC both exploit the global structure of the dataset and handle noises and outliers explicitly. The similarity matrix obtained is more accurate.

The performance of LLRR is even better than LRR and rLRR in most cases. LLRR solves the problem when the data samples are insufficient and considers the effect of the unobserved data samples. The representation model of LLRR is better than that of the original LRR model. The representation matrix obtained is more accurate than that of the original LLR model. FLLRR gets the highest clustering accuracy compared with other subspace clustering approaches. By using the Frobenius norm instead of the nuclear norm, the representation matrix obtained is a low rank matrix. This matrix is even better than that of the original LLRR model. This fact further proves that the Frobenius norm is a better convex surrogate than the nuclear norm. To clearly show the difference between the proposed model and the LLRR model, we show the matrix Z obtained by the two algorithms using the data samples in group 1 in Fig. 6. From Fig. 6 we can find that those coefficients in the matrix Z within the diagonal blocks obtained by the FLLRR model are larger and those coefficients in off-diagonal blocks are smaller. This property is beneficial to clustering. From these experimental results, it can be seen that the proposed algorithm is better than the nuclear norm based low rank subspace clustering methods. As pointed out by Lu et al. [49], the clustering capability of low rank based subspace clustering method may not come from the low rankness of the representation matrix. In [49], the authors propose another criterion which is enforced block diagonal condition and prove that matrix solved obeying this condition can be used for subspace clustering. Nuclear norm is a special case of this condition. Frobenius norm is another special case of

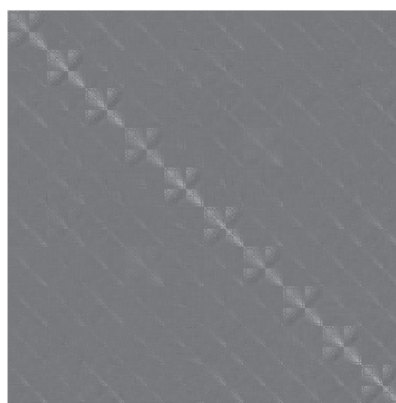
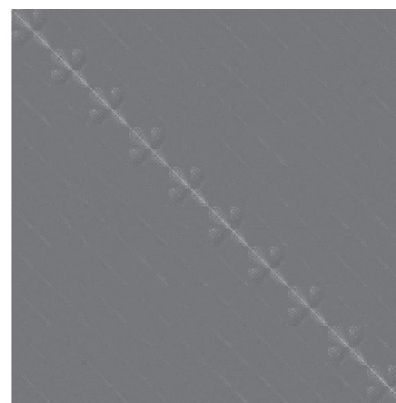
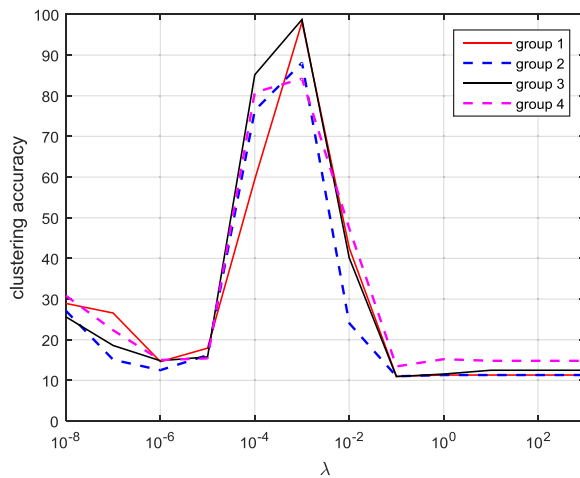
(a) matrix Z obtained by LLRR(b) matrix Z obtained by FLLRR**Fig. 6.** The matrix Z obtained by the two algorithms.

Table 5

The running time of LLRR and FLLRR (s).

Approach	Group 1	Group 2	Group 3	Group 4
LLRR [36]	433.7	434.7	401.1	384.9
FLLRR	280.4	282.8	279.1	236.1

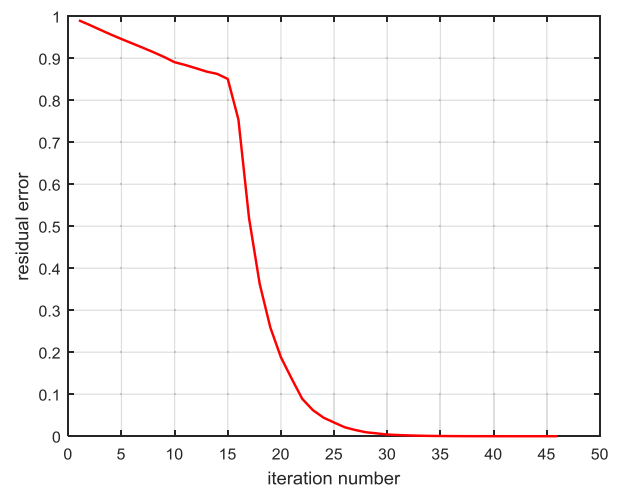
**Fig. 7.** The relationship between λ and the clustering accuracies.

this condition. Thus matrix solved based on Frobenius norm minimization can also be used for subspace clustering. However, there is another advantage about Frobenius norm minimization based method. The coefficients of highly correlated data samples are almost the same. It means that the intra-cluster difference of the coefficients is smaller and the inter-cluster difference of the coefficients is larger. The clustering accuracy can be improved when the coefficients of the data samples within the same cluster are similar to each other. This grouping effect cannot be obtained by using nuclear norm minimization. So Frobenius norm based latent low rank representation can get better subspace clustering accuracy than nuclear norm based methods such as LatLRR or rLRR.

In order to compare the efficiency of different approaches, the running time of LLRR and FLLRR is recorded. The environment for running the clustering approaches is: MATLAB R2015a, Intel (R) Core (TM) i5-4460S CPU @ 2.90GHz, 8 GB RAM. The running time of LLRR and FLLRR is given in Table 5. From the table we can clearly see that the running time of FLLRR is significantly less than that of LLRR. This is because no SVD operations are needed and no auxiliary variables are introduced in the optimization algorithm of FLLRR.

5.1.2. Parameter sensitivity analysis

We investigate the effects of using different values of parameter λ on the performance of our proposed model. The values of λ are chosen from $\{10^{-8}, 10^{-7}, 10^{-6}, 10^{-5}, 10^{-4}, 10^{-3}, 10^{-2}, 10^{-1}, 1, 10^2, 10^3\}$ and the corresponding clustering accuracies are recorded. The experiments are done on the four groups used in the previous section. The relationship between λ and the clustering accuracies is shown in Fig. 7. From Fig. 7 we can observe the following facts. First, the clustering accuracy is low when using too large or too small values of λ . Second, the highest clustering accuracy is obtained when the value of λ is around 10^{-3} . Third, although different groups are used as data samples, the trends are highly similar to each other. The value of λ should be chosen according to the noise energy presented in the data samples. Too large or too small values of λ will cause negative effect to the proposed model. In

**Fig. 8.** The relationship between the residual error and the iteration number.

the handwritten digit clustering experiments to be presented, the values of λ are chosen appropriately.

5.1.3. Convergence analysis

In this subsection we mainly analyze the convergence property of the proposed model. Since the proposed model only uses one Lagrange multiplier, it can be proved theoretically that the proposed model convergences. The experiments on all the databases also show that the proposed model convergences to the optimal solution. To clearly show the results, we give the relationship between the residual error and the iteration number in Fig. 8. The experiment is done using the data samples in group 1 in the Section 5.1.1. From Fig. 8 we can observe that, the residual error quickly decreases to 0 in a few iterations. In the face clustering experiments, the proposed model usually convergences within 60 iterations. The proposed model has good convergence property.

5.2. Handwritten digit clustering

Handwritten digit clustering refers to the problem of clustering multiple handwritten digits with different rotations and translations. The USPS dataset and MNIST dataset are used as the dataset for evaluating the performance of different subspace clustering approaches. Before presenting the clustering results, we want to make a few discussions about the characteristic of the handwritten digits. Unlike frontal face images, the images of handwritten digits are less aligned. Besides, the information contained in each handwritten image is not as rich as that of the face images since each image can be considered as a binary image. Furthermore, as can be seen from the figures, the difference between the subspaces where the handwritten digits are drawn is not as large as that of the face images. Some of the digits are very similar to the others. These characteristics make the clustering of handwritten digits more difficult than clustering face images. The experiment settings of handwritten digit clustering are as follows. The data samples are chosen from the training samples and testing samples separately. Different number of data samples per class is used. We have done the clustering experiments when the number of data samples per class is 10, 20, 30, 40 and 50. In each clustering experiment, 15 random selections of the data samples from the database are done and the average clustering accuracies and the standard deviations are recorded. The clustering accuracies of different methods on the USPS database and MNIST database are presented in Fig. 9.

From Fig. 9 we can give the following observations. The performances of most algorithms increase as the number of data samples per class is increased. When more data samples are chosen for

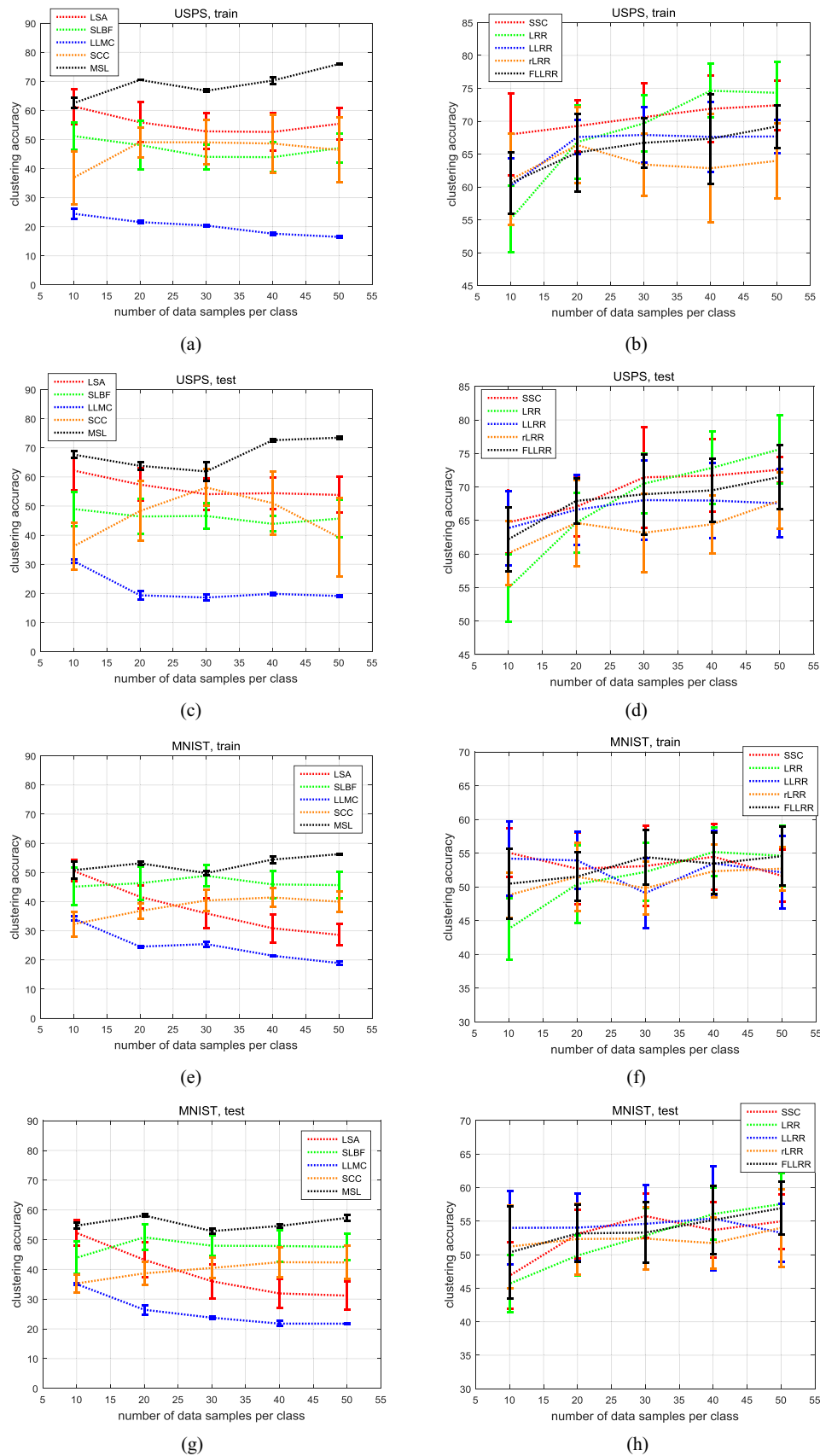


Fig. 9. The performances of different algorithms on the USPS and MNIST database.

Table 6

The Silhouette index of different algorithms when clustering the river segments.

Number of clusters	LLRR	FLLRR
2	−0.461	−0.003
3	−0.091	−0.037
4	−0.080	0.056
5	−0.078	−0.002
6	−0.049	0.051

clustering, more information can be utilized. The algorithms LLMC and LSA are exceptions. The performances of these two algorithms decrease as the number of data samples per class is increased. The performances of the algorithms based on sparse representation and low rank coding are comparable to each other. When the number of data samples per class is small, the performances of the algorithms LLRR, rLRR and FLLRR are better. When the number of data samples per class is large, the performances of the algorithms SSC and LRR are better. The proposed algorithm gets higher clustering accuracies compared to the algorithms LLRR and rLRR when different number of data samples per class is used. The clustering accuracies of the algorithm MSL are very high compared to other methods and the stability of the algorithm is better. The algorithm LLMC always gets the lowest clustering accuracies on all the databases.

5.3. River segment clustering

In this section we have done several experiments using river segments. The images of river segments share some similarity with the images of handwritten digits but the images of river segments are even more difficult to cluster than the images of handwritten digits for the following reasons. First, some of the images of river segments are severely interrupted by noises. Second, some of the images of river segments are blurred and the river can be hardly seen from the images. Third, the difference between the linear subspaces from which the river segments are drawn is smaller which means that all river segments share some kind of similarity with each other. Since the ground truth label information is not known before hand, the clustering accuracy cannot be used. In order to compare the performance of different algorithms quantitatively, the Silhouette index is used. The larger the Silhouette index, the better the clustering results. We choose the number of clusters from {2, 3, 4, 5, 6} and the corresponding Silhouette index is recorded. The parameters of different algorithms are chosen so that the best performance can be obtained. The Silhouette index of different algorithms is shown in Table 6. From Table 6 we can see that, the Silhouette index of FLLRR is higher than that of LLRR. This result further indicates the superiority of the algorithm FLLRR.

6. Conclusions and future research directions

In this paper a new low rank subspace clustering model is proposed. This method combines latent low rank representation and Frobenius norm minimization. Experiments on real data sets such as face images and handwritten digits show the effectiveness of the proposed algorithm and its superiority over the state-of-the-art. The computational complexity of the proposed algorithm is reduced a lot compared to the LLRR model. The proposed algorithm has good convergence property. When clustering face images, the proposed model gets the highest clustering accuracies. When clustering handwritten digit images, the proposed model is better than LLRR model and rLRR model. When clustering river segments, the proposed model can get better clustering results.

There are several possible future research directions we want to point out. There is one parameter to be determined in the LLRR

and FLLRR model. In this paper, this parameter is chosen heuristically. Finding a method to choose the optimal parameter for LLRR and FLLRR will be researched in the future. Latent low rank representation has many improvements. For example, graph Laplacian regularization can be added to the latent low rank representation and the clustering accuracy is improved [37–39]. This regularization may also be added to the FLLRR model and higher clustering accuracy may be obtained. The computational complexity of the FLLRR model is still relatively high. When applying this model to large scale databases, the running time will be long. Recently, an online low-rank representation method is proposed [58] and the computational complexity of the original LRR model is reduced significantly. The clustering accuracy is also improved by using the online version. Similarly, an online implementation of the LLRR model and the proposed model should also be developed to reduce the computational complexity and improve the capability of the algorithms to deal with large scale databases. Finally, more effort should be devoted to research a better clustering algorithm for clustering handwritten digits and river segments.

Acknowledgment

This work is partially supported by the Key Laboratory of Port, Waterway and Sedimentation Engineering of the Ministry of Transport, Nanjing Hydraulic Research Institute; State Key Laboratory of Urban Water Resource and Environment under Grant ES201409; and Key Laboratory of Rivers and Lakes Governance and Flood Protection of Yangtse River Water Conservancy Committee under Grant CKWV2013225/KY.

Appendix

The proof of the inequality (11) is given as follows.

Proof. The first inequality is apparent. Since $\|\mathbf{X}\|_F = (\sum_{i=1}^r \sigma_i^2(\mathbf{X}))^{\frac{1}{2}} \geq (\sigma_1^2(\mathbf{X}))^{\frac{1}{2}} = \sigma_1(\mathbf{X}) = \|\mathbf{X}\|$. To prove the second inequality, we can take square of the both sides and observe that

$$\begin{aligned} \|\mathbf{X}\|_*^2 &= \left(\sum_{i=1}^r \sigma_i(\mathbf{X}) \right)^2 = \sum_{i=1}^r \sigma_i^2(\mathbf{X}) + \sum_{i=1}^r \sum_{j=1, j \neq i}^r \sigma_i(\mathbf{X}) \sigma_j(\mathbf{X}) \\ &\geq \sum_{i=1}^r \sigma_i^2(\mathbf{X}) = \|\mathbf{X}\|_F^2 \end{aligned} \quad (\text{A.1})$$

and using the fact that the singular values of the matrix are all positive. The third inequality can be proved using Cauchy–Schwarz inequality. It can be proved as follows:

$$\begin{aligned} \|\mathbf{X}\|_*^2 &= \left(\sum_{i=1}^r \sigma_i(\mathbf{X}) \right)^2 = \left(\sum_{i=1}^r \sigma_i(\mathbf{X}) \cdot 1 \right)^2 \leq \sum_{i=1}^r \sigma_i^2(\mathbf{X}) \sum_{i=1}^r 1^2 \\ &= r \sum_{i=1}^r \sigma_i^2(\mathbf{X}) = r \|\mathbf{X}\|_F^2 \end{aligned} \quad (\text{A.2})$$

The fourth inequality can be proved as follows:

$$r \|\mathbf{X}\|_F^2 = r \sum_{i=1}^r \sigma_i^2(\mathbf{X}) \leq r \sum_{i=1}^r \sigma_1^2(\mathbf{X}) = r \cdot r \sigma_1^2(\mathbf{X}) = r^2 \sigma_1^2(\mathbf{X}) \quad (\text{A.3})$$

and using the fact that $\sigma_1(\mathbf{X})$ is the largest singular value of the matrix. $\square$

References

- [1] H. Abdi, L.J. Williams, *Principal component analysis*, Wiley Interdiscip. Rev. Comput. Stat. 2 (4) (2010) 433–459.

- [2] E. Candes, X. Li, Y. Ma, J. Wright, Robust principal component analysis, *J. ACM* 58 (3) (2009) 1–37.
- [3] D.D. Lee, H.S. Seung, Learning the parts of objects by nonnegative matrix factorization, *Nature* 401 (6755) (1999) 788–791.
- [4] G. Liu, Z. Lin, X. Tang, Y. Yu, Unsupervised object segmentation with a hybrid graph model, *IEEE Trans. Pattern Anal. Mach. Intell.* 32 (5) (2010) 910–924.
- [5] K. Kanatani, Motion segmentation by subspace separation and model selection, in: *Proceedings of the IEEE International Conference on Computer Vision*, 2, 2001, pp. 586–591.
- [6] W. Hong, J. Wright, K. Huang, Y. Ma, Multiscale hybrid linear models for lossy image representation, *IEEE Trans. Image Process.* 15 (12) (2006) 3655–3671.
- [7] R. Vidal, Y. Ma, S. Sastry, Generalized principal component analysis (GPCA), *IEEE Trans. Pattern Anal. Mach. Intell.* 27 (12) (2005) 1–15.
- [8] S. Rao, R. Tron, R. Vidal, Y. Ma, Motion segmentation in the presence of outlying, incomplete, or corrupted trajectories, *IEEE Trans. Pattern Anal. Mach. Intell.* 32 (10) (2010) 1832–1845.
- [9] A. Yang, J. Wright, Y. Ma, S. Sastry, Unsupervised segmentation of natural images via lossy data compression, *Comput. Vis. Image Underst.* 110 (2) (2008) 212–225.
- [10] C. Zhang, R. Bitmead, Subspace system identification for training-based MIMO channel estimation, *Automatica* 41 (9) (2005) 1623–1632.
- [11] T. Zhang, A. Szlam, G. Lerman, Median k-Flats for hybrid linear modeling with many outliers, *Proceedings of the IEEE International Workshop on Subspace Methods*, 2009.
- [12] P.S. Bradley, O.L. Mangasarian, k-Plane clustering, *J. Global Optim.* 16 (1) (2000) 23–32.
- [13] P. Agarwal, N. Mustafa, k-Means projective clustering, in: *Proceedings of the ACM SIGMOD-SIGACT-SIGART Symposium on Principles of Database Systems*, 2004, pp. 155–165.
- [14] Y. Sugaya, K. Kanatani, Geometric structure of degeneracy for multi-body motion segmentation, in: *Proceedings of the Workshop on Statistical Methods in Video Processing*, 2004.
- [15] Y. Sugaya, K. Kanatani, Multi-stage unsupervised learning for multi-body motion segmentation, *IEICE T. Inf. Syst.* 87 (7) (2004) 1935–1942.
- [16] M. Tipping, C. Bishop, Mixtures of probabilistic principal component analyzers, *Neural Comput.* 11 (2) (1999) 443–482.
- [17] T.E. Boult, L.G. Brown, Factorization-based segmentation of motions, in: *Proceedings of the IEEE Workshop on Motion Understanding*, 1991, pp. 179–186.
- [18] Y. Ma, A. Yang, H. Derksen, R. Fossum, Estimation of subspace arrangements with applications in modeling and segmenting mixed data, *SIAM Rev.* 40 (2008) 413–458.
- [19] J. Costeira, T. Kanade, A multibody factorization method for independently moving objects, *Int. J. Comput. Vision* 29 (3) (1998) 159–179.
- [20] C.W. Gear, Multibody grouping from motion images, *Int. J. Comput. Vision* 29 (2) (1998) 133–150.
- [21] M. Fischler, R. Bolles, Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography, *Commun. ACM* 24 (6) (1981) 381–395.
- [22] Y. Ma, H. Derksen, W. Hong, J. Wright, Segmentation of multivariate mixed data via lossy data coding and compression, *IEEE Trans. Pattern Anal. Mach. Intell.* 29 (9) (2007) 1546–1562.
- [23] J. Yan, M. Pollefeys, A general framework for motion segmentation: independent, articulated, rigid, non-rigid, degenerate and non-degenerate, in: *Proceedings of the European Conference on Computer Vision*, 2006, pp. 94–106.
- [24] T. Zhang, A. Szlam, Y. Wang, G. Lerman, Hybrid linear modeling via local best-fit flats, *Int. J. Comput. Vision* 100 (3) (2012) 1927–1934.
- [25] L. Zelnik-Manor, M. Irani, Degeneracies, dependencies and their implications in multi-body and multi-sequence factorization, in: *Proceedings of the IEEE Conference on Computer Vision Pattern Recognition*, 2, 2003, pp. 287–293.
- [26] A. Goh, R. Vidal, Segmenting motions of different types by unsupervised manifold clustering, in: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2007.
- [27] G. Chen, G. Lerman, Spectral curvature clustering (SCC), *Int. J. Comput. Vision* 81 (3) (2009) 317–330.
- [28] E. Elhamifar, R. Vidal, Sparse subspace clustering: algorithm, theory, and applications, *IEEE Trans. Pattern Anal. Mach. Intell.* 35 (11) (2013) 2765–2781.
- [29] 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 (1) (2013) 171–184.
- [30] J. Wright, A. Yang, A. Ganesh, S. Sastry, Y. Ma, Robust face recognition via sparse representation, *IEEE Trans. Pattern Anal. Mach. Intell.* 31 (2) (2009) 210–227.
- [31] S. Chen, D. Donoho, M. Saunders, Atomic decomposition by basis pursuit, *SIAM J. Sci. Comput.* 20 (1) (1998) 33–61.
- [32] J. Yang, J. Wright, T. Huang, Y. Ma, Image super-resolution via sparse representation, *IEEE Trans. Image Process.* 19 (11) (2010) 2861–2873.
- [33] T. Zhang, G. Lerman, A novel M-estimator for robust PCA, *J. Mach. Learn. Res.* 15 (1) (2011) 749–808.
- [34] E. Candes, T. Tao, The power of convex relaxation: near-optimal matrix completion, *IEEE Trans. Inf. Theory* 56 (5) (2010) 2053–2080.
- [35] E. Candes, Y. Plan, Matrix completion with noise, *P. IEEE* 98 (6) (2010) 925–936.
- [36] M. Yin, J. Gao, Z. Lin, Laplacian regularized low-rank representation and its applications, *IEEE Trans. Pattern Anal. Mach. Intell.* 38 (3) (2016) 504–517.
- [37] Z. Zhang, S. Yan, M. Zhao, Similarity preserving low-rank representation for enhanced data representation and effective subspace learning, *Neural Netw.* 53 (2014) 81–94.
- [38] Z. Zhang, S. Yan, M. Zhao, Robust image representation and decomposition by Laplacian regularized latent low-rank representation, in: *Proceedings of the International Joint Conference on Neural Networks*, 2013, pp. 1–7.
- [39] M. Yin, J. Gao, Z. Lin, Q. Shi, Y. Guo, Dual graph regularized latent low-rank representation for subspace clustering, *IEEE Trans. Image Process.* 24 (12) (2015) 4918–4933.
- [40] Z. Zhang, F. Li, M. Zhao, L. Zhang, S. Yan, Joint low-rank and sparse principal feature coding for enhanced robust representation and visual classification, *IEEE Trans. Image Process.* 25 (6) (2016) 2429–2443.
- [41] Z. Zhang, C. Liu, M. Zhao, A sparse projection and low-rank recovery framework for handwriting representation and salient stroke feature extraction, *ACM Trans. Intell. Syst. Technol.* 6 (1) (2015) 9:1–9:26.
- [42] C. Lu, J. Tang, S. Yan, Z. Lin, Nonconvex nonsmooth low rank minimization via iteratively reweighted nuclear norm, *IEEE Trans. Image Process.* 25 (2) (2016) 829–839.
- [43] S. Li, Y. Fu, Learning balanced and unbalanced graphs via low-rank coding, *IEEE Trans. Knowl. Data Eng.* 27 (5) (2015) 1274–1287.
- [44] G. Liu, S. Yan, Latent low-rank representation for subspace segmentation and feature extraction, in: *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, 2011, pp. 1615–1622.
- [45] Z. Ding, M. Shao, Y. Fu, Missing modality transfer learning via latent low-rank constraint, *IEEE Trans. Image Process.* 24 (11) (2015) 4322–4334.
- [46] S. Li, K. Li, Y. Fu, Temporal subspace clustering for human motion segmentation, in: *Proceedings of the International Conference on Computer Vision*, 2015, pp. 4453–4461.
- [47] S. Li, Y. Fu, Unsupervised transfer learning via low-rank coding for image clustering, in: *Proceedings of the International Joint Conference on Neural Networks*, 2016, pp. 1795–1802.
- [48] H. Zhang, Z. Yi, X. Peng, fLRR: fast low-rank representation using Frobenius-norm, *Electron. Lett.* 50 (13) (2014) 936–938.
- [49] C. Lu, H. Min, Z. Zhao, L. Zhu, D. Huang, S. Yan, Robust and efficient subspace clustering segmentation via least squares regression, in: *Proceedings of the European Conference on Computer Vision*, 2012, pp. 347–360.
- [50] R. Vidal, P. Favaro, Low rank subspace clustering (LRSC), *Pattern Recognit. Lett.* 43 (1) (2014) 47–61.
- [51] P. Ji, M. Salzmann, H. Li, Efficient dense subspace clustering, in: *Proceedings of the IEEE Winter Conference on Applications of Computer Vision*, 2014, pp. 461–468.
- [52] X. Peng, Z. Yi, H. Tang, Robust subspace clustering via thresholding ridge regression, in: *Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence*, 2015, pp. 1053–1066.
- [53] X. Peng, Z. Yu, Z. Yi, H. Tang, Constructing the L2-graph for robust subspace learning and subspace clustering, *IEEE T. Cybernet.* 47 (4) (2017) 1053–1066.
- [54] X. Peng, C. Lu, Z. Yi, H. Tang, Connections between nuclear norm and Frobenius norm based representation, *IEEE Trans. Neural Netw. Learn. Syst.* (99) 1–7.
- [55] B. Recht, M. Fazel, P.A. Parrilo, Guaranteed minimum-rank solutions of linear matrix equations via nuclear norm minimization, *SIAM Rev.* 52 (3) (2010) 471–501.
- [56] K.-C. Lee, J. Ho, D. Kriegman, Acquiring linear subspaces for face recognition under variable lighting, *IEEE Trans. Pattern Anal. Mach. Intell.* 27 (5) (2005) 684–698.
- [57] G. Liu, H. Xu, J. Tang, Q. Liu, S. Yan, A deterministic analysis for LRR, *IEEE Trans. Pattern Anal. Mach. Intell.* 38 (3) (2016) 417–430.
- [58] J. Shen, P. Li, H. Xu, Online low-rank subspace clustering by basis dictionary pursuit, in: *Proceedings of the Thirty-third International Conference on Machine Learning*, 48, New York, USA, 2016.



Song Yu: He received bachelor degree and master degree from college of electronic and information engineering in Nanjing University of Aeronautics and Astronautics in 2010 and 2013 respectively. Now he is a Ph.D. candidate in electronic and information engineering. His research interests include pattern recognition and subspace clustering.



Wu Yiquan: He received Ph.D. degree in 1998 from Nanjing University of Aeronautics and Astronautics. He is working as a professor and doctoral supervisor in college of electronic and information engineering in Nanjing University of Aeronautics and Astronautics. His research interests include image processing and machine learning.