Enhancing Low-Rank Subspace Clustering by Manifold Regularization

Junmin Liu, *Member, IEEE*, Yijun Chen, Jiangshe Zhang, and Zongben Xu

Abstract—Recently, low-rank representation (LRR) method has achieved great success in subspace clustering, which aims to cluster the data points that lie in a union of low-dimensional subspace. Given a set of data points, LRR seeks the lowest rank representation among the many possible linear combinations of the bases in a given dictionary or in terms of the data itself. However, LRR only considers the global Euclidean structure, while the local manifold structure, which is often important for many real applications, is ignored. In this paper, to exploit the local manifold structure of the data, a manifold regularization characterized by a Laplacian graph has been incorporated into LRR, leading to our proposed Laplacian regularized LRR (LapLRR). An efficient optimization procedure, which is based on alternating direction method of multipliers, is developed for LapLRR. Experimental results on synthetic and real data sets are presented to demonstrate that the performance of LRR has been enhanced by using the manifold regularization.

Index Terms—Subspace clustering, low-rank representation, manifold regularization.

I. Introduction

MANY real-word applications involve grouping, usually in an unsupervised manner, a set of objects into several subsets such that objects in the same subset are more similar than those in different groups, i.e. *data clustering* [1]. It is a common and challenging problem in data mining [2]–[4], machine learning and pattern recognition [5], [6]. However, the high-dimensionality of real data, ranging from hundreds to thousands, leads to the curse of dimensionality problem, and thus makes that directly performing clustering in the data space is infeasible. To deal with this, *subspace clustering* (SC) [2],

J. Liu is with the School of Mathematics and Statistics, Xi'an Jiaotong University, Xi'an 710049, China, and also with the Beijing Center for Mathematics and Information Interdisciplinary Sciences, Beijing 100190, China (e-mail: jmliu192@gmail.com).

Y. Chen is with the Department of Reference, Ningxia Library, Yinchuan 750011, China, and also with the School of Electronic and Information Engineering, Xi'an Jiaotong University, Xi'an 710049, China (e-mail: yijunchen@gmail.com).

J. Zhang and Z. Xu are with the School of Mathematics and Statistics, Xi'an Jiaotong University, Xi'an 710049, China (e-mail: jszhang@mail.xjtu.edu.cn; zbxu@mail.xjtu.edu.cn).

which is based on the assumption that high-dimensional data points lie in a union of low-dimensional subspaces, has been introduced by extending the single subspace to several subspace model. Due to the effectiveness of SC, it has received a lot of attention in recent years [7]–[9].

Over the past decades, many subspace clustering methods or models have been proposed, such as the Median K-flats [11], Generalized Principal Component Analysis (GPCA) [12], Robust Algebraic Segmentation [13], K-subspaces [10], Random Sample Consensus (RANSAC) [14], mixture of probabilistic principal component analysis [15], Gaussian Mixture Model [16], and so on. Following the perspective in [8], these methods can be classified into four categories: *statistical, iterative, algebraic, and spectral clustering-based* methods. Compared to the spectral clustering-based methods, most methods of the first three categories are sensitive to initialization and data errors (noise or outliers), and are difficult for optimization. In contrast, the spectral clustering-based methods are very easy to implement and can be solved efficiently by standard linear algebra methods. In addition, some spectral clustering-based methods proposed recently in [17] and [18] have achieved the state-of-the-art results in subspace clustering (or segmentation). Therefore, we here would like to focus on the spectral clustering-based methods.

The spectral clustering-based methods usually work by first learning an affinity matrix of data points, and then obtaining the clustering by applying spectral clustering methods such as K-means [19], Normalized Cuts (NCut) [20] to the affinity matrix. The key in applying spectral clustering-based methods is how to learn a good affinity matrix. Therefore, different affinity matrix learning methods will yield different subspace clustering methods with different properties. To learn a good affinity matrix on which the discriminating structures of the data points should be revealed, many algorithms have been proposed over the last few years. Some construct the affinity matrix only based on the local structure of the data (such as k-nearest neighbor structure), e.g., Local Subspace Affinity (LSA) [21], Spectral Local Best-fit Flats [22], Local Linear Manifold Clustering (LLMC) [23], to name a few. Some build it by trying to capture the *global structures* in the sense of exploiting the relations of the *whole* data set. One of the representative approaches is the Spectral Curvature Clustering (SCC) [24] method. However, only exploiting the local or global structures of the data to learn the affinity matrix limit their performances in *robust subspace clustering* [18].

Recently, inspired by the advances in compressed sensing [25], [26] and matrix completion [27], [28], Low-rank

Representation (LRR) [18], which is based on low-rank representation model, has been introduced to subspace clustering. LRR tries to build the affinity matrix by finding the lowest rank representation of the whole data and with each data point representing by a linear combinations of the bases in a given dictionary, typically choosing the data matrix itself as the dictionary. In the sense of simultaneously optimizing the data representation, LRR is claimed to capture the *global* structure of the data. It is demonstrated [17], [18] that LRR is one of the state-of-the-art methods in SC and superior to LSA, GPCA, RANSAC, and LLMC.

LRR exhibits good performance in segmenting motions in videos and clustering images of human faces. However, only the *global Euclidean structure*[1] of the data is considered. Many previous studies [29]–[37] have shown that integrating global and local structure is very important to data clustering [29] or classification [30]–[32], and that videos and images of human faces are more likely to reside on a low-dimensional submanifold of the high-dimensional ambient Euclidean space [33]–[37]. Thus, it is necessary to consider the intrinsic manifold structure for further improving the performance of LRR. To exploit such manifold information, we introduce the manifold regularization [38] to LRR and obtain its improved variant, called Laplacian regularized LRR (LapLRR). The manifold regularization is characterized by the Laplacian graph which captures the local geometrical structure of the data manifold such that nearby points in the intrinsic geometry of the data space are likely to have similar low-rank representations. By adding an additional manifold structure learning term to LRR, our proposed LapLRR is expected to have more discriminating power than LRR since they try to discover both the global Euclidean and local manifold structures of the data. In addition, a sequential optimization framework based on the *alternating direction method of multipliers* (ADMM) is developed to solve the optimization problem. It should be noted that a similar low-rank model with manifold regularization called manifold matrix factorization (MMF) was recently proposed by Zhang and Zhao in [39], but different with ours in the two main aspects. First, MMF takes the low-rank property of the representation coefficient matrix X as a constraint, whereas LapLRR incorporates it into the objective function. Second, the columns of the basis matrix A in MMF are unknown and restricted to be orthonormal, whereas the basis matrix in LapLRR is not necessary to be orthogonal and usually known (e.g. set to be the data X). Additionally, the optimization methods of MMF are based on the orthogonal property of X. If this property is lost, the methods will not work. On the other hand, LapLRR solves the problem by ADMM, which does not require any extra constrains of X.

The remaining of the paper is organized as follows: In Section II, we give a brief description of the subspace clustering problem and review of LRR. Section III introduces our proposed Laplacian regularized LRR models. The optimization algorithm of LapLRR is provided in Section IV.

Some experiments are shown in Section V. Finally, some conclusions are drawn in Section VI.

II. RELATED WORK

In this section, we will first describe the subspace clustering problem, and then give a brief review of LRR.

A. Subspace Clustering Problem

The generic problem of subspace clustering is the following. Let $\mathcal{Y} = \{y_i \in \mathcal{R}^d\}_{i=1}^N$ be a given set of N data points drawn from an unknown union of K linear subspaces $\mathcal{S}_1$, $\mathcal{S}_2$, ..., $\mathcal{S}_K$ of dimensions d_1, d_2, ..., d_K and bases $A_1 \in \mathcal{R}^{d \times d_1}$, $A_2 \in \mathcal{R}^{d \times d_2}$, ..., $A_K \in \mathcal{R}^{d \times d_K}$. The task of subspace clustering is to find the number of subspaces K, their dimensions $\{d_i\}_{i=1}^K$, the subspace bases $\{A_i\}_{i=1}^K$, and the segmentation of the data set $\mathcal{Y}$ according to the K subspaces, i.e. $\mathcal{Y} = \mathcal{Y}_1 \cup \mathcal{Y}_2 \cup \cdots \cup \mathcal{Y}_K$, where $\mathcal{Y}_i$ is a collection of data points drawn from $\mathcal{S}_i$.

The main challenges in subspace clustering are to 1) handle the unclean data corrupted by noise or outliers, which may distort the true subspaces structure, 2) require prior knowledge of the subspace parameters, such as the number of subspaces K and the dimensions $\{d_i\}_{i=1}^K$, which significantly affect the accuracy of results, 3) lack of theoretical guarantees for the optimality of the method. Out of many existing subspace clustering methods, LRR can simultaneously deal with the above challenges, and have been considered as one of the-state-of-art methods in subspace clustering. Our approach is an extension of LRR method. Therefore, in the following, we give a brief description of LRR.

B. LRR

LRR is a recently proposed subspace clustering method [17]. The basic idea of LRR is to seek the lowest rank representation among the many possible linear combinations of the bases in a given dictionary for the *whole* data set. This formulates the following problem[2]:

$$\min_X \ \|X\|_* \quad s.t. \quad Y = AX \tag{1}$$

where $A = [A_1, A_2, \ldots, A_K]$ is the bases matrix, $\|\cdot\|_*$ is the nuclear norm of a matrix, defined as the sum of the singular values of a matrix. Typically, when the data matrix itself is chosen as the dictionary, we have the following problem

$$\min_X \ \|X\|_* \quad s.t. \quad Y = YX, \tag{2}$$

In practice, the observed data Y are noisy, or there are model errors with the linear representation. To deal with the noise or model errors, an extension of (1) is to solve the following problem

$$\min_X \ \|X\|_* + \lambda \|E\|_p \quad s.t. \quad Y = AX + E \tag{3}$$

where $E \in \mathcal{R}^{d \times N}$ is the observation noise or model errors. $\|E\|_p$ denotes certain regularization strategy related with the property of E, for example

---

[1]A space, which is measured by the 2-norm, is clearly a *Euclidean space*. In this paper, the structure exploited by minimizing the approximate error of the whole data in Frobenius norm (or equally in 2-norm) is called global Euclidean structure.

[2]Note that the nuclear norm is a strong surrogate of the rank of a matrix.

- when E represents the Gaussian noise, $\|E\|_p \equiv \|E\|_F^2$, where $\|\cdot\|_F$ is the Frobenius norm.
- when E represents the random corruptions, $\|E\|_p \equiv \|E\|_1$, where $\|\cdot\|_1$ is the ℓ_1 norm.
- when E represents the sample-specific corruptions, $\|E\|_p \equiv \|E\|_{2,1}$, where $\|E\|_{2,1} = \sum_{j=1}^{N}\sqrt{\sum_{i=1}^{d}E_{ij}^2}$ is called $\ell_{2,1}$ norm.

When the representation coefficient matrix X is obtained, one can define the affinity matrix W as

$$W = |X| + |X^T|. \tag{4}$$

Then, the clustering is given by applying the spectral clustering method on W.

III. ENHANCING LRR BY MANIFOLD REGULARIZATION

In this section, we introduce the enhanced version of LRR, called Laplacian regularized LRR (LapLRR), which is based on a manifold regularization. Therefore, we begin with a description of manifold regularization.

A. Manifold Regularization

By assuming the data points drawn from a union of *linear* subspaces, we have the following representation

$$y_i = A x_i, \qquad i = 1, \ldots, N. \tag{5}$$

Here y_i is the observed data points, x_i is the new representation of y_i. Maintaining the above relation is optimal for the Euclidean structures in the data space. Generally, the spaces for many naturally occurring data, such as images, live on or close to a submanifold of the ambient space. Previous works [4], [41], and [42] have shown that both the Euclidean and manifold structures in the data space are important to low-dimensional representation from Eq. (5). However, many existing methods for subspace clustering problem only explore the Euclidean structure while fail to discover the the intrinsic geometry structure of the data manifold. Therefore, we here hope that the intrinsic geometry structure of the data manifold can be exploited to enhance this new representation. To this end, a natural assumption is that if two data points such as y_i and y_j are close in the intrinsic geometry of the data manifold, then the representations of this two data points, x_i and x_j, are also close to each other. This assumption is commonly referred to as *manifold assumption* [37], which has been applied to improve various kinds of algorithms [4], [16], [38], [41], [42].

Accurately estimating the global manifold structure of the data space is very challenge due to the insufficient number of samples and the high dimensionality of the ambient space. Therefore, many methods resort to capture the local manifold structure. Lots of efforts on *manifold learning* [35], [37] have shown that the local geometric structure of the data manifold can be effectively modeled through a nearest neighbor graph on the sampled data points.

In the following, we construct a nearest neighbor graph to characterize the local geometry of the data manifold. Given N data points $\{y_1, y_2, \ldots, y_N\} \subset \mathcal{R}^d$ sampled from the underlying submanifold, we can build a nearest neighbor graph $\mathcal{G}$ with its ith node corresponding to the data point y_i $(i = 1, 2, \ldots, N)$. For each node y_i, one can put an edge between it and its k nearest neighbors. Let $\mathcal{N}_k(y_i) = \{y_i^1, y_i^2, \ldots, y_i^k\}$ be the set of its k nearest neighbors. Thus, the similarity matrix S of graph $\mathcal{G}$ can be defined as follows:

$$S_{ij} = \begin{cases} \dfrac{y_i^T y_j}{\|y_i\|_2^2 \cdot \|y_j\|_2^2}, & \text{if } y_i \in \mathcal{N}_k(y_j) \text{ or } y_j \in \mathcal{N}_k(y_i), \\ 0, & \text{otherwise.} \end{cases}$$

To satisfy the manifold assumption, i.e. if two points y_i and y_j are close to each other, then their representations x_i and x_j are close as well, a reasonable choice is to minimize the following objective

$$\begin{aligned} \mathcal{T} &= \frac{1}{2}\sum_{i,j=1}^{N}\|x_i - x_j\|_2^2 S_{ij} \\ &= \sum_{i=1}^{N} x_i^T x_i D_{ii} - \sum_{i,j=1}^{N} x_i^T x_j S_{ij} \\ &= \mathrm{Tr}(XDX^T) - \mathrm{Tr}(XSX^T) \\ &= \mathrm{Tr}(XLX^T) \end{aligned} \tag{6}$$

where D is a diagonal matrix with the ith diagonal element $D_{ii} = \sum_j S_{ij}$, and $L = D - S$. The matrix L is usually called *graph Laplacian* [40]. The constant $1/2$ used in $\mathcal{T}$ is to simplify deductions. It is apparent that the objective function impose the smoothness of the representation coefficients, or the prior assumption that if neighboring points y_i and y_j are similar (a relatively bigger S_{ij}), their low-dimensional representations x_i and x_j should be very close. Hence, minimizing (6) is an attempt to ensure the manifold assumption. This manifold regularization $\mathcal{T}$ has been frequently used in enhancing various kinds of algorithms [16], [38], [41], [42].

B. LapLRR

LRR has yielded impressive results in SC. However, the local manifold geometry is not considered in LRR. Many studies [37], [41], [42] have shown that the local geometry structure of the data manifold is important to preserve the locality and the similarity among the data points. Thus, we incorporate the above manifold regularization term $\mathcal{T}$ into LRR, and formulate the Laplacian regularized LRR (LapLRR) problem

$$\min_{X} \quad \frac{1}{2}\|AX - Y\|_F^2 + \lambda_1\|X\|_* + \frac{\lambda_2}{2}\mathcal{T} \tag{7}$$
$$s.t. \quad X \geq 0$$

where $\lambda_1 > 0$ and $\lambda_2 > 0$ are the regularization parameters, the nonnegative constraints $X \geq 0$ is easy for interpretation of the representation. In LapLRR, the Euclidean space of data is exploited by keeping the reconstruction error $\|AX - Y\|_F^2$ as small as possible, the global structure of the data is captured by the nuclear norm regularization $\|X\|_*$, while the local manifold structure is exploited through the manifold regularization term $\mathcal{T}$. Hence, by considering both the global Euclidean and local manifold structures of data, we expect more discriminating information can be learned by LapLRR.

Algorithm 1 ADMM

Set $t = 0$, Choose $\mu > 0$, $\boldsymbol{v}^0$, $\boldsymbol{d}^0$

while stopping criterion is not met **do**

    1. $\boldsymbol{u}^{t+1} \leftarrow \arg\min_{\boldsymbol{u}} f(\boldsymbol{u}) + \frac{\mu}{2} \|\boldsymbol{G}\boldsymbol{u} - \boldsymbol{v}^t - \boldsymbol{d}^t\|_2^2$

    2. $\boldsymbol{v}^{t+1} \leftarrow \arg\min_{\boldsymbol{v}} g(\boldsymbol{u}) + \frac{\mu}{2} \|\boldsymbol{G}\boldsymbol{u}^{t+1} - \boldsymbol{v} - \boldsymbol{d}^t\|_2^2$

    3. $\boldsymbol{d}^{t+1} \leftarrow \boldsymbol{d}^t - (\boldsymbol{G}\boldsymbol{u}^{t+1} - \boldsymbol{v}^{t+1})$

    4. $t \leftarrow t + 1$

end while

IV. OPTIMIZATION

In this section, we propose to make use of the ADMM for solving the problem (7) of LapLRR. We begin by introducing the general framework of ADMM, and then deduce the iterative formulas of LapLRR.

A. ADMM

The ADMM [43], also called alternating direction augmented Lagrangian (ADAL), has recently attracted more attention in a wide range of research fields, such as compressive sensing [44], image restoration [43], matrix completion, due to its simple form and decoupling of variables. It is usually used to solve the problems with a convex, nonsmooth objective function and with structured linear constraints.

Consider the following structured optimization problem with linear constraints:

$$\min_{\boldsymbol{u},\boldsymbol{v}} f(\boldsymbol{u}) + g(\boldsymbol{v}) \tag{8}$$
$$s.t. \quad \boldsymbol{G}\boldsymbol{u} = \boldsymbol{v}$$

where both $f(\boldsymbol{u}) : \mathcal{R}^m \to \mathcal{R}$ and $g(\boldsymbol{v}) : \mathcal{R}^n \to \mathcal{R}$ are convex functions. For this problem, the augmented Lagrangian function is given by

$$\mathcal{L}(\boldsymbol{u},\boldsymbol{v},\boldsymbol{\alpha}) = f(\boldsymbol{u}) + g(\boldsymbol{v}) + \boldsymbol{\alpha}^T(\boldsymbol{G}\boldsymbol{u} - \boldsymbol{v}) + \frac{\mu}{2}\|\boldsymbol{G}\boldsymbol{u} - \boldsymbol{v}\|_2^2$$
$$= f(\boldsymbol{u}) + g(\boldsymbol{v}) + \frac{\mu}{2}\|\boldsymbol{G}\boldsymbol{u} - \boldsymbol{v} - \boldsymbol{d}\|_2^2 + \text{constant}$$

where $\boldsymbol{\alpha}$ is the Lagrange multipliers, $\boldsymbol{d} = -\boldsymbol{\alpha}/\mu$, and $\mu > 0$ is the penalty parameter. Different from the classical augmented Lagrangian method that attempts to jointly minimize the variable $\boldsymbol{u}$ and $\boldsymbol{v}$, ADMM alternately minimizes $\mathcal{L}(\boldsymbol{u},\boldsymbol{v},\boldsymbol{\alpha})$ with respect to $\boldsymbol{u}$ and $\boldsymbol{v}$ in a Gauss-Seidel manner. The general procedures of ADMM are summarized in Algorithm 1[3].

The convergence of ADMM has been guaranteed by the following theorem [45].

Theorem 4.1 (see [45]): Consider problem (8) with $\boldsymbol{G}$ having full columns rank and f, g being closed, proper, convex functions. Then, for arbitrary $\mu > 0$ and $\boldsymbol{u}_0, \boldsymbol{v}_0, \boldsymbol{d}_0$, if problem (8) has a solution, the sequences $\{\boldsymbol{u}^t, \boldsymbol{v}^t, \boldsymbol{d}^t\}$ generated by Algorithm 1 converges to it; otherwise, at least one of the sequences $\{(\boldsymbol{u}^t, \boldsymbol{v}^t)\}$ and $\{\boldsymbol{d}^t\}$ diverges.

---

[3]See [43] for more details about this version of ADMM.

B. Application of ADMM to LapLRR

In this subsection, we apply ADMM to the LapLRR problem. Similar to [43], the problem (7) can be rewritten as an unconstrained problem

$$\min_{\boldsymbol{X}} \quad \frac{1}{2}\|\boldsymbol{A}\boldsymbol{X} - \boldsymbol{Y}\|_F^2 + \lambda_1\|\boldsymbol{X}\|_* + \frac{\lambda_2}{2}\text{Tr}(\boldsymbol{X}\boldsymbol{L}\boldsymbol{X}^T) + \iota_{\mathcal{R}_+}(\boldsymbol{X})$$

where $\iota_{\mathcal{R}_+}$ is the indicator function, defined as

$$\iota_{\mathcal{R}_+}(x) = \begin{cases} 0, & \text{if } x \geq 0 \\ +\infty & \text{otherwise.} \end{cases} \tag{9}$$

By introducing some auxiliary variables, the above problem has the following equivalent form

$$\min_{\boldsymbol{X}} \frac{1}{2}\|\boldsymbol{Z}_1 - \boldsymbol{Y}\|_F^2 + \lambda_1\|\boldsymbol{Z}_2\|_* + \lambda_2\text{Tr}(\boldsymbol{Z}_3\boldsymbol{L}\boldsymbol{Z}_3^T) + \iota_{R_+}(\boldsymbol{Z}_4)$$
$$s.t. \quad \boldsymbol{A}\boldsymbol{X} = \boldsymbol{Z}_1, \quad \boldsymbol{X} = \boldsymbol{Z}_2, \quad \boldsymbol{X} = \boldsymbol{Z}_3, \quad \boldsymbol{X} = \boldsymbol{Z}_4 \tag{10}$$

which has an augmented Lagrangian function of the form

$$\mathcal{L}(\boldsymbol{X}, \boldsymbol{Z}_1, \boldsymbol{Z}_2, \boldsymbol{Z}_3, \boldsymbol{Z}_4)$$
$$= \frac{1}{2}\|\boldsymbol{Z}_1 - \boldsymbol{Y}\|_F^2 + \lambda_1\|\boldsymbol{Z}_2\|_* + \frac{\lambda_2}{2}\text{Tr}(\boldsymbol{Z}_3\boldsymbol{L}\boldsymbol{Z}_3^T) + \iota_{R+1}(\boldsymbol{Z}_4)$$
$$+ \frac{\mu}{2}\|\boldsymbol{A}\boldsymbol{X} - \boldsymbol{Z}_1 - \boldsymbol{D}_1\|_F^2 + \frac{\mu}{2}\|\boldsymbol{X} - \boldsymbol{Z}_2 - \boldsymbol{D}_2\|_F^2$$
$$+ \frac{\mu}{2}\|\boldsymbol{X} - \boldsymbol{Z}_3 - \boldsymbol{D}_3\|_F^2 + \frac{\mu}{2}\|\boldsymbol{X} - \boldsymbol{Z}_4 - \boldsymbol{D}_4\|_F^2. \tag{11}$$

Then, we apply the alternating minimization idea to update $\boldsymbol{X}$, $\boldsymbol{Z}_1, \boldsymbol{Z}_2, \boldsymbol{Z}_3, \boldsymbol{Z}_4$, i.e. update one of the five variables with the other fixed.

Given the current point $\boldsymbol{X}^t, \boldsymbol{Z}_1^t, \boldsymbol{Z}_2^t, \boldsymbol{Z}_3^t, \boldsymbol{Z}_4^t, \boldsymbol{D}_1^t, \boldsymbol{D}_2^t, \boldsymbol{D}_3^t, \boldsymbol{D}_4^t$, we update $\boldsymbol{X}^{t+1}$ by minimizing $\mathcal{L}$ with respect to $\boldsymbol{X}$, i.e.

$$\min_{\boldsymbol{X}} \quad \frac{\mu}{2}\|\boldsymbol{A}\boldsymbol{X} - \boldsymbol{Z}_1^t - \boldsymbol{D}_1^t\|_F^2 + \frac{\mu}{2}\|\boldsymbol{X} - \boldsymbol{Z}_2^t - \boldsymbol{D}_2^t\|_F^2$$
$$+ \frac{\mu}{2}\|\boldsymbol{X} - \boldsymbol{Z}_3 - \boldsymbol{D}_3\|_F^2 + \frac{\mu}{2}\|\boldsymbol{X} - \boldsymbol{Z}_4 - \boldsymbol{D}_4^t\|_F^2$$

which produces the updating formula as

$$\boldsymbol{X}^{t+1} \leftarrow (\boldsymbol{A}^T\boldsymbol{A} + 3\boldsymbol{I})^{-1}(\boldsymbol{A}^T\zeta_1^t + \zeta_2^t + \zeta_3^t + \zeta_4^t) \tag{12}$$

where $\boldsymbol{I}$ is the identity matrix, $\zeta_1^t = \boldsymbol{Z}_1^t + \boldsymbol{D}_1^t$, $\zeta_2^t = \boldsymbol{Z}_2^t + \boldsymbol{D}_2^t$, $\zeta_3^t = \boldsymbol{Z}_3^t + \boldsymbol{D}_3^t$, and $\zeta_4^t = \boldsymbol{Z}_4^t + \boldsymbol{D}_4^t$.

To update $\boldsymbol{Z}_1$, we solve

$$\min_{\boldsymbol{Z}_1} \quad \frac{1}{2}\|\boldsymbol{Z}_1 - \boldsymbol{Y}\|_F^2 + \frac{\mu}{2}\|\boldsymbol{A}\boldsymbol{X}^{t+1} - \boldsymbol{Z}_1 - \boldsymbol{D}_1^t\|_F^2 \tag{13}$$

which yields the following updating form

$$\boldsymbol{Z}_1^{t+1} \leftarrow \frac{1}{1+\mu}\left[\boldsymbol{Y} + \mu(\boldsymbol{A}\boldsymbol{X}^{t+1} - \boldsymbol{D}_1^t)\right] \tag{14}$$

To update $\boldsymbol{Z}_2$, we have the augmented Lagrangian subproblem

$$\min_{\boldsymbol{Z}_2} \quad \lambda_1\|\boldsymbol{Z}_2\|_* + \frac{\mu}{2}\|\boldsymbol{X}^{t+1} - \boldsymbol{Z}_2 - \boldsymbol{D}_2^t\|_F^2 \tag{15}$$

which can be solved by the well-known *singular value thresholding* (SVT) operator [18]. Define the SVT operator $\mathcal{D}_\tau$ as

$$\mathcal{D}_\tau(\boldsymbol{X}) = \mathcal{U}\mathcal{S}_\tau\mathcal{V}^T \tag{16}$$

Algorithm 2 LapLRR

---

initialize: set $t = 0$, choose $\lambda_1 > 0$, $\lambda_2 > 0$, $\mu > 0$, $U^0 = D^0 = 0$.

while stopping criterion is not met **do**

  1. $X_{t+1} = (A^T A + 3I)^{-1}(A^T \zeta_1^t + \zeta_2^t + \zeta_3^t + \zeta_4^t)$

  2. $Z_1^{t+1} = \frac{1}{1+\mu}\left[Y + \mu(AX^{t+1} - D_1^t)\right]$

  3. $Z_2^{t+1} = \mathcal{D}_{\lambda_1/\mu}(X^{t+1} - D_2^t)$

  4. $Z_3^{t+1} = (X^{t+1} - D_3^t)(\frac{\lambda_2}{\mu}L + I)^{-1}$

  5. $Z_4^{t+1} = \max(X^{t+1} - D_4^t, 0)$

  6. Update the Lagrange multipliers:

    6.1 $D_1^{t+1} = D_1^t - (AX_{t+1} - Z_1^{t+1})$

    6.2 $D_2^{t+1} = D_2^t - (X_{t+1} - Z_2^{t+1})$

    6.3 $D_3^{t+1} = D_3^t - (X_{t+1} - Z_3^{t+1})$

    6.4 $D_4^{t+1} = D_4^t - (X_{t+1} - Z_4^{t+1})$

  7. $t \leftarrow t + 1$

end while

output: X

---

where $X = \mathcal{U}\Sigma\mathcal{V}^T$ is the singular value decomposition, and $\mathcal{S}_\tau[x] = \text{sgn}(x)\max(|x| - \tau, 0)$ is the shrinkage operator, we have the updating formula of Z_2 as

$$Z_2^{t+1} \leftarrow \mathcal{D}_{\lambda_1/\mu}(X^{t+1} - D_2^t). \tag{17}$$

It is easy to get the iteration of Z_3 as

$$Z_3^{t+1} \leftarrow (X^{t+1} - D_3^t)(\frac{\lambda_2}{\mu}L + I)^{-1} \tag{18}$$

by minimizing the following subproblem with respect to Z_3, given by

$$\min_{Z_3} \quad \lambda_2 \text{Tr}(Z_3 L Z_3^T) + \frac{\mu}{2}\|X^{t+1} - Z_3 - D_3^t\|_F^2. \tag{19}$$

To update Z_4, we need to solve

$$\min \quad \iota_{R_+1}(Z_4) + \frac{\mu}{2}\|X^{t+1} - Z_4 - D_4^t\|_F^2 \tag{20}$$

which yields the updating rule

$$Z_4^{t+1} \leftarrow \max(X^{t+1} - D_4^t, 0). \tag{21}$$

Finally, we update the Lagrangian multipliers D_1, D_2, D_3, D_4 by

$$D_1^{t+1} = D_1^t - (AX_{t+1} - Z_1^{t+1})$$
$$D_2^{t+1} = D_2^t - (X_{t+1} - Z_2^{t+1})$$
$$D_3^{t+1} = D_3^t - (X_{t+1} - Z_3^{t+1})$$
$$D_4^{t+1} = D_4^t - (X_{t+1} - Z_4^{t+1})$$

With all the above updating formulas collected, we obtain the ADMM scheme for LapLRR in Algorithm 2 above.

The convergence of Algorithm 2 for problem (10) is guaranteed by Theorem 4.1 since it can be expressed as an instance of problem (8). For instance, by letting

$$U \equiv X, \quad V \equiv \begin{bmatrix} Z_1 \\ Z_2 \\ Z_3 \\ Z_4 \end{bmatrix}, \quad G \equiv \begin{bmatrix} A \\ I \\ I \\ I \end{bmatrix}, \quad f(U) \equiv 0, \tag{22}$$

and $g(V) \equiv \frac{1}{2}\|Z_1 - Y\|_F^2 + \lambda_1\|Z_2\|_* + \lambda_2\text{Tr}(Z_3 L Z_3^T) + \iota_{R_+}(Z_4)$, we then can write problem (10) in the form of problem (8), and obtain the augmented Lagrangian function (11). In addition, G is a full column rank matrix, and functions f, g are closed, proper, convex. These meet the conditions in Theorem 4.1. Hence, the convergence of Algorithm 2 is guaranteed by Theorem 4.1.

V. EXPERIMENTS

In this section, we carry out several experiments on both the synthetic and real world data to demonstrate the efficacy of our proposed approach. We also compare the clustering results obtained by LapLRR to those obtained by LRR, and also to those obtained by the K-means, Principle Component Analysis (PCA), Graph regularized Sparse Coding (GSC) [29] and Graph regularized Nonnegative Matrix Factorization (GNMF) [41]. For PCA, LRR, LapLRR, GSC and GNMF, the K-means method is applied to find the clustering results. And two metrics, namely the accuracy (AC) and the Normalized Mutual Information (NMI)[4], are used to evaluate the results. In our experiments, the parameters of each methods are manually tuned to achieve the best performance.

A. Synthetic Data

To validate the effectiveness of our proposed LapLRR, a synthetic data set has been created and used for experiments in this subsection. A major advantage of using the synthetic data is that all of the details of ground truth are perfectly known such that the performance of the algorithms can be quantitatively evaluated. Similar to [46], we construct 5 independent subspaces $\{S\}_{i=1}^5 \subset \mathcal{R}^{200}$ of dimension 4 and bases $\{A_i \subset \mathcal{R}^{200\times 4}\}_{i=1}^5$. By collecting the each bases in a matrix, we get the bases matrix $A = [A_1, A_2, A_3, A_4, A_5] \subset \mathcal{R}^{200\times 20}$, which are generated by randomly selecting 20 columns of a random orthogonal matrix of dimension 200×200. Then, we sample 200 data points with each subspace 40 data points by $Y = AX$, where x_{ij} is a random number uniformly sampled from $[0, 1]$ if y_i and y_j belong to the same subspace, otherwise, $x_{ij} = 0$. The generated representation matrix $X \subset \mathcal{R}^{20\times 200}$ are shown in Fig. 1(a). Data points with different numbers, from zero percent to 100 percent at 10 percent intervals, are randomly chosen to corrupt in a similar manner as in [46]. Each *level of corruption* (LoC) is based on 20 randomly generated realizations.

We compare the performance of LapLRR and LRR by using the data matrix X as the bases matrix. The average clustering AC and NMI obtained by them are reported in Table I.

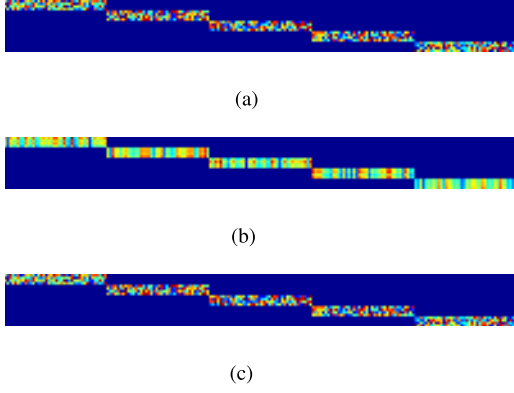[4]Please see [16], [41] for the detailed definitions of AC and NMI.

(a)

(b)

(c)

Fig. 1. Comparison of (a) the true representation matrix, (b) the estimated one by LRR, and (c) the estimated one by LapLRR. (a) True representation matrix. (b) Estimated representation matrix by LRR. (c) Estimated representation matrix by LapLRR.

TABLE I
CLUSTERING RESULTS ON SYNTHETIC DATA
OF K-MEANS, PCA, LRR AND LapLRR

LoC	AC(%)				NMI(%)			
	K-means	PCA	LRR	LapLRR	K-means	PCA	LRR	LapLRR
1	**100**	**100**	**100**	**100**	**100**	**100**	**100**	**100**
2	96.4	96.5	96.7	**99.3**	94.0	93.9	92.1	**97.8**
3	90.7	92.7	90.3	**96.9**	86.9	88.5	81.2	**95.2**
4	83.3	85.1	83.9	**94.0**	77.8	79.9	72.7	**90.8**
5	76.4	79.6	81.8	**93.0**	67.3	70.3	70.0	**89.6**
6	70.0	71.8	85.2	**91.1**	56.5	59.6	74.8	**86.4**
7	61.7	59.8	79.1	**88.4**	46.4	43.5	67.6	**81.3**
8	48.9	52.6	75.0	**81.6**	30.0	34.8	61.1	**72.8**
9	32.2	39.0	58.5	**72.4**	12.0	18.3	40.0	**62.7**
10	27.8	25.8	42.4	**56.3**	7.3	5.6	21.2	**41.3**
11	27.3	28.4	**34.4**	21.8	5.7	6.3	**10.7**	2.3
Avg.	65.0	66.5	75.2	**81.3**	53.1	54.6	62.8	**74.6**

From this table, we can see that our proposed LapLRR method provide better performance than PCA, LRR and K-means. Since the ground truth of bases matrix A and representation matrix X is available, it is possible to evaluate recoverability of the representation matrix X by using the A as the bases matrix. A visualization comparison of the true representation matrix X and the estimated ones by LRR and LapLRR is depicted in Fig. 1. It can be found from this figure that our proposed LapLRR method can better approximate the true representation matrix.

B. Handwritten Digit Databases

In this subsection, we use two publicly available handwritten digit databases, namely USPS[5] and MNIST[6],

[5]Available at: http://www.gaussianprocess.org/gpml/data/.
[6]Available at: http://yann.lecun.com/exdb/mnist/.
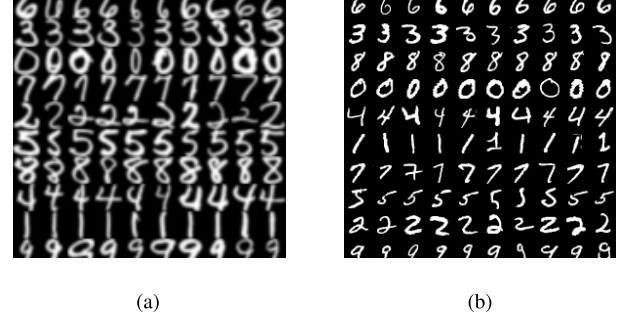

(a)   (b)

Fig. 2. Sample handwritten digit images from the (a) USPS and (b) MNIST databases. (a) USPS. (b) MNIST.

which are shown to roughly reside in a low-dimensional subspace.

- USPS Database: This handwritten database contains totally 9298 digit images of "0" through "9", each of which is of size 16×16 pixels, with 256 gray levels per pixel. In the experiment, each image is represented by a 256-dimensional vector.
- MNIST Database: This handwritten database has 70,000 digit images of "0" through "9". The images of each class (digit) are of size 28×28. Thus, each digit image is represented by a 784-dimensional vector.

Some sample handwritten digit images from the two databases are shown in Fig. 2.

For each database, we randomly select 100 digit images with each class in this experiment. Similar to [16], the clustering results with different cluster numbers are based on 20 test runs on different randomly chosen classes for each given cluster number K. Table II reports the averaged results using these two databases.

As we can see, our proposed LapLRR achieved the best clustering accuracy than K-means, GSC, and GNMF methods in both USPS and MNIST databases. In addition, the both GSC and GNMF, which also incorporate the manifold regularization into the objective functions, perform better than LRR and K-means. This indicates that the manifold regularization term is capable of enhancing the performance of clustering algorithms.

Fig. 3 shows the representation matrices learned by LapLRR and LRR in the USPS database when $K = 10$. It can be found that the block structure of the representation matrix learned by LapLRR is more clear than those learned by LRR. Therefore, our proposed LapLRR can better capture the discriminative structure of the image space.

C. PIE Face Database

The PIE face database is widely used to test the performance of clustering methods. This database contains 2856 face images with size of 32×32. There are 68 people (or classes), each of which has 42 facial images under different light and illumination conditions. Fig. 4 shows some sample images.

Similar to the experiments in [29], we randomly select the cluster number K (ranging from 4 to 68) classes (persons) from the 38 classes in the whole database

TABLE II

CLUSTERING RESULTS ON USPS AND MNIST DIGIT DATABASES OF K-MEANS, LRR, LAPLRR, GSC AND GNMF

| K	USPS Digit Database								MNIST Digit Database							
	AC(%)				NMI(%)				AC(%)				NMI(%)			
	K-means	LRR	LapLRR	GSC	K-means	LRR	LapLRR	GSC	K-means	LRR	LapLRR	GNMF	K-means	LRR	LapLRR	GNMF
2	94.1	94.8	**96.4**	96.1	71.0	73.1	**80.5**	79.3	92.6	94.1	**96.1**	96.2	69.5	72.2	**82.1**	82.3
3	89.2	92.1	**94.3**	93.9	71.3	75.9	**81.8**	80.7	76.1	77.8	**88.3**	86.3	55.7	57.4	**77.0**	74.0
4	83.4	87.4	**90.0**	89.4	67.6	72.4	**78.9**	77.7	73.0	73.9	**80.3**	73.7	57.3	57.3	**69.3**	66.7
5	80.4	84.3	**87.1**	85.8	65.1	70.9	**77.2**	75.4	67.0	66.3	**71.4**	69.3	58.0	57.6	**67.3**	65.9
6	78.6	83.9	**88.6**	84.8	65.9	73.2	**79.6**	75.8	63.3	63.5	**70.1**	64.2	54.1	54.0	**66.0**	61.4
7	74.3	79.3	**82.5**	81.0	62.2	69.6	**75.7**	72.9	58.4	58.6	**67.1**	58.8	54.2	53.9	**65.2**	58.7
8	72.3	77.6	**80.4**	77.0	61.2	68.8	**74.5**	71.4	59.5	59.3	**66.3**	57.4	54.1	54.2	**65.2**	59.0
9	69.2	76.7	**79.3**	76.7	59.9	67.7	**73.4**	70.4	54.8	55.1	**62.5**	54.6	51.7	51.3	**62.5**	56.3
10	65.7	71.9	**75.4**	72.5	59.1	66.3	**72.2**	68.7	53.9	54.4	**61.1**	50.6	51.7	51.7	**62.1**	54.4
Avg.	78.6	83.1	**86.0**	84.1	64.8	70.9	**77.1**	74.7	66.5	67.0	**73.7**	67.9	56.3	56.6	**68.5**	64.3

(a)   (b)
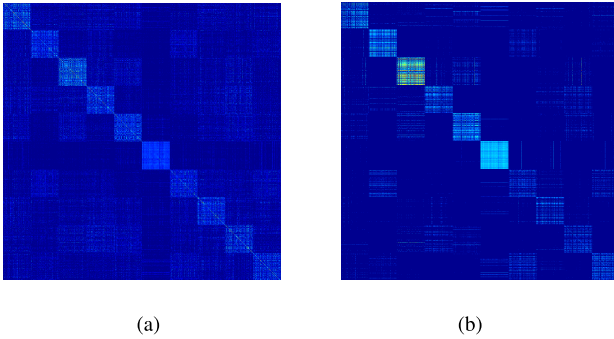
Fig. 3. The representation matrices learned by (a) LRR and (b) LapLRR in the USPS handwritten database when $K = 10$. (a) LRR. (b) LapLRR.

Fig. 4. Sample face images from the PIE databases. Each row shows the images corresponding to one people.

TABLE III

CLUSTERING RESULTS ON PIE FACE DATABASES OF K-MEANS, PCA, LRR, LAPLRR, GSC, AND GNMF

LoC	AC(%)					
	K-means	PCA	LRR	LapLRR	GSC	GNMF
4	63.1	59.5	**100**	**100**	**100**	**100**
12	32.1	32.5	81.3	**100**	99.2	**100**
20	36.0	39.0	87.1	92.6	**92.9**	80.7
28	35.7	40.0	85.7	**90.3**	84.7	79.1
36	33.6	38.0	88.1	**88.9**	81.3	81.0
44	35.1	33.0	79.8	**86.8**	77.1	74.0
52	34.7	34.8	80.4	**87.2**	76.2	79.4
60	33.1	32.9	86.9	**87.9**	81.5	77.1
68	33.4	32.3	84.0	**88.4**	79.9	75.4
Avg.	37.4	38.0	85.9	**91.3**	85.9	83.0

for this experiment. For each value of K, we run 20 generated realizations on different randomly chosen classes.

For the limitation of space and also because AC and NMI disclose similar pattern of behavior, we only present the results evaluated by AC. The averaged AC obtained by K-means, PCA, LRR, LapLRR, GSC and GNMF are reported in Table III, and the representation matrices leared by LRR and LapLRR are shown in Fig. 6. From the table and figure, we can see that our method outperforms the other five methods, which proves the effectiveness of our LapLRR again.

D. COIL20 Image Database

We use the COIL20 image database to test the clustering performance. This database has 32×32 gray scale images of 20 objects viewed from varying angles, see Fig. 5 for some

sample images. Here, we carry out this experiment with the cluster number $K = 2, \ldots, 20$. For each K, the best parameters of the six methods are selected. After PCA projection, ninety-nine percenter energy of this data are retained. The averaged AC obtained by K-means, PCA, LRR, GSC, GNMF and our proposed LapLRR are shown in Table IV.

From Table IV, it is observed that the three algorithms LapLRR, GSC and GNMF, which integrate the local manifold structure in the representation, can be ranked from best to worst as LapLRR, GSC and GNMF according to the mean AC. And we can draw a similar conclusion with the above three experiments that by integrating the local manifold structure in the representation, the performance of the clustering algorithms can be significantly enhanced, especially for the LRR algorithm.
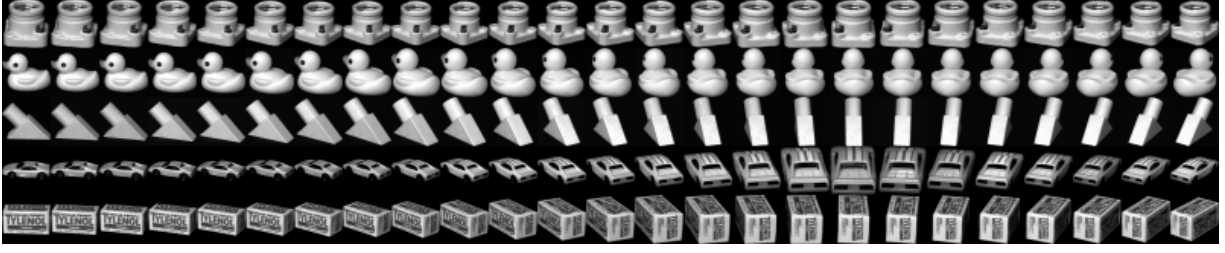
Fig. 5.   Sample images from the COIL20 image databases.
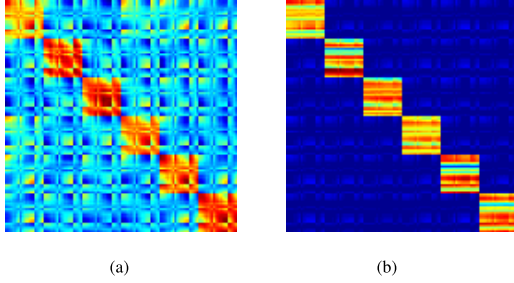

| | (a) | | (b) |

Fig. 6.   The representation matrices learned by (a) LRR and (b) LapLRR in the PIE face database when $K = 6$. (a) LRR. (b) LapLRR.

TABLE IV

CLUSTERING RESULTS ON COIL20 IMAGE DATABASES OF K-MEANS, PCA, LRR, GSC, GNMF, AND OUR PROPOSED LapLRR

LoC	AC(%)					
	K-means	PCA	LRR	LapLRR	GSC	GNMF
2	88.3	88.5	90.3	**98.6**	98.1	96.7
3	85.8	85.5	89.2	**97.3**	97.0	94.8
4	84.8	84.7	88.2	96.2	**96.4**	94.0
5	87.5	87.5	89.5	**96.3**	96.2	94.3
6	74.8	79.2	83.2	**95.2**	94.0	94.1
7	74.0	70.3	75.4	**92.8**	91.6	91.3
8	73.2	71.5	77.8	**89.1**	86.7	86.5
9	74.2	73.1	76.1	**90.5**	89.2	88.8
10	71.0	69.6	74.2	**89.4**	87.7	87.5
11	70.0	70.2	71.3	**87.9**	86.5	86.9
12	68.2	68.9	70.7	**87.3**	86.3	86.4
13	67.7	67.4	69.4	**87.1**	85.8	86.0
14	65.1	65.3	66.0	**86.2**	84.0	84.2
15	66.5	65.7	67.8	**87.0**	85.4	84.9
16	66.3	65.0	67.2	**87.4**	85.6	85.0
17	64.3	64.5	66.7	**85.9**	84.1	84.2
18	63.4	64.0	65.2	**84.8**	83.3	83.1
19	62.7	63.8	64.6	**82.1**	81.5	81.2
20	62.2	63.0	64.2	**79.6**	78.8	77.6
Avg.	72.1	72.0	74.6	**89.5**	88.3	87.8

VI. CONCLUSIONS

Naturally occurring data are often high dimensionality and complex structures such that single subspace can not character them. Multisubspace learning, especially the subspace clustering techniques, is an effective way to address this problem. Extending on the recently work in [18], we propose a novel subspace clustering method, called Laplacian regularized low-rank representation (LapLRR), by introducing an manifold regularization. Our proposed LapLRR can exploit both the global Euclidean and local manifold structures of the data, and thus makes itself learned more discriminating information. Experimental results on both synthetic and real data sets show that our proposed LapLRR achieves excellent performances for handwritten digit clustering, face clustering, and image clustering problems.

REFERENCES

[1] A. K. Jain, M. N. Murty, and P. J. Flynn, "Data clustering: A review," *ACM Comput. Surv.*, vol. 31, no. 3, pp. 264–323, 1999.

[2] L. Parsons, E. Haque, and H. Liu, "Subspace clustering for high dimensional data: A review," *ACM SIGKDD Explorations Newslett.*, vol. 6, no. 1, pp. 90–105, Jun. 2004.

[3] E. Cesario, G. Manco, and R. Ortale, "Top-down parameter-free clustering of high-dimensional categorical data," *IEEE Trans. Knowl. Data Eng.*, vol. 19, no. 12, pp. 1607–1624, Dec. 2007.

[4] D. Cai, X. He, and J. Han, "Document clustering using locality preserving indexing," *IEEE Trans. Knowl. Data Eng.*, vol. 17, no. 12, pp. 1624–1637, Dec. 2005.

[5] A. Y. Yang, J. Wright, Y. Ma, and S. S. Sastry, "Unsupervised segmentation of natural images via lossy data compression," *Comput. Vis. Image Understand.*, vol. 110, no. 2, pp. 212–225, May 2008.

[6] R. Vidal, R. Tron, and R. Hartley, "Multiframe motion segmentation with missing data using PowerFactorization and GPCA," *Int. J. Comput. Vis.*, vol. 79, no. 1, pp. 85–105, Aug. 2008.

[7] K. Kailing, H. Kriegel, and P. Kröger, "Density-connected subspace clustering for high-dimensional data," in *Proc. 4th SAIM Int. Conf. Data Mining*, Lake Buena Vista, FL, USA, 2004, pp. 246–257.

[8] R. Vidal, "Subspace clustering," *IEEE Signal Process. Mag.*, vol. 28, no. 2, pp. 52–68, Mar. 2011.

[9] M. Soltanolkotabi and E. J. Candès, "A geometric analysis of subspace clustering with outliers," *Ann. Statist.*, vol. 40, no. 4, pp. 2195–2238, 2012.

[10] J. Ho, M.-H. Yang, J. Lim, K.-C. Lee, and D. Kriegman, "Clustering appearances of objects under varying illumination conditions," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, vol. 1. Jun. 2003, pp. 11–18.

[11] T. Zhang, A. Szlam, and G. Lerman, "Median K-flats for hybrid linear modeling with many outliers," in *Proc. IEEE 12th Int. Conf. Comput. Vis. Workshops*, Sep./Oct. 2009, pp. 234–241.

[12] Y. Ma, A. Yang, H. Derksen, and R. Fossum, "Estimation of subspace arrangements with applications in modeling and segmenting mixed data," *SIAM Rev.*, vol. 50, no. 3, pp. 413–458, Aug. 2008.

[13] S. Rao, A. Yang, S. Sastry, and Y. Ma, "Robust algebraic segmentation of mixed rigid-body and planar motions in two views," *Int. J. Comput. Vis.*, vol. 88, no. 3, pp. 425–446, Jul. 2010.

[14] M. A. Fischler and R. C. Bolles, "Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography," *Commun. ACM*, vol. 24, no. 6, pp. 381–395, Jun. 1981.

[15] M. E. Tipping and C. M. Bishop, "Mixtures of probabilistic principal component analyzers," *Neural Comput.*, vol. 11, no. 2, pp. 443–482, Feb. 1999.

[16] X. He, D. Cai, Y. Shao, H. Bao, and J. Han, "Laplacian regularized Gaussian mixture model for data clustering," *IEEE Trans. Knowl. Data Eng.*, vol. 23, no. 9, pp. 1406–1418, Sep. 2011.

[17] E. Elhamifar and R. Vidal, "Sparse subspace clustering: Algorithm, theory, and applications," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 11, pp. 2765–2781, Nov. 2013. [Online]. Available: http://arxiv.org/abs/1203.1005

[18] G. Liu, Z. Lin, S. Yan, J. Sun, Y. Yu, and Y. Ma, "Robust recovery of subspace structures by low-rank representation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 1, pp. 171–184, Jan. 2013.

[19] L. Jiang, M. K. Ng, and J. Z. Huang, "An entropy weighting k-means algorithm for subspace clustering of high-dimensional sparse data," *IEEE Trans. Knowl. Data Eng.*, vol. 19, no. 8, pp. 1026–1041, Aug. 2007.

[20] J. Shi and J. Malik, "Normalized cuts and image segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 8, pp. 888–905, Aug. 2000.

[21] J. Yan and M. Pollefeys, "A general framework for motion segmentation: Independent, articulated, rigid, non-rigid, degenerate and non-degenerate," in *Proc. 9th Eur. Conf. Comput. Vis.*, May 2006, pp. 94–106.

[22] T. Zhang, A. Szlam, Y. Wang, and G. Lerman, "Hybrid linear modeling via local best-fit flats," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Oct. 2010, pp. 1927–1934.

[23] A. Goh and R. Vidal, "Segmenting motions of different types by unsupervised manifold clustering," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2007, pp. 1–6.

[24] G. Chen and G. Lerman, "Spectral curvature clustering (SCC)," *Int. J. Comput. Vis.*, vol. 81, no. 3, pp. 317–330, 2009.

[25] E. J. Candes, J. Romberg, and T. Tao, "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information," *IEEE Trans. Inf. Theory*, vol. 52, no. 2, pp. 489–509, Feb. 2006.

[26] D. L. Donoho, "Compressed sensing," *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1289–1306, Apr. 2006.

[27] E. Candes and B. Recht, "Exact matrix completion via convex optimization," *Found. Comput. Math.*, vol. 9, no. 6, pp. 717–772, Dec. 2009.

[28] E. Candes, X. Li, Y. Ma, and J. Wright, "Robust principal component analysis?" *J. ACM*, vol. 58, no. 3, May 2011, Art. ID 11.

[29] M. Zheng et al., "Graph regularized sparse coding for image representation," *IEEE Trans. Image Process.*, vol. 20, no. 5, pp. 1327–1560, May 2011.

[30] Q. Gao, J. Liu, H. Zhang, X. Gao, and K. Li, "Joint glaobal and local structure discriminant analysis," *IEEE Trans. Inf. Forensics Security*, vol. 8, no. 4, pp. 626–635, Apr. 2013.

[31] N. Guan, D. Tao, Z. Luo, and B. Yuan, "Manifold regularized discriminative nonnegative matrix factorization with fast gradient descent," *IEEE Trans. Image Process.*, vol. 20, no. 7, pp. 2030–2048, Jul. 2011.

[32] J. Chen, J. Ye, and Q. Li, "Integrating global and local structures: A least squares framework for dimensionality reduction," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Minneapolis, MN, USA, Jun. 2007, pp. 1–8.

[33] A. Shashua, "On photometric issues in 3D recognition from a single 2D image," *Int. J. Comput. Vis.*, vol. 21, no. 1, pp. 99–122, Jan. 1997.

[34] K.-C. Lee, J. Ho, M.-H. Yang, and D. Kriegman, "Video-based face recognition using probabilistic appearance manifolds," in *Proc. IEEE Comput. Soc. Comput. Vis. Pattern Recognit.*, vol. 1. Jun. 2003, pp. 313–320.

[35] H. S. Seung and D. D. Lee, "The manifold ways of perception," *Science*, vol. 290, no. 5500, pp. 2268–2269, 2000.

[36] J. B. Tenenbaum, V. D. Silva, and J. C. Langford, "A global geometric framework for nonlinear dimensionality reduction," *Science*, vol. 290, no. 5500, pp. 2319–2323, Dec. 2000.

[37] X. He and P. Niyogi, "Locality preserving projections," *Adv. Neural Inf. Process. Syst.*, vol. 16, no. 16, pp. 153–160, 2004.

[38] M. Belkin, P. Niyogi, and V. Sindhwani, "Manifold regularization: A geometric framework for learning from labeled and unlabeled examples," *J. Mach. Learn. Res.*, vol. 7, pp. 2399–2434, Nov. 2006.

[39] Z. Zhang and K. Zhao, "Low-rank matrix approximation with manifold regularization," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 7, pp. 1717–1729, Jul. 2013.

[40] F. R. K. Chung, *Spectral Graph Theory*. Providence, RI, USA: AMS, 1997.

[41] D. Cai, X. He, J. Han, and T. S. Huang, "Graph regularized nonnegative matrix factorization for data representation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 8, pp. 1548–1560, Aug. 2011.

[42] C. Wang, X. He, J. Bu, Z. Chen, C. Chen, and Z. Guan, "Image representation using Laplacian regularized nonnegative tensor factorization," *Pattern Recognit.*, vol. 44, nos. 10–11, pp. 2516–2526, Oct./Nov. 2011.

[43] M. V. Afonso, J. M. Bioucas-Dias, and M. A. T. Figueiredo, "An augmented Lagrangian approach to the constrained optimization formulation of image inverse problems," *IEEE Trans. Image Process.*, vol. 20, no. 3, pp. 681–695, Mar. 2011.

[44] J. F. Yang and Y. Zhang, "Alternating direction algorithms for ℓ_1-problems in compressive," *SIAM J. Sci. Comput.*, vol. 33, no. 1, pp. 250–278, 2011.

[45] J. Eckstein and D. P. Bertsekas, "On the Douglas-Rachford splitting method and the proximal point algorithm for maximal monotone operators," *Math. Program.*, vol. 55, no. 3, pp. 293–318, Jun. 1992.

[46] G. Liu, Z. Lin, and Y. Yu, "Robust subspace segmentation by low-rank representation," in *Proc. ICML*, 2010, pp. 663–670.

Junmin Liu (M'12) is an Assistant Professor with the Department of Information Science, School of Mathematics and Statistics, Xi'an Jiaotong University, Xi'an, China. He received the Ph.D. degree from Xi'an Jiaotong University in 2013. His research interests are focused on spectral unmixing, pattern recognition, remotely sensed image fusion, and intelligence computation.

Yijun Chen received the B.S. degree from Ningxia University, Yinchuan, China, in 2007. She is currently pursuing the M.S. degree with the School of Electronic and Information Engineering, Xi'an Jiaotong University, Xi'an, China. Her main research interests include data mining and digital library.
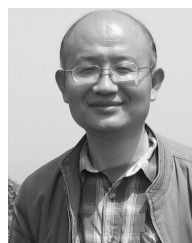
Jiangshe Zhang was born in 1962. He received the M.S. and Ph.D. degrees in applied mathematics from Xi'an Jiaotong University, Xi'an, China, in 1987 and 1993, respectively, where he is currently a Professor with the Department of Statistics. He has authored and co-authored one monograph and over 80 conference and journal publications on robust clustering, optimization, short-term load forecasting for electric power system, and remote sensing image processing. His current research interests include Bayesian learning, global optimization, ensemble learning, and deep learning.

Zongben Xu received the Ph.D. degree in mathematics from Xi'an Jiaotong University, Xi'an, China, in 1987. He currently serves as the Vice President of Xi'an Jiaotong University, the Chief Scientist of the National Basic Research Program of China (973 Project), and the Director of the Institute for Information and System Sciences at Xi'an Jiaotong University. He was a recipient of the National Natural Science Award of China in 2007, the National Award on Scientific and Technological Advances of China in 2011, and the CSIAM Su Buchin Applied Mathematics Prize in 2008. He delivered a 45-min talk on the International Congress of Mathematicians in 2010. He was elected as a member of the Chinese Academy of Science in 2011. His current research interests include intelligent information processing and applied mathematics.