

# Laplacian Regularized Low-Rank Representation and Its Applications

Ming Yin, Junbin Gao, and Zhouchen Lin, *Senior Member, IEEE*

**Abstract**—Low-rank representation (LRR) has recently attracted a great deal of attention due to its pleasing efficacy in exploring low-dimensional subspace structures embedded in data. For a given set of observed data corrupted with sparse errors, LRR aims at learning a lowest-rank representation of all data jointly. LRR has broad applications in pattern recognition, computer vision and signal processing. In the real world, data often reside on low-dimensional manifolds embedded in a high-dimensional ambient space. However, the LRR method does not take into account the non-linear geometric structures within data, thus the locality and similarity information among data may be missing in the learning process. To improve LRR in this regard, we propose a general Laplacian regularized low-rank representation framework for data representation where a hypergraph Laplacian regularizer can be readily introduced into, i.e., a Non-negative Sparse Hyper-Laplacian regularized LRR model (NSHLRR). By taking advantage of the graph regularizer, our proposed method not only can represent the global low-dimensional structures, but also capture the intrinsic non-linear geometric information in data. The extensive experimental results on image clustering, semi-supervised image classification and dimensionality reduction tasks demonstrate the effectiveness of the proposed method.

**Index Terms**—Low-rank representation, graph, Hyper-Laplacian, manifold structure, Laplacian Matrix, regularization

## 1 INTRODUCTION

LOW-RANK representation (LRR) [30], [33], as a promising method to capture the underlying low-dimensional structures of data, has attracted great interest in the pattern analysis and signal processing communities. Specifically, the problems involving the estimation of low-rank matrices have drawn considerable attention in recent years. LRR has been widely used in subspace segmentation [11], [49], image destriping [35], image clustering [51], [56] and video background/foreground separation [1]. The low-rank regularizer in LRR has a deep link with the recent theoretical advances on robust principal component analysis (RPCA) [9], [10], which leads to new and powerful modeling options for many applications.

Under the practical assumption that observed data all lie near some low-dimensional subspaces, the data matrix stacked from all the vectorized observations should be approximately of low rank [44]. The conventional principal component analysis (PCA) [24] is one of ways to recover the best low-rank representation in terms of  $\ell_2$  errors. It is well known that PCA, widely applied in computer vision and

other research areas, is fragile to outliers or sparse/extreme errors introduced by occlusion and disguise [26]. Researchers have long been working on different kinds of approaches aiming to robustify PCA and to recover a low-dimensional subspace from corrupted data, such as  $\ell_1$  PCA [17] and robust PCA (RPCA) [10]. These robust PCA schemes show that  $\ell_1$ -type noise models can offer better robustness than the conventional  $\ell_2$ -type ones. Although most of robust PCA approaches can achieve resilience with respect to grossly corrupted observation data, unfortunately none of them yields a polynomial time algorithm under broad conditions [10]. Since these methods are modified by adding extra regularization terms, the problems usually become non-convex and corresponding optimization algorithms can easily get stuck at local minima, resulting in performance degradation.

The LRR method [11], [30], [31], [33] focuses on low-rank data representation, based on the hypothesis that data approximately jointly span several low-dimensional subspaces. The authors of [32] generalize LRR model to take care of largely contaminated outliers by incorporating a  $\ell_1/\ell_2$  noise model and prove that under mild technical conditions, the LRR model exactly recovers the subspace of samples and detect the outliers as well. Thus LRR can accurately recover the row space of the original data and detect outliers under mild conditions [30]. In general, the resulting problem, which minimizes a combination of the nuclear norm and the  $\ell_1$ -norm, is convex and can be solved in polynomial time [10]. In order to handle the cases where the number of observed data is insufficient or data themselves are too badly corrupted, Liu and Yan [33] further proposed a latent low-rank representation approach. In the latent LRR, hidden data can be regarded as the input data matrix after being transposed. This idea has been recently used in designing a classifier for image classification [4]. As for LRR, only the

- M. Yin is with School of Automation, Guangdong University of Technology, Guangzhou 510006, China, and the Key Laboratory of Autonomous Systems and Networked Control, Ministry of Education, Guangzhou 510640, P.R. China. E-mail: yiming@gdut.edu.cn.
- J. Gao is with the School of Computing and Mathematics, Charles Sturt University, Bathurst, N.S.W. 2795, Australia. E-mail: jbgao@csu.edu.au.
- Z. Lin is with Key Laboratory of Machine Perception (MOE), School of EECS, Peking University, Beijing 100871, P.R. China, and the Cooperative Medianet Innovation Center, Shanghai Jiaotong University, Shanghai 200240, P.R. China. E-mail: zlin@pku.edu.cn.

Manuscript received 10 Dec. 2013; revised 18 Apr. 2015; accepted 20 July 2015. Date of publication 28 July 2015; date of current version 10 Feb. 2016.

Recommended for acceptance by I.S. Dhillon.

For information on obtaining reprints of this article, please send e-mail to: reprints@ieee.org, and reference the Digital Object Identifier below.

Digital Object Identifier no. 10.1109/TPAMI.2015.2462360

row space information is recovered and the column space information of the input data matrix is not sufficiently exploited for learning subspace structures. To address this issue, Yin et al. [49] proposed a novel approach, termed double LRR, which simultaneously learns the row space and column space information embedded in a given dataset.

If all the data in a high-dimensional space actually lie on the union of several linear subspaces, LRR can easily pick up the low-dimensional structures embedded in the data. However, in real world applications the assumption cannot be ensured. For instance, it is well known that facial images are sampled from a non-linear low-dimensional manifold which is embedded in a high-dimensional ambient space [22]. Therefore, in this case, the LRR method may fail to discover the intrinsic geometric and discriminating structures of data [6], [35], [52], which is essential to actual applications.

To preserve local geometric structures embedded in a high-dimensional space, numerous researchers have considered the manifold learning methods, such as the locally linear embedding (LLE) [37], ISOMAP [40], locality preserving projection (LPP) [22], neighborhood preserving embedding (NPE) [21] and Laplacian Eigenmap (LE) [2]. All these algorithms are motivated by the idea of the so-called local invariance [19], which aims to estimate geometric and topological properties of an unknown manifold from random points (scattered data) lying around it.

In practice, it is reasonable to assume that if two data points are close in the intrinsic manifold of the data distribution, then the representations of these two points in a new space are close to each other too [20]. In recent years, this observation inspires Zheng et al. [52] to propose a graph regularized sparse coding to learn a sparse representation that explicitly takes into account the local manifold structures of data. Similarly, Gao et al. [18] also proposed two Laplacian regularized sparse codings, termed as Laplacian sparse coding (LSc) and Hypergraph Laplacian sparse coding (HLSc) by incorporating a similarity preserving term into the objective of sparse coding. In [6], Cai et al. developed a graph based approach for non-negative matrix factorization [25] of data representation in order to address the failure in representing geometric structures in data. To exploit the intrinsic geometry of the data distribution, He et al. [20] proposed a Laplacian regularized Gaussian Mixture Model (LapGMM) based on manifold structures for data clustering. What is more, from the matrix factorization perspective, a novel low-rank matrix factorization model that incorporates manifold regularization into matrix factorization is proposed in [51]. More recently, to comprehensively consider the high spectral correlation between the observed sub-images in different bands and the local manifold structures of the hyperspectral data space, Lu et al. [35] proposed a novel graph-regularized LRR (GLRR) destriping approach by incorporating the LRR technique. Their formulation shares the similar idea as that used in our proposed model, but with the purpose of removing striping noise in hyperspectral images. In contrast, we consider data clustering and classification, and further incorporate both the sparsity and non-negative constraints in our model.

Motivated by the above works, in this paper, we propose a non-negative sparse hyper-Laplacian regularized low-rank representation model, or NSHLRR for short, for image

representation. Note that here we use a hypergraph, instead of a normal graph, to describe similarity structures among data in order to introduce a general manifold regularization to the LRR model. Although the manifolds that data points reside near are non-linear, every data point can still be linearly represented by points in its neighborhood because any manifold can be locally approximated by linear subspaces. So we still can characterize the local geometry of data by the linear coefficients that reconstruct each data point from its neighbors, which are mostly on the same manifold unless the to-be-represented data point is far away along the manifold. Hence the optimal representation by our proposed model is close to block-diagonal, which facilitates non-linear manifold clustering.

In summary, our main contributions in this paper lie in the following three aspects:

- 1) We extend the LRR model by introducing a regularization term based on the manifold structures of data, and propose a novel and more general Laplacian regularized LRR. More specifically, as the data manifolds are unknown *a priori*, the local geometric structures of data are modeled by a nearest neighbour graph and then the graph structures are incorporated into an optimization problem targeting on finding the lowest-rank representation matrix.
- 2) Instead of considering pairwise relations, we model the data manifold structures by a hypergraph to explore the high order relations among data points. It is easy to see that the regularized LRR model under a normal graph structure is actually a special case of our proposed model. Our experimental results show that the proposed model outperforms the state-of-the-art methods in the semi-supervised image classification and clustering tasks.
- 3) We constrain the representation coefficients to be non-negative to facilitate learning local manifold structures. Recently, many literatures [25], [56] have shown that the non-negative constraint leads to a parts-based representation for data, which can make the model more physically interpretable. Since we use locally linear subspaces to approximate the non-linear manifold, it is better if the sample to be represented is at the center of the locally linear subspace so that the approximation is valid. This is distinct from the model proposed in [35].

The remainder of this paper is organized as follows. In Section 2, we give a brief review on related works. Section 3 is dedicated to introducing the proposed NSHLRR. In Section 4, we apply the linearized alternating direction method (ADM) with adaptive penalty (LADMAP) [29] to solve the optimization problem of NSHLRR. Section 5 presents experimental results on image clustering and semi-supervised image classification tasks. The results on applying the proposed model to hypergraph-based applications are provided in Sections 6 and 7. Finally, Section 8 concludes our paper.

## 2 RELATED WORKS

In recent years, the success of low-rank matrix and graph representations has inspired researchers to extend these

representations by jointly considering the two approaches. Before we introduce our model, in this section, we review the recent development of LRR [30], [31], [33] and graph based analysis [2], [47], [52].

## 2.1 Low-Rank Matrix Recovery Formulation

The LRR model [30] is based on the assumption that data are approximately sampled from a union of multiple low-dimensional subspaces. Given a set of data samples, LRR aims at finding the lowest-rank representation of all data jointly. It has been demonstrated that LRR is very effective in exploring low-dimensional subspace structures embedded in data [30].

Consider the case that data  $Y$  are drawn from a union of multiple subspaces given by  $\bigcup_{i=1}^k \mathcal{S}_i$ , where  $\mathcal{S}_1, \mathcal{S}_2, \dots, \mathcal{S}_k$  are low-dimensional subspaces. The LRR model for the given data  $Y$  is defined as the following rank minimization problem

$$\min_{Z,E} \text{rank}(Z) + \lambda \|E\|_0 \text{ s.t. } Y = AZ + E, \quad (1)$$

where the columns of  $A$  are a set of known bases or dictionary items,  $E$  denotes the error components,  $\|\cdot\|_0$  is the  $\ell_0$  pseudo-norm, i.e., the number of nonzeros in a matrix or vector, and  $\lambda$  is a penalty parameter for balancing the low-rank term and the reconstruction fidelity.

As it is difficult to solve the above optimization problem (1) due to the discrete nature of the rank function and the  $\ell_0$  pseudo-norm, a convex relaxation of the above optimization problem is proposed as

$$\min_{Z,E} \|Z\|_* + \lambda \|E\|_1 \text{ s.t. } Y = AZ + E, \quad (2)$$

where  $\|Z\|_*$  is the nuclear norm, defined as the sum of all singular values of  $Z$ , which is the convex envelope of the rank function [9], and  $\|E\|_1$  is the  $\ell_1$  norm, defined as the sum of absolutes of all entries, which is the convex envelope of the  $\ell_0$  pseudo-norm. In fact, it has been proved in [30] that in the noise free case the solution to (2) is also the solution to (1).

Given a low rank solution  $Z$  to (2) or (1), the clean representation  $AZ$  for data  $Y$  is also of low rank. Denote  $X = AZ$ . Instead of solving the optimization problem with respect to the representation coefficient matrix  $Z$ , we may solve for the clean data  $X$  directly. This links LRR with robust PCA [10], whose optimization problem is

$$\min_{X,E} \|X\|_* + \lambda \|E\|_1 \text{ s.t. } Y = X + E. \quad (3)$$

RPCA is broadly used in the computer vision community [46], [50]. RPCA only aims to recover the low-rank clean data from given noisy data. Compared to model (3), the LRR model is able to reveal the implicit data membership. To see this, let  $X = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n]$  be the matrix whose columns are clean data samples drawn from independent subspaces  $\{\mathcal{S}_i\}$ . Given a dictionary  $A = [\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_m]$ , or more generally a dictionary of atoms, linearly spanning the whole data space, we can represent the data as  $X = AZ$ , where  $Z = [\mathbf{z}_1, \mathbf{z}_2, \dots, \mathbf{z}_n]$  is the coefficient matrix with each  $\mathbf{z}_i$  being the representation of  $\mathbf{x}_i$ . Under an appropriately designed dictionary  $A$ , the optimal solution  $Z^*$  of LRR can actually reveal some underlying affinity

between data points, which can be exploited to uncover the intrinsic structures of data. In [30], the dictionary  $A$  is chosen as the given data  $Y$ . In this case, the coefficient matrix element  $z_{ij}$  can be explained as the ‘‘similarity’’ between data points  $\mathbf{y}_i$  and  $\mathbf{y}_j$ .

A number of methods have been proposed for solving low-rank matrix problems, such as singular value thresholding [8], accelerated proximal gradient (APG) [28], and augmented Lagrange Multiplier Method (ALM) [27]. Lin et al. [29] proposed a fast method, termed linearized alternating direction method with adaptive penalty, which uses less auxiliary variables and no matrix inversions, and hence converges faster than the usual alternating direction method [43]. Most recently, in order to handle the multi-block cases, Lin et al. [34] further proposed a new algorithm, called LADMPSAP.

## 2.2 Graph Based Manifold Analysis

Without loss of generality, suppose  $Y = [\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_n]$  are sampled from an underlying manifold  $\mathcal{M}$ . Since a non-linear manifold can be locally approximated by linear subspaces, it is reasonable to assume that the relationship between a data point and its neighbors is linear [3]. Thus, the local geometry of these data points can be effectively characterized by the linear coefficients that reconstruct each data point from its neighbors. Given data  $Y \in \mathbb{R}^{d \times n}$ , we construct a  $k$  nearest neighbor (kNN) graph  $G$  with  $n$  vertices, where each vertex denoting a data point. We also define a symmetric weight matrix  $W \in \mathbb{R}^{n \times n}$ , in which  $W_{ij}$  is the weight of the edge joining vertices  $i$  and  $j$ . That is, a weighted undirected graph  $G = (V, E; W)$  is constructed with the weight matrix  $W$ , where  $V = \{v_i\}_{i=1}^n$  is the vertex set with each node  $v_i$  corresponding to a data point  $\mathbf{y}_i$ , and  $E = \{e_{ij}\}$  is the edge set with each edge  $e_{ij}$  associating nodes  $v_i$  and  $v_j$  with a weight  $W_{ij}$ . The value of  $W_{ij}$  is set as follows:

$$W_{ij} = \begin{cases} 1 & \text{if } \mathbf{y}_i \in N_k(\mathbf{y}_j) \text{ or } \mathbf{y}_j \in N_k(\mathbf{y}_i), \\ 0 & \text{otherwise,} \end{cases} \quad (4)$$

where  $N_k(\mathbf{y}_i)$  denotes the set of  $k$  nearest neighbors of  $\mathbf{y}_i$ .

Graph embedding is based on the natural assumption that if two data points  $\mathbf{y}_i$  and  $\mathbf{y}_j$  are close in the intrinsic geometry of the data distribution, their embedding/mappings in a new space are also close to each other. This assumption can be implemented in terms of similarity. So graph embedding aims at describing each vertex of the graph by a low-dimensional vector that preserves affinity between the vertices, where the similarity is measured by the edge weight. In mathematics, this relationship defined by the manifold assumption is formulated as

$$\min_{\{\mathbf{z}_k\}} \sum_{ij} \|\mathbf{z}_i - \mathbf{z}_j\|^2 W_{ij}, \quad (5)$$

where  $\mathbf{z}_i$  and  $\mathbf{z}_j$  are the mappings of  $\mathbf{y}_i$  and  $\mathbf{y}_j$  under some transformation, respectively. This rule plays a critical role in developing various kinds of algorithms, including dimensionality reduction [47], clustering [5] and semi-supervised learning [56]. The degree matrix is defined as  $D$ , which is a diagonal matrix, whose  $i$ th diagonal entry  $D_{ii}$

corresponds to the summation of all the similarities related to  $\mathbf{y}_i$ , i.e.,  $D_{ii} = \sum_j W_{ij}$ . Then the graph Laplacian matrix is defined by [14]

$$L = D - W.$$

It is easy to prove that the graph embedding (5) can be rewritten as

$$\min \text{tr}(ZZ^T).$$

In many applications, however, the relationship among the data of interest is more complicated than pairwise, which cannot be simply represented by a normal graph. For a long time, the hypergraph relations have often been transformed into a normal graph, which is easier to handle [18], [36], [54]. If this complex relationship is simplified into pairwise one, it is inevitable to lose some useful information for targeted learning tasks. To effectively represent the complex relationship among given data, the concept of hypergraph is employed in manifold learning tasks [54], such as clustering, classification and dimensionality reduction via embedding. In fact, a hypergraph is an extension of a normal graph, in which an edge can “connect” more than two vertices.

For the convenience of subsequent presentation, we here introduce some basic notations on hypergraphs. Given a hypergraph  $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ ,  $\mathcal{V}$  denotes the vertex set and  $\mathcal{E}$  represents the hyperedge set, in which each  $\mathbf{e} \in \mathcal{E}$  is a subset of vertices. The weight corresponding to a hyperedge  $\mathbf{e}$  is denoted by  $W(\mathbf{e})$ , which is usually a non-negative number. The degree of a vertex  $v$  is defined as  $d(v) = \sum_{\mathbf{e} \in \mathcal{E}} W(\mathbf{e})h(v, \mathbf{e})$ , where  $h(v, \mathbf{e})$  is the entry of the incidence matrix  $H$  of  $\mathcal{G}$ . The value of the entry  $h(v, \mathbf{e})$  is set as follows:

$$h(v, \mathbf{e}) = \begin{cases} 1, & \text{if } v \in \mathbf{e}; \\ 0, & \text{otherwise.} \end{cases}$$

Denote the degree of a hyperedge  $\mathbf{e}$  by  $d(\mathbf{e}) = \sum_{v \in \mathcal{V}} h(v, \mathbf{e})$  and the diagonal degree matrix  $D_{\mathcal{E}}$  consists of diagonal entries  $d(\mathbf{e})$ . Similarly, denote  $D_{\mathcal{V}}$  the diagonal matrix whose diagonal entries correspond to the degree of each vertex. Then the unnormalized hyper-Laplacian matrix [54] can be defined as

$$L^h = D_{\mathcal{V}} - HW_{\mathcal{E}}D_{\mathcal{E}}^{-1}H^T, \quad (6)$$

where  $W_{\mathcal{E}}$  is the diagonal matrix whose diagonal entries are the edge weights  $W(\mathbf{e})$ .

### 3 LAPLACIAN REGULARIZED LOW-RANK REPRESENTATION

In this section, a novel LRR model with Laplacian regularization is proposed, in which we consider the local manifold structures of given data. Note that, under the new model, the coefficients  $Z$  not only capture the global structures of the whole data, but preserve the local geometric structures among data points. Different from the graph regularized sparse representation [18], [52], our proposed model can effectively represent data by themselves without relying on a large dictionary.

#### 3.1 Non-Negative Sparse Hyper-Laplacian Regularized LRR

If we use data  $Y$  itself as the dictionary  $A$  in (2), there are two explanations to the coefficient matrix  $Z$ . Firstly, the element  $z_{ij}$  of  $Z$  reflects the “similarity” between data pair  $\mathbf{y}_i$  and  $\mathbf{y}_j$ , hence  $Z$  is sometimes called an affinity matrix. Secondly, we can explain each column  $\mathbf{z}_i$  of  $Z$  as a new representation of the data point  $\mathbf{y}_i$  in terms of other samples in  $Y$ . In other words,  $\mathbf{z}_i$  can be used as a representation for  $\mathbf{y}_i$ .

To introduce richer information over  $Z$ , people consider imposing some helpful regularizations on the coefficient matrix, such as sparsity [16], [45] and a non-negative constraint [25]. For example, given that the sparsity criterion can better capture the local structure around each data vector, a non-negative sparse LRR model can be formulated as follows:

$$\begin{aligned} \min_{Z, E} \|Z\|_* + \lambda \|Z\|_1 + \gamma \|E\|_1, \\ \text{s.t. } Y = YZ + E, Z \geq 0. \end{aligned} \quad (7)$$

In fact, in order to deal with grossly corrupted data, Zhuang et al. [56] proposed a non-negative low-rank and sparse (NNLRS) graph for semi-supervised learning with  $\ell_{2,1}$ -norm for the error term  $E$ .

Motivated by the graph based manifold learning [3], [14], we here can incorporate the Laplacian regularization into the objective function (7) so that similar data points have similar representation coefficients. Specifically denote by  $W$  the similarity measure for the dataset, for example as defined in (4). As the columns of  $Z$  are new representations of the data under certain dictionary, the distance between  $\mathbf{z}_i$  and  $\mathbf{z}_j$  is actually one of dissimilarity measures for the original data points  $\mathbf{y}_i$  and  $\mathbf{y}_j$ . Thus we can add a similarity matching regularization term  $\sum_{i,j} \|\mathbf{z}_i - \mathbf{z}_j\|^2 W_{ij}$  to the objective function (7).

However, to assume the pairwise relationship between complicated data points is far from complete, especially on high order relationship in gene data, web images and co-authorship articles, etc. This relationship cannot be simply represented by a normal graph since the information that helps our grouping task would be lost. Therefore, we consider a more general case, instead of *pairs*, by introducing a hyper-Laplacian regularizer.

According to the hypergraph definition, we enforce the LRR coefficients corresponding to data points within the same hyperedge be similar to each other. Instead of using the normal weight  $W_{ij}$ , we weight the summation of pairwise distances among the given data points within each hyperedge  $\mathbf{e}$  by  $W(\mathbf{e})/d(\mathbf{e})$ .

Then the NSHLRR is formulated as

$$\begin{aligned} \min_{Z, E} \|Z\|_* + \lambda \|Z\|_1 + \beta \sum_{(i,j) \subset \mathbf{e} \in \mathcal{E}} \|\mathbf{z}_i - \mathbf{z}_j\|^2 \frac{W(\mathbf{e})}{d(\mathbf{e})} + \gamma \|E\|_1, \\ \text{s.t. } Y = YZ + E, Z \geq 0. \end{aligned} \quad (8)$$

The non-negative constraint on  $Z$  aims to guarantee that each data point is in the middle of its neighbors, which is more meaningful and can better embody the dependency among the data points than otherwise.

As can be seen, it is easy to transform the formulation of hyper-Laplacian regularizer to the case of graph Laplacian regularizer, when each hyperedge only contains two vertices, i.e.,  $d(\mathbf{e}) = 2$ . In fact, a hypergraph is a generalized model of the normal graph, where the edges are arbitrary non-empty subsets of the vertex set. If the subset of the vertex set is degenerated to only two vertices, we have a special case of (8), i.e., a Non-negative Sparse Laplacian regularized LRR model, termed NSLLRR, defined as follow:

$$\begin{aligned} \min_{Z, E} \|Z\|_* + \lambda \|Z\|_1 + \beta \sum_{ij} \|\mathbf{z}_i - \mathbf{z}_j\|^2 W_{ij} + \gamma \|E\|_1, \\ \text{s.t. } Y = YZ + E, Z \geq 0. \end{aligned} \quad (9)$$

Therefore, we are actually considering a generalized version of the graph regularized LRR model. The advantage of this generalized model is as follows. By introducing a hypergraph regularization term, we can effectively preserve high order affinity among samples belonging to the same hyperedge so as to keep the manifold structures of data. In particular, it is more powerful to explore all-round relations among more data points.

Finally, by some algebraic manipulations we have the matrix form of hyper-Laplacian regularized LRR model

$$\begin{aligned} \min_{Z, E} \|Z\|_* + \lambda \|Z\|_1 + \beta \text{tr}(ZL^h Z^T) + \gamma \|E\|_1, \\ \text{s.t. } Y = YZ + E, Z \geq 0, \end{aligned} \quad (10)$$

where  $L^h$  is the hyper-Laplacian matrix defined in (6) and  $\lambda$ ,  $\beta$  and  $\gamma$  are penalty parameters for balancing the regularization terms.

From (10), it is easy to rewrite the above formulation (9) as

$$\begin{aligned} \min_{Z, E} \|Z\|_* + \lambda \|Z\|_1 + \beta \text{tr}(ZLZ^T) + \gamma \|E\|_1, \\ \text{s.t. } Y = YZ + E, Z \geq 0, \end{aligned} \quad (11)$$

where  $L$  is the Laplacian matrix for the graph built on  $W$  and other parameters are similar to those in the NSHLRR model (10). In this model, we introduce the sum of distance of pairwise LRR coefficients to the traditional LRR model, where the distance is weighted by the similarity between the given data points.

In our model (10), the term  $Y = YZ$  ensures each sample can be linearly represented by its neighbors, i.e., the representation can still be written as  $Y = YZ + E$  by taking noise effect into account. In fact, this term is not contradictory to the assumption of a non-linear manifold. The data itself could reside on a union of non-linear manifolds. However, except near the intersections, the non-linear manifolds can all be locally approximated by linear subspaces, centered at every sample point. In addition, the Laplacian and the sparsity constraints encourage choosing nearby samples (which most likely belong to the same cluster), rather than the far-away samples (which may belong to other clusters), to represent the sample at the center of the locally linear manifold. So the optimum  $Z$  will be close to block-diagonal. Then the subsequent normalized cut (Ncut) can correct some of the errors in  $Z$ , resulting in pleasing manifold clustering results.

As for the non-negative constraint, i.e.,  $Z \geq 0$ , our idea originates from non-negative matrix factorization which shows better results in clustering tasks. It has been noted in many literatures [25], [56] that the non-negative constraint actually leads to a parts-based representation for data which can render the model more physically interpretable. In LRR, the columns of the coefficient matrix  $Z$  can be regarded as the representation of  $\mathbf{y}_i$ 's. Since we use locally linear manifold to approximate the non-linear manifold, it is better that the sample to be represented is at the center of the locally linear manifold so that the approximation is valid. To this end, we require  $Z$  to be non-negative. The advantage of this constraint is also demonstrated by our subsequent experimental results.

In light of the definition of hypergraph, the hyper-Laplacian regularized LRR model is a more general one compared to the Laplacian regularized LRR model. It is obvious that the NSHLRR model (10) becomes the NSLLRR model (11) when there exists a hyperedge for every vertex pair and each hyperedge just includes two vertices. On the other hand, the NSLLRR model (11) can also be changed into the NSHLRR model (10) under some special definition of the weight  $W$  [18]. To this end, we can unify the way to apply some manifold constraint into the LRR framework for better exploiting the geometric structures of data space.

In the above viewpoint, the graph-regularized LRR proposed in [35] is actually a special case of our NSHLRR model merely considering the *pairwise* relation between data points. The similar idea can be seen in another recent work [53]. However, the authors used the second explanation of  $Z$ , i.e.,  $z_{ij}$ , is the similarity between the data points  $\mathbf{y}_i$  and  $\mathbf{y}_j$ . Hence the corresponding regularized term is defined as  $\sum |z_{ij}| \|\mathbf{y}_i - \mathbf{y}_j\|^2$ . It can also be easily extended to incorporate all-round relations.

### 3.2 Laplacian Regularized LRR for Unsupervised Learning

According to the principle of LRR,  $Z$  is actually a new representation of each data point using the dataset itself as the dictionary. Specifically,  $Z$  reflects the affinity between the data pair  $\mathbf{y}_i$  and  $\mathbf{y}_j$ , which can be used for data clustering [33]. In recent years, many similarity-based clustering algorithms have emerged, such as K-means and spectral clustering method [38], which do not require explicit assumptions on the distribution of data. In this regard, our proposed method should greatly benefit image clustering as it can effectively learn similarity among samples by incorporating data neighborhood information.

### 3.3 Laplacian Regularized LRR for Semi-Supervised Learning

When we consider the case that a hyperedge just includes two vertices, NSLLRR not only can perform well in unsupervised learning, but also can be adopted for semi-supervised learning effectively. In the pattern recognition community, semi-supervised learning has been attracting considerable attention over the past decades [48], [56] because labeled samples are hard or expensive to acquire whereas unlabeled ones are easy and inexpensive. Recently, graph-based learning methods have been widely used to

develop high performance algorithms for semi-supervised learning tasks [13], [40], [42], [56].

In general, most of graph-based semi-supervised learning algorithms use a common assumption on the well-posedness of the constructed graph. However, this assumption is not always true because some parameters need manual settings when constructing graphs. Furthermore, the graph construction process is critical for the performance of algorithms. Given that our proposed model well preserves the locality and the similarity of data, we expect that our proposed method also has the potential to convey more discriminative information than the traditional graph-based methods do.

In this section, we present a novel semi-supervised learning framework incorporating the above proposed Laplacian regularized LRR model. The task is to assign the unlabeled samples with proper labels. A popular way is to compute the similarities between samples first, and then propagate the labels from the known samples to unknown ones, using the principle that similar samples are likely to be in the same class. Therefore, one of the critical steps is to compute the affinity matrix of data at first. Then, the objective of graph based semi-supervised learning can be formulated as follows [47]:

$$\min_{\{\mathbf{f}_k\}} \sum_{ij} W_{ij}^l \|\mathbf{f}_i - \mathbf{f}_j\|^2,$$

where  $W_{ij}^l$  is graph weight possibly related to label information,  $\mathbf{f}_i$  and  $\mathbf{f}_j$  denote the probability distributions of  $\mathbf{y}_i$  and  $\mathbf{y}_j$  being different classes, respectively.

Under semi-supervised learning framework, we denote a label matrix  $B = [B_l; B_u] \in \mathbb{R}^{|V| \times c}$  where  $B_l$  represents the label information for the labeled data,  $B_u$  is the information for the unlabeled data and  $c$  denotes the number of the classes. If a sample  $\mathbf{y}_i$  is associated with label  $k \in \{1, \dots, c\}$ , then  $B_{ik} = 1$ . Otherwise,  $B_{ik} = 0$ . Furthermore, we can define a classification function  $F = [F_l; F_u] \in \mathbb{R}^{|V| \times c}$ .  $F_l$  and  $F_u$  include the class probability vectors for the labeled and unlabeled samples, respectively. Thus, mathematically the objective of graph based semi-supervised learning is readily transferred into

$$\min_{\{\mathbf{f}_k\}} \sum_{ij} W_{ij}^l \|\mathbf{f}_i - \mathbf{f}_j\|^2 = \text{tr}(F^T L_w F), \quad (12)$$

where  $L_w$  is the graph Laplacian matrix corresponding to a weighted graph  $G^l = (V, E; W^l)$ . The graph  $G^l$  is constructed via using relationship among the labeled and unlabeled data points, where edge weights encode the similarity between samples. To achieve this target, our work focuses on how to effectively learn the graph. Specifically, in this paper, instead of explicitly designing weights  $W^l$ , we directly choose  $L_w$  to be the Laplacian of the symmetrized affinity matrix  $\frac{1}{2}(|Z| + |Z^T|)$ , where  $Z$  is learned by the Laplacian regularized LRR method. In other words, the graph we are using is the one learned from the data.

In order to efficiently predict the labels of unlabeled samples, we adopt a Gaussian field, rather than a random field, over the label set as the basis for semi-supervised classification. That is, we form a Gaussian field to assign a

probability distribution of  $\mathbf{f}_i$ . Then the minimum energy function (12) is harmonic. Namely, it should simultaneously satisfy  $L_w F = 0$  on unlabeled data points, denoted by  $(L_w F)_u = 0$ , and  $F_l = B_l$  on labeled ones. Please refer to [55] for more details.

Based on the above analysis, we finally learn a Gaussian Harmonic Function (GHF) [55], which combines Gaussian field and harmonic function, to realize the label propagation, formulated as follows:

$$\min_{F \in \mathbb{R}^{|V| \times c}} \text{tr}(F^T L_w F), \text{ s.t. } (L_w F)_u = 0, F_l = B_l.$$

In our case, the matrix  $L_w$  is positively definite, hence the above question is convex. The solution is given by  $F_u = -\Delta_{uu}^{-1} \Delta_{ul} F_l$  through separating  $L_w$  into blocks for labeled and unlabeled nodes [55], such as

$$L_w = \begin{bmatrix} \Delta_{ll} & \Delta_{lu} \\ \Delta_{ul} & \Delta_{uu} \end{bmatrix}.$$

Hence an error rate of graph based semi-supervised learning method can be computed using the propagated label information. The above algorithm is actually a generalization of the case of binary classes in [55] to the multiple classes. The theoretical foundation is based on the maximum principle of harmonic function [15], and the solution is unique and satisfies  $0 < F_u(j) < 1$  for each unlabeled data point  $\mathbf{y}_j$ .

#### 4 LADMAP FOR SOLVING LAPLACIAN REGULARIZED LRR

In recent years, a lot of algorithms have been proposed for solving low-rank optimization problems [28], [29], [39], [41], [44]. In particular, the alternating direction method (ADM) has drawn considerable attention [43]. ADM updates the variables alternately by minimizing the augmented Lagrangian function in a Gauss-Seidel manner. We also apply ADM to solve our optimization problems. Unfortunately, directly applying ADM to solve (10) results in sub-problems as follows

$$\min_Z \|Z\|_* + \frac{\lambda}{2} \|\mathcal{C}(Z) - D\|_F^2, \quad (13)$$

where  $\mathcal{C}$  is a linear mapping and  $\|\cdot\|_F$  is the matrix Frobenius norm defined as  $\|Y\|_F^2 = \sum_{i=1}^m \sum_{j=1}^n |Y_{ij}|^2$ . When  $\mathcal{C}$  is the identity mapping, (13) has a closed form solution [8]. However, when  $\mathcal{C}$  is not the identity mapping, (13) can only be solved iteratively. For example, Lu et al. [35] solved (13) by the accelerated gradient method [23], making the overall algorithm inefficient.

To remedy this issue in solving (10), we adopt the Linearized ADM with Adaptive Penalty [29] instead. We first introduce an auxiliary variable  $J$  in order to make the objective function of (10) separable. Thus the optimization problem can be rewritten as follows:

$$\begin{aligned} \min_{Z, J, E} & \|Z\|_* + \lambda \|J\|_1 + \beta \text{tr}(ZL^h Z^T) + \gamma \|E\|_1, \\ \text{s.t. } & Y = YZ + E, Z = J, J \geq 0. \end{aligned} \quad (14)$$

The augmented Lagrangian function of problem (14) is

$$\begin{aligned} \mathfrak{L}(J, E, Z, Y_1, Y_2) &= \|Z\|_* + \lambda\|J\|_1 + \beta\text{tr}(ZL^h Z^T) + \gamma\|E\|_1 \\ &\quad + \langle M_1, Y - YZ - E \rangle + \langle M_2, Z - J \rangle \\ &\quad + \frac{\mu}{2}(\|Y - YZ - E\|_F^2 + \|Z - J\|_F^2), \end{aligned} \quad (15)$$

where  $M_1$  and  $M_2$  are Lagrange multipliers and  $\mu > 0$  is a penalty parameter. The variables are updated alternately by minimizing the augmented Lagrangian function, with other variables fixed. If the augmented Lagrangian function is difficult to minimize with respect to a variable, the smooth component of it can be linearized. Hence this algorithm bears the name Linearized Alternating Direction Method [29]. Then the Lagrange multipliers are updated using the feasibility errors. The iteration stops when the convergence conditions are met. We provide details of solving (10) with LADMAP in the following.

#### 4.1 Updating $Z$

Updating  $Z$  by minimizing  $\mathfrak{L}(J_k, E_k, Z, M_1^k, M_2^k)$  is equivalent to minimizing the following objective function:

$$\begin{aligned} \mathfrak{L}_1 &= \|Z\|_* + \beta\text{tr}(ZL^h Z^T) + \frac{\mu}{2}\left\|Z - J_k + \frac{1}{\mu}M_2^k\right\|_F^2 \\ &\quad + \frac{\mu}{2}\left\|Y - YZ - E_k + \frac{1}{\mu}M_1^k\right\|_F^2, \end{aligned} \quad (16)$$

which does not have a closed-form solution. By the spirit of LADMAP [29], we denote the smooth component of  $\mathfrak{L}_1$  by

$$\begin{aligned} q(Z, E_k, J_k, M_1^k, M_2^k) &= \beta\text{tr}(ZL^h Z^T) + \frac{\mu}{2}\left\|Z - J_k + \frac{1}{\mu}M_2^k\right\|_F^2 \\ &\quad + \frac{\mu}{2}\left\|Y - YZ - E_k + \frac{1}{\mu}M_1^k\right\|_F^2. \end{aligned}$$

Then according to LADMAP, minimizing  $\mathfrak{L}_1$  can be replaced by solving the following problem:

$$\min_Z \|Z\|_* + \langle \nabla_Z q(Z_k), Z - Z_k \rangle + \frac{\eta_1}{2}\|Z - Z_k\|_F^2, \quad (17)$$

where  $q(Z, E_k, J_k, M_1^k, M_2^k)$  is approximated by its linearization  $\langle \nabla_Z q(Z_k), Z - Z_k \rangle$  at  $Z_k$  plus a proximal term  $\frac{\eta_1}{2}\|Z - Z_k\|_F^2$  and  $\nabla_Z q(Z_k)$  is the gradient of  $q$  w.r.t.  $Z$ . As long as  $\eta_1 > 2\beta\|L\|_2 + \mu(1 + \|Y\|_2^2)$ , where  $\|\cdot\|_2$  is the spectral norm of a matrix, i.e., the largest singular value, the above replacement is valid. Then (17) has a closed-form solution given by

$$Z_{k+1}^* = \Theta_{(\eta_1)^{-1}}(Z_k - \nabla_Z q(Z_k)/\eta_1), \quad (18)$$

where  $\Theta_\varepsilon(A) = US_\varepsilon(\Sigma)V^T$  is the singular value thresholding operator (SVT) [8], in which  $U\Sigma V^T$  is the singular value decomposition (SVD) of  $A$  and  $\mathcal{S}_\varepsilon(x) = \text{sgn}(x)\max(|x| - \varepsilon, 0)$  is the soft thresholding operator [8].

#### 4.2 Updating $J$ and $E$

We view  $(J, E)$  as a larger block of variables. We can update  $(J, E)$  by minimizing  $\mathfrak{L}(J, E, Z_{k+1}, M_1^k, M_2^k)$ , which

naturally splits into to subproblems for  $J$  and  $E$ , respectively, as  $J$  and  $E$  are independent of each other in this minimization problem.

After some simple manipulation, the problem for updating  $E$  is

$$\min_E \gamma\|E\|_1 + \frac{\mu}{2}\left\|E - \left(Y - YZ + \frac{1}{\mu}M_1^k\right)\right\|_F^2, \quad (19)$$

which has the following closed-form solution [8]

$$E_{k+1} = \mathcal{S}_{\frac{\gamma}{\mu}}\left(Y - YZ_{k+1} + \frac{1}{\mu}M_1^k\right). \quad (20)$$

Similarly, the problem for updating  $J$  is

$$\min_{J \geq 0} \lambda\|J\|_1 + \frac{\mu}{2}\left\|J - \left(Z + \frac{1}{\mu}M_2\right)\right\|_F^2, \quad (21)$$

which has the following closed-form solution [8]:

$$J_{k+1} = \max\left\{\mathcal{S}_{\frac{\lambda}{\mu}}\left(Z_{k+1} + \frac{1}{\mu}M_2^k\right), 0\right\}. \quad (22)$$

For faster convergence,  $\mu$  can be adjusted using the adaptive updating strategy as suggested in [29] (see step 5) in Algorithm 1). The detailed procedure of solving the proposed hyper-Laplacian regularized problem is described in Algorithm 1. Note that the Laplacian regularized LRR problem 11 can also be efficiently solved by Algorithm 1 simply replacing  $L^h$  with  $L$ .

#### 4.3 Convergence and Complexity Analysis

The algorithm described above converges to a globally optimal solution of (10) as it is a direct application of LADMAP.

When updating  $Z$  by singular value thresholding, see (18), we may predict the rank  $r$  of  $Z_{k+1}$  using the rank prediction strategy described in [27], [41], which grows from a small value and stabilizes at the true rank when the iteration goes on. Moreover, as in [29] we may use the Lanczos method to compute the leading singular values and singular vectors, which only requires multiplication of  $\tilde{Z}_k = Z_k - \nabla_Z q(Z_k)/\eta_1$ , and its transpose, with vectors, which can be efficiently computed by successive matrix-vector multiplication, rather than forming  $\tilde{Z}_k$  explicitly and then multiplying it with vectors. With such treatments, the complexity of each iteration of Algorithm 1 is  $O(rn^2)$ , where  $n$  is the number of samples.

### 5 GRAPH REGULARIZED LRR FOR IMAGE CLUSTERING AND IMAGE CLASSIFICATION

Firstly, we consider the special case of our proposed model, i.e., normal graph based LRR, in this section. That is, each hyperedge for vertex pairs only includes two vertices and our model particularly refers to NSLLRR in this section. In order to investigate the performance of our approach, we conducted comprehensive experiments on both unsupervised learning (image clustering) and semi-supervised

learning (image classification) tasks. All of the experiments were done on a PC with an Intel i3 3.30 GHz CPU and 8 GB memory, running Windows 7.

---

**Algorithm 1.** LADMAP for Solving Non-Negative Sparse Laplacian Regularized LRR

---

**Input:**  $Y$ ,  $\lambda$ ,  $\beta$ ,  $\gamma$  and the number of nearest neighbors.

**Output**  $Z_{k+1}, E_{k+1}$ .

**Initialization:** Compute  $L^h$ ,  $Z_0 = E_0 = J_0 = M_1^0 = M_2^0 = 0$ ,  $\lambda = 0.02$ ,  $\beta = 1.0$ ,  $\gamma = 5.0$ ,  $\rho_0 = 2.5$ ,  $\mu_0 = 10^{-6}$ ,  $\mu_{\max} = 10^6$ ;  $\varepsilon_1 = 10^{-6}$ ,  $\varepsilon_2 = 10^{-2}$ .

**While** not converged ( $k = 0, 1, \dots$ ) **do**

1) update  $Z$ :

$$Z_{k+1} = \Theta_{(\eta_1)^{-1}}(Z_k - \nabla_Z q(Z_k) / \eta_1),$$

where

$$\begin{aligned} \nabla_Z q(Z_k) &= \beta(Z_k L^{hT} + Z_k L^h) + \mu_k \left( Z_k - J_k + \frac{M_2^k}{\mu_k} \right) \\ &\quad + \mu_k Y^T \left( Y Z_k - Y + E_k - \frac{M_1^k}{\mu_k} \right), \\ \eta_1 &= 2\beta \|L^h\|_2 + \mu_k(1 + \|Y\|_2^2), \end{aligned}$$

2) update  $E$ :

$$E_{k+1} = S_{\frac{\gamma}{\mu_k}} \left( Y - Y Z_{k+1} + \frac{1}{\mu_k} M_1^k \right),$$

3) update  $J$ :

$$J_{k+1} = \max \left\{ S_{\frac{\lambda}{\mu_k}} \left( Z_{k+1} + \frac{1}{\mu_k} M_2^k \right), 0 \right\},$$

4) update Lagrange multipliers  $M_1$  and  $M_2$ :

$$\begin{aligned} M_1^{k+1} &= M_1^k + \mu_k(Y - Y Z_{k+1} - E_{k+1}), \\ M_2^{k+1} &= M_2^k + \mu_k(Z_{k+1} - J_{k+1}), \end{aligned}$$

5) update  $\mu$ :

$$\mu_{k+1} = \min(\mu_{\max}, \rho_k \mu_k),$$

where

$$\rho_k = \begin{cases} \text{if } \max\{\eta_1 \|Z_{k+1} - Z_k\|, \\ \mu_k \|J_{k+1} - J_k\|, \mu_k \|E_{k+1} - E_k\|\} \leq \varepsilon_2, \\ \rho_0, \\ 1, \text{ otherwise,} \end{cases}$$

6) check convergence: if

$$\begin{aligned} \|Y - Y Z_{k+1} - E_{k+1}\| / \|Y\| < \varepsilon_1 \text{ and} \\ \max\{\eta_1 \|Z_{k+1} - Z_k\|, \mu_k \|J_{k+1} - J_k\|, \mu_k \|E_{k+1} - E_k\|\} < \varepsilon_2, \end{aligned}$$

then stop.

**End while**

---

## 5.1 A Synthetic Example

Firstly, we test with a synthetic dataset. We consider a set of data points constructed in two moons pattern, as shown in Fig. 1a. There are two natural clusters, i.e., the

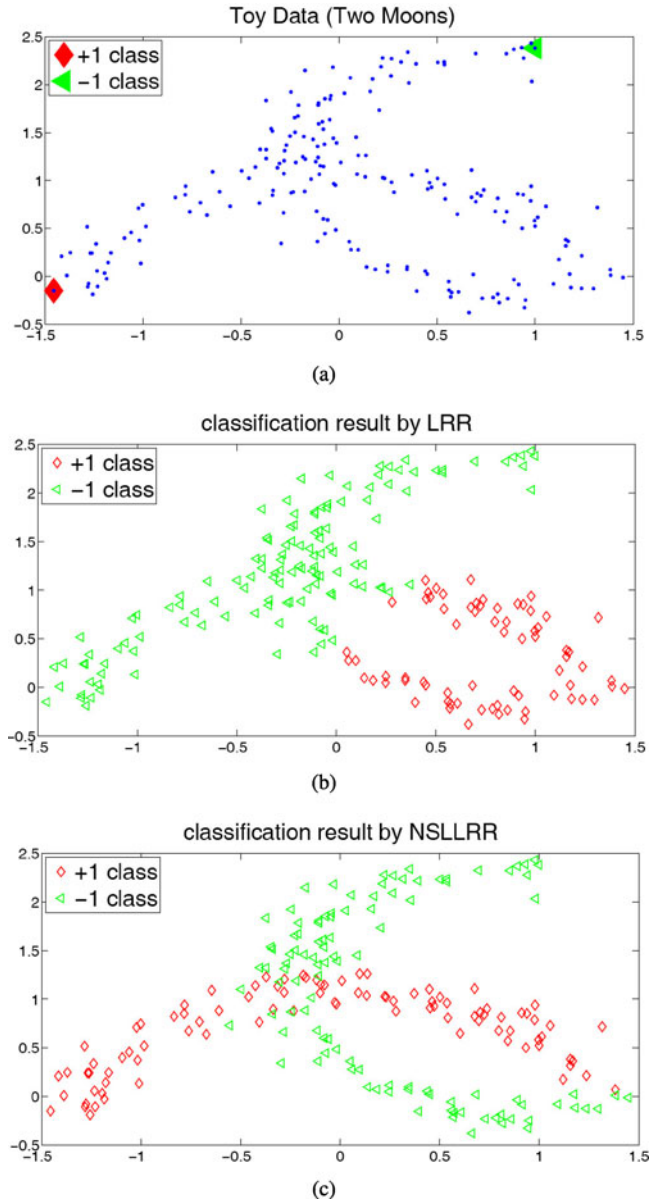


Fig. 1. Clustering on a toy dataset. (a) A synthetic dataset with two intersecting half moons. (b) Clustering result given by LRR, whose accuracy is 57 percent. (c) Clustering result given by NSLLRR, whose accuracy is 96.5 percent.

two half moons intersecting each other. By using the conventional LRR method, we can achieve the result shown in Fig. 1b where the algorithm failed to distinguish the two half moons and the clustering accuracy is only 57 percent. By using our proposed Laplacian regularized LRR, i.e., NSLLRR, we can better separate the mixed clusters. As illustrated in Fig. 1c, the clustering accuracy is as high as 96.5 percent. This clearly demonstrates that the local structural information coded in the Laplacian regularizer is helpful for non-linear manifold clustering. In particular, this pleasing result is beneficial from the block-diagonal structure of the affinity matrix. As shown in Fig. 2, we can see the affinity matrix obtained by NSLLRR has a relative distinct block-diagonal structure which is helpful for data clustering. Note that the affinity matrices in Fig. 2 are both permuted according to the order of two true clusters.



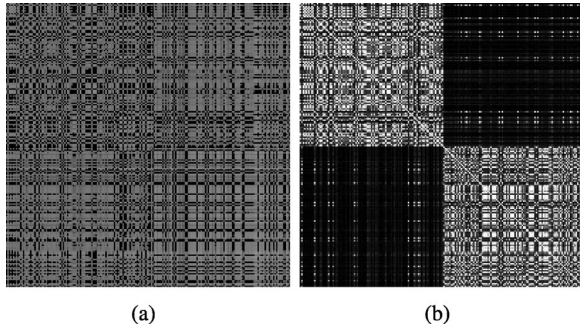


Fig. 2. Illustration of the affinity matrix produced by different methods. a) LRR. b) NSLLRR.

## 5.2 Unsupervised Learning: Image Clustering

Data clustering is to group samples into different groups. To quantitatively evaluate the clustering performance, we adopt two metrics, accuracy (AC) and normalized mutual information (NMI) [7], in our experiments. Given a data point  $\mathbf{x}_i$ , let  $\mathcal{F}$  and  $\hat{\mathcal{F}}$  be the ground truth label and the label produced by a clustering approach, respectively, then the AC measure is defined by

$$AC = \frac{\sum_{i=1}^n \delta(\hat{\mathcal{F}}(i), \text{Match}_{(\hat{\mathcal{F}}, \mathcal{F})}(i))}{n},$$

where  $n$  is the number of samples in total, function  $\delta(\mathbf{a}, \mathbf{b})$  is set to 1 if and only if  $\mathbf{a} = \mathbf{b}$ , and  $\text{Match}(\cdot)$  is the best matching function that permutes  $\hat{\mathcal{F}}$  to match  $\mathcal{F}$ , which is usually fulfilled by the Kuhn-Munkres algorithm [12].

The NMI measure between two index sets  $K$  and  $K'$  is defined as

$$\text{NMI}(K, K') = \frac{\text{MI}(K, K')}{\max(\text{H}(K), \text{H}(K'))},$$

where  $\text{H}(K)$  and  $\text{H}(K')$  denote the entropy<sup>1</sup> of  $K$  and  $K'$ , respectively, and

$$\text{MI}(K, K') = \sum_{y \in K} \sum_{x \in K'} p(x, y) \log_2 \left( \frac{p(x, y)}{p(x)p(y)} \right),$$

where  $p(y)$  and  $p(x)$  denote the marginal probability distribution functions of  $K$  and  $K'$ , respectively, and  $p(x, y)$  is the joint probability distribution function of  $K$  and  $K'$ . Usually,  $\text{NMI}(K, K')$  ranges from 0 to 1, for which the value 1 means that the two clusters are identical and the value 0 means that two are independent. Different from AC, NMI is invariant with the permutation of labels. Namely, it does not require the matching between two clusters in advance.

To extensively assess the clustering performance of our proposed approach, the following five methods for image clustering are compared with:

- 1) K-means clustering algorithm (K-means);
- 2) Normalized cut [38];
- 3) Principle component analysis;
- 4) Traditional LRR [31];
- 5) Locally linear embedding [42].

1. The entropy  $\text{H}(K)$  is defined as  $\text{H}(K) = \sum_{y \in K} -p(y) \cdot \log_2(p(y))$  where  $p(y)$  is the probability that a sample belongs to a cluster.

K-means usually serves as a baseline of image clustering. PCA is a well known unsupervised dimensionality reduction method, which is applied to discard the noisy information corresponding to the small eigenvalues of the data covariance matrix. The Ncut method is a spectral clustering based algorithm which can be used to seek the cluster indicator information for data. Both the traditional LRR and NSLLRR algorithms can learn an affinity matrix measuring the relationship among data points. However, the clustering information, obtained by above methods except K-means, cannot be directly used as the final clustering results because they lack the clear cluster structures. Therefore, here K-means is adopted to assign final labels on the learned low dimensional representation space. The NSLLRR can be regarded as a non-linear manifold clustering algorithm. So it would be informative to compare it with some existing graph-based clustering methods. Here, we compare it with a classical method, locally linear embedding [42], to demonstrate the clustering effectiveness of our proposed method.

In this experiment, the real world CMU-PIE<sup>2</sup> face image database is utilized to compare the performance of related approaches. CMU-PIE is a popular face database, which is widely used in many kinds of learning tasks. It includes 68 subjects with 41,368 face images in total. In this dataset, the size of each sample is  $32 \times 32$  pixels and images of each subject are captured under 13 different poses, 43 different illumination conditions and four different expressions. We only select part of images by fixing the pose and expression so that, for each subject, we have 21 images under different lighting conditions. For computational convenience, we first apply PCA to reduce the data dimensionality to 64. The clustering experiments are conducted with different cluster numbers. That is, we use the first  $k$  classes in the dataset for the corresponding data clustering experiments. The detailed clustering results are reported in Table 1. The bold numbers denote the best result at corresponding cluster numbers. We can see that our proposed method outperforms other algorithms, especially when the number of clusters is large, which is more difficult. It also shows that K-means is generally worse than other methods. As can be seen, the results by our proposed NSLLRR are consistently much better in terms of the NMI values when compared with the LLE method. However, it can be observed that there exists a fluctuation in clustering accuracy and NMI from the proposed method. The phenomenon could be due to applying K-means, whose result may be different under different initializations of cluster centers, to the learned low dimensional representation during the clustering step. The similar phenomenon occurs for the LLE method and LRR as well.

There are several regularization parameters affecting the performance of NSLLRR. In the following, we study the influence of parameters  $\lambda$ ,  $\gamma$  and  $\beta$  on NMI by setting them at different values. We vary a parameter while keeping others fixed. The NMI results with different parameter settings on the CMU-PIE dataset are given in Fig. 3. From the results, we can conclude that our proposed Laplacian regularized LRR outperforms other algorithms over a large

TABLE 1  
Clustering Results on the CMU-PIE Data Set (# Is the Number of Clusters)

Cluster Num. #	Accuracy						NMI					
	K-means	PCA	Ncut	LLE-graph	LRR	NSLLRR	K-means	PCA	Ncut	LLE-graph	LRR	NSLLRR
4	48.60	52.30	99.10	88.10	100	<b>100</b>	46.81	46.81	98.60	77.33	100	<b>100</b>
12	41.60	57.14	86.40	63.89	<b>89.68</b>	76.98	52.66	52.66	91.20	82.10	<b>95.35</b>	87.81
20	38.40	36.40	78.30	73.57	81.19	<b>86.19</b>	59.12	58.88	88.60	86.95	90.74	<b>94.46</b>
28	35.40	34.80	78.20	78.74	77.21	<b>86.91</b>	60.48	61.06	89.60	90.30	89.89	<b>93.99</b>
36	34.90	34.6	75.60	75.00	68.92	<b>76.59</b>	61.69	60.56	88.90	91.13	82.15	<b>93.55</b>
44	33.20	33.70	74.20	68.18	71.86	<b>80.20</b>	62.30	62.57	89.50	88.27	84.16	<b>93.70</b>
52	33.20	33.70	72.60	<b>74.45</b>	71.25	74.18	63.44	61.58	89.10	89.39	84.81	<b>93.01</b>
60	33.10	33.20	71.10	69.92	65.87	<b>78.73</b>	64.06	63.58	88.50	88.59	80.39	<b>93.43</b>
68	31.70	32.90	70.80	69.91	65.27	<b>85.08</b>	63.30	66.60	88.30	89.92	79.12	<b>96.62</b>

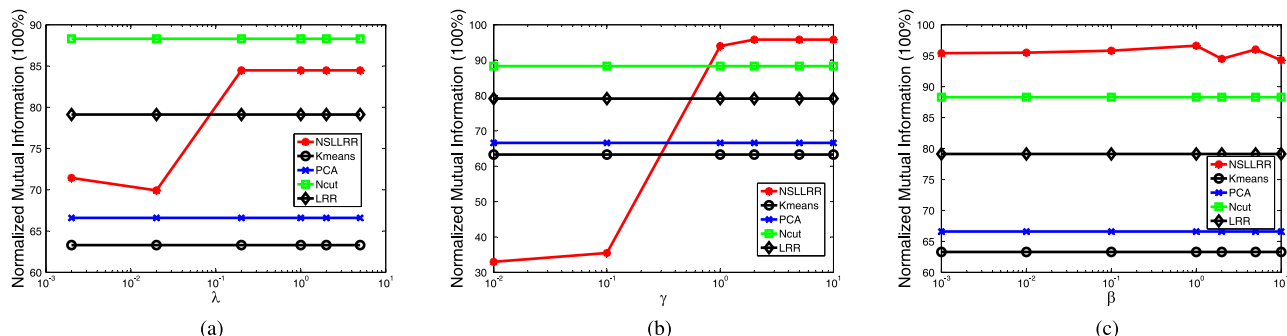


Fig. 3. Performance of NSLLRR on the CMU-PIE face database, with different parameter settings. a)  $\lambda$  varies. b)  $\gamma$  varies. c)  $\beta$  varies.

range of parameter values. We can also note that the performance of NSLLRR is insensitive to the choice of  $\beta$ . Moreover, the results on  $\lambda$  and  $\gamma$  show that both sparsity and low-rankness are critical for data clustering. Thus, we set  $\lambda = 0.2, \gamma = 5$  and  $\beta = 5$  in our experiments.

### 5.3 Semi-Supervised Classification

In order to evaluate our proposed model on semi-supervised learning task, we select two publicly available image datasets for our experiments, i.e., CMU-PIE and USPS.<sup>3</sup> CMU-PIE is a face database and USPS is a handwritten digit database. CMU-PIE includes many face images under different illumination conditions and with different expressions, as described in the last section. USPS is composed of 9,298 handwritten digit images of size  $16 \times 16$  pixels. Each image is represented by a 256-dimensional vector. Some samples of these datasets are shown in Fig. 4.

In order to show the efficacy of NSLLRR, the following five methods for image classification are chosen for performance comparison.

- 1)  $k$ -nearest-neighbor based classification method [40]: We adopt the euclidean distance as the similarity measure and set the numbers of nearest neighbors to be 5.
- 2) LLE-graph based classification method [42]: We construct an LLE-graph with the numbers of nearest neighbors being 8.
- 3)  $\ell_1$ -graph based classification method [13]: Cheng et al. proposed to construct a robust and datum-adaptive

graph by solving the sparsest representation of every sample, using other samples as a dictionary. The coefficient matrix is then symmetrized.

- 4) LRR-graph based classification method: Following [30], we construct the LRR-graph by solving the



(a) CMU-PIE samples.



(b) USPS samples.

Fig. 4. Samples of test databases.

3. <http://www-i6.informatik.rwth-aachen.de/keysers/usps.html>

TABLE 2  
Classification Error Rates (Percent) on CMU-PIE, Based on Various Graphs under Different Percentages of Labeled Samples

Labeling percentage#	kNN [40]	LLE-graph	$\ell_1$ graph	LRR-graph [30]	NNLRS-graph [56]	NSLLRR-graph
10	34.84	33.06	22.88	47.30	<b>11.11</b>	35.66
20	37.46	35.05	22.94	21.60	22.81	<b>9.28</b>
30	35.3	32.52	22.33	11.80	17.86	<b>5.83</b>
40	35.81	32.51	23.14	7.80	16.25	<b>3.36</b>
50	34.39	31.41	23.01	4.30	19.25	<b>2.10</b>
60	35.63	32.76	25.76	2.50	21.56	<b>1.62</b>

TABLE 3  
Classification Error Rates (Percent) on USPS, Based on Various Graphs under Different Percentages of Labeled Samples

Labeling percentage#	kNN [40]	LLE-graph	$\ell_1$ graph	LRR-graph [30]	NNLRS-graph [56]	NSLLRR-graph
10	11.97	17.10	43.27	67.00	11.57	<b>10.55</b>
20	12.31	22.92	41.27	48.30	9.30	<b>5.90</b>
30	5.88	21.26	38.31	33.60	4.95	<b>4.80</b>
40	7.87	19.21	34.86	25.80	7.44	<b>4.01</b>
50	17.19	18.41	29.42	15.50	11.27	<b>3.20</b>
60	11.04	14.80	23.36	10.00	6.09	<b>2.05</b>

lowest-rank coefficient matrix and then symmetrizing it as  $\ell_1$ -graph does. The parameters of LRR are the same as those in [30].

- 5) Non-negative low-rank and sparse-graph based classification method: Following [56], we construct a graph for which the weights are obtained by seeking a non-negative low-rank and sparse matrix that represents each data points as a linear combination of others. The coefficient matrix is also symmetrized.

Similar to the LRR-graph based method, our proposed approach also seeks a low-rank coefficient matrix. We further require that the coefficient matrix is sparse and nonnegative. We call our approach as the NSLLRR-graph based classification method in our experiments.

The purpose of semi-supervised learning task is to assign labels to unlabeled data, according to the labeled data and the overall data distribution. We make the percentage of labeled samples range from 10 to 60 percent. Empirically, we found that  $\lambda = 0.08$ ,  $\gamma = 2$  and  $\beta = 5$  are good choices for these parameters. Therefore, we did not tune these parameters in the experiments. The classification results on CMU-PIE and USPS are reported in Tables 2 and 3, respectively. The bold numbers in each table denote the best performance under corresponding labeling percentages. From these results, we can see that our proposed NSLLRR almost consistently outperforms other methods, especially on the CMU-PIE database. This suggests that our proposed method is effective for semi-supervised learning. This also shows that the manifold regularizer can enhance the robustness of the LRR graph-based method.

## 6 HYPERGRAPH REGULARIZED LRR FOR SEMI-SUPERVISED LEARNING

In the real world, there usually exists a co-occurrence relationship that involves more than two samples [36]. This can be more effectively represented by a hypergraph. In this section, we consider some categorical data whose samples have one or more attributes, i.e., the Zoo dataset, which comes

from the UCI Machine Learning Depository.<sup>4</sup> The Zoo dataset covers 101 animals with 17 Boolean-valued attributes, where the attributes contain **hair**, **feathers**, **eggs**, **milk**, **legs**, **tail**, etc. Moreover, these animal data have been manually grouped into seven classes beforehand. Note that we have found that there are two repeated samples of “frog” in the dataset. So we manually remove the repeated items in our experiment. That is, the test dataset contains 100 samples in our experiment. In order to evaluate our proposed hypergraph regularized LRR, we conduct a semi-supervised classification task on the Zoo dataset, where we define a hypergraph with 36 hyper-edges.

As in [36], we construct a hypergraph by taking samples as vertices and creating a hyperedge for each value of the attributes of Zoo. For instance, attribute 2 (hair) is a boolean variable. So two hyperedges are created for this attribute: F020 and F021 (F denotes feature, 02 represents the second attribute and 0 or 1 corresponds to the value of the attribute). Then, each hyperedge includes the instances that share the same attribute value. After this pre-processing, we find that this data set contains 36 hyperedges.

In our experiment, the labels of some vertices in the hypergraph are supposed to be known. For fair comparison, the labeled vertices are randomly chosen to ensure that all labels appear. That is, the labeled data should cover all classes in the entire dataset. The labeling percentage ranges from 10 to 60 percent. For each labeling percentage, we run 10 times for each algorithm with the labeled data sets randomly chosen and then average the error rates. We empirically set the parameters as  $\lambda = 0.01$ ,  $\gamma = 0.1$  and  $\beta = 0.001$ . To also assess the effectiveness of learning the lowest rank representation, we compare our proposed method, NSHLRR, with LRR-graph methods and normal graph based NSLLRR. Moreover, we compare with kNN-graph based method which is usually used as a baseline. As before, we also select five nearest neighbors for each data point.

4. <http://archive.ics.uci.edu/ml/datasets/Zoo>

TABLE 4  
Classification Error Rates (Percent) on the Zoo Dataset,  
Based on Various Methods under Different  
Percentages of Labeled Samples

Labeling percentage#	kNN-graph	LRR-graph [30]	normal NSLLRR	hyper NSHLRR
10	68.00	43.01	41.00	<b>27.00</b>
20	56.00	36.00	17.00	<b>13.00</b>
30	50.00	26.00	<b>11.00</b>	15.00
40	38.00	16.00	<b>10.00</b>	12.00
50	33.00	11.00	11.00	<b>9.00</b>
60	25.00	10.00	7.00	<b>7.00</b>

The classification results are reported in Table 4 and Fig. 5 shows the plot of classification error rate versus the percentage of labeled points on the Zoo dataset. As can be seen, the hypergraph regularized NSHLRR method consistently outperforms LRR and the kNN graph based method. Among these methods, kNN graph based method produces the worst classification performance. NSHLRR is significantly better than all other methods except NSLLRR in the cases of 30 and 40 percent labeling percentages. This result shows that a hypergraph is more sensible to modeling data with more complex relationship than a normal graph approach does, since the complicated interactions among data can be well represented by a hypergraph. The experiment demonstrates that our proposed Laplacian regularized LRR model can not only preserve the globally non-linear structures, but also represent the locally linear structures, by incorporating a hypergraph regularizer, and has the capability of learning high order relationship (i.e., three-way or higher) in the semi-supervised settings.

### 7 DIMENSIONALITY REDUCTION THROUGH HYPERGRAPH EMBEDDING

In this section, we consider dimensionality reduction task via hypergraph embedding that projects the vertices of  $\mathcal{G}$  onto a low dimensional euclidean space, where the vertices are best characterized by the similarity relationship between the vertex pairs in hypergraphs [36]. The general graph-spectral approach for embedding is implemented by Laplacian

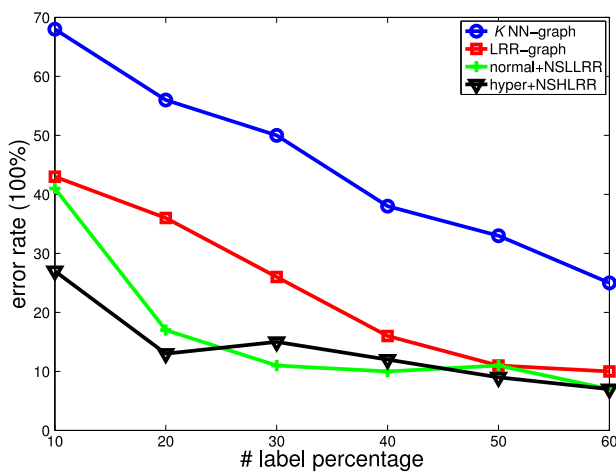
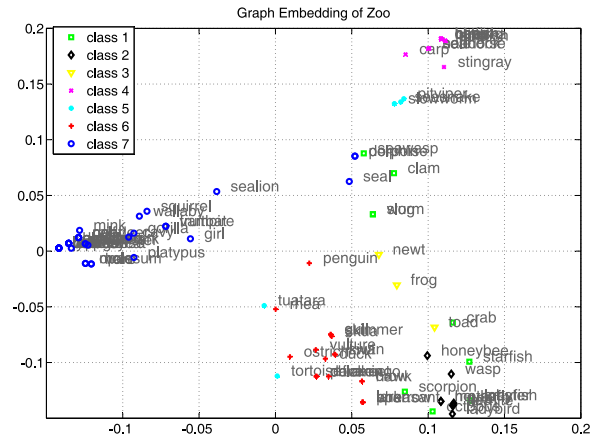
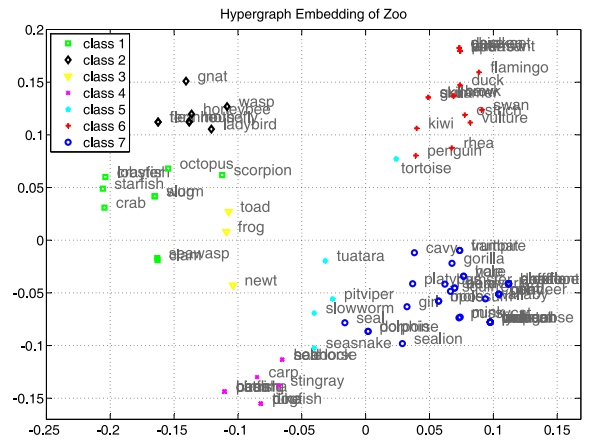


Fig. 5. Classification error rate versus the labeling percentage on the Zoo dataset, based on LRR, normal-graph NSLLRR and NSHLRR.



(a) With normal graph-based NSLLRR.



(b) With hypergraph-based NSHLRR.

Fig. 6. The embeddings of the Zoo dataset with eigenvectors corresponding to the second and third smallest eigenvalues. (a) The cluster accuracy is 74 percent. (b) The cluster accuracy is 91 percent.

Eigenmaps [2], [47], in which the tailing eigenvectors of the Laplacian matrix of  $\mathcal{G}$ , except for the eigenvector corresponding to eigenvalue 0, is often exploited for embedding.

To evaluate the effectiveness of our proposed method on hypergraph embedding, we follow the Laplacian Eigenmaps framework [2]. First we use both hypergraph-based NSHLRR and normal-graph-based NSLLRR to learn the similarity  $Z$  of high dimensional data, respectively. Then we do graph embedding by Laplacian Eigenmaps by using the Laplacian matrix derived from the learned  $Z$ . Thus, the embedding of vertex  $v$  in a  $k$ -dimensional space is just a row vector of the matrix that contains columns of the eigenvectors associated with the  $k$  smallest eigenvalues, except for the eigenvector corresponding to eigenvalue 0, of the Laplacian matrix of  $\mathcal{G}$  [54].

In this experiment, we test both hypergraph-based NSHLRR and normal-graph-based NSLLRR methods on the Zoo dataset. First we visualize the dimensionality reduction results in Fig. 6 by looking at the embeddings along the second and third smallest eigenvectors direction, i.e., in a two-dimensional space. As can be seen from the figure, visualization for both methods demonstrates obvious separation among different classes, however classes 1, 5 and 7 under the normal-graph-based NSLLRR are scattered much more largely than under the hypergraph-based NSHLRR.

In order to quantitatively validate the above visual observation, we compute the cluster accuracy of two methods on the Zoo dataset after graph embedding based on a post K-mean process. The cluster accuracy of the hypergraph-based NSHLRR is 91 percent while the normal-graph-based NSLLRR only achieves 74 percent cluster accuracy.

Since some sample (e.g., **tortoise**) has a special attribution, the embedding with the hypergraph-based NSHLRR fails to give better grouping. Nevertheless, both the visualization in Fig. 6 and the quantity result demonstrate overall efficacy of the hypergraph-based NSHLRR.

## 8 CONCLUSION

In this paper, we propose a generalized Laplacian regularized low-rank representation framework, in which we explicitly consider the manifold structures of data space by introducing a Laplacian regularization term. In particular, a hypergraph Laplacian regularizer is introduced into the conventional LRR objective function to form the hypergraph Laplacian regularized LRR model. Our proposed method not only can represent the global low-dimensional structures, but also capture the intrinsic non-linear geometric information in data. The extensive experimental results on image clustering, classification and embedding show the efficacy of our proposed method. However, a theoretical analysis the benefits of using nonlinearities and hyperedges will be further investigated as the future work and the application of the proposed model to a broader range of problems is to be explored.

## ACKNOWLEDGMENTS

The research project is supported by the Australian Research Council (ARC) through the grant DP130100364. Ming Yin is supported by the Foundation of Key Laboratory of Autonomous Systems and Networked Control, Ministry of Education, P.R. China (No. 2013A06), and Guangdong Natural Science Foundation (No. 2014A030313511) also in part supported by National Natural Science Foundation (NSF) of China (grant nos. 61333013 and 61322306). Zhouchen Lin is supported by National Basic Research Program of China (973 Program) (grant no. 2015CB352502), National Natural Science Foundation (NSF) of China (grant nos. 61272341 and 61231002), and Microsoft Research Asia Collaborative Research Program. Zhouchen Lin is the corresponding author.

## REFERENCES

- [1] D. Babacan, M. Luessi, R. Molina, and A. K. Katsaggelos, "Sparse Bayesian methods for low-rank matrix estimation," *IEEE Trans. Signal Process.*, vol. 60, no. 8, pp. 3964–3977, Aug. 2012.
- [2] M. Belkin and P. Niyogi, "Laplacian eigenmaps for dimensionality reduction and data representation," *Neural Comput.*, vol. 15, pp. 1373–1396, 2003.
- [3] M. Belkin and P. Niyogi, "Laplacian eigenmaps and spectral techniques for embedding and clustering," in *Proc. Adv. Neural Inf. Process. Syst.*, 2001, vol. 14, pp. 585–591.
- [4] G. F. Bull and J. Gao, "Transposed low rank representation for image classification," in *Proc. Int. Conf. Digital Image Comput. Techn. Appl.*, 2012, pp. 1–7.
- [5] D. Cai, X. He, and J. Han, "Spectral regression: A unified approach for sparse subspace learning," in *Proc. IEEE Int. Conf. Data Mining*, 2007, pp. 73–82.
- [6] D. Cai, X. He, J. Han, and T. S. Huang, "Graph regularized non-negative matrix factorization for data representation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 8, pp. 1548–1560, Aug. 2011.
- [7] D. Cai, X. He, X. Wang, H. Bao, and J. Han, "Locality preserving nonnegative matrix factorization," in *Proc. Int. Joint Conf. Artif. Intell.*, 2009, pp. 1010–1015.
- [8] J. F. Cai, E. J. Candès, and Z. Shen, "A singular value thresholding algorithm for matrix completion," *SIAM J. Optimization*, vol. 20, no. 4, pp. 1956–1982, 2008.
- [9] E. Candès and B. Recht, "Exact matrix completion via convex optimization," *Found. Comput. Math.*, vol. 9, pp. 717–772, 2008.
- [10] E. Candès, X. Li, Y. Ma, and J. Wright, "Robust principal component analysis?" *J. ACM*, vol. 58, article 11: 1–37, 2011.
- [11] C.-F. Chen, C.-P. Wei, and Y.-C. F. Wang, "Low-rank matrix recovery with structural incoherence for robust face recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, 2012, pp. 2618–2625.
- [12] S. S. Chen, D. L. Donoho, and M. A. Saunders, "Atomic decomposition by basis pursuit," *SIAM Rev.*, vol. 43, no. 1, pp. 129–159, Jan. 2001.
- [13] B. Cheng, J. Yang, S. Yan, Y. Fu, and T. S. Huang, "Learning with  $\ell_1$ -graph for image analysis," *IEEE Trans. Image Process.*, vol. 19, no. 4, pp. 858–866, Apr. 2010.
- [14] F. R. K. Chung, *Spectral Graph Theory*. Providence, RI, USA: AMS, 1997.
- [15] P. G. Doyle and J. L. Snell, *Random Walks and Electric Networks*. Washington, DC, USA: MAA, 1984.
- [16] D. L. Donoho, "Compressed sensing," *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1289–1306, Apr. 2006.
- [17] J. Gao, "Robust  $\ell_1$  principal component analysis and its Bayesian variational inference," *Neural Comput.*, vol. 20, no. 2, pp. 555–572, 2008.
- [18] S. Gao, I. W.-H. Tsang, and L.-T. Chia, "Laplacian sparse coding, hypergraph Laplacian sparse coding, and applications," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 1, pp. 92–104, Jan. 2013.
- [19] R. Hadsell, S. Chopra, and Y. Lecun, "Dimensionality reduction by learning an invariant mapping," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recog.*, 2006, pp. 1735–1742.
- [20] 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.
- [21] X. He, D. Cai, S. Yan, and H.-J. Zhang, "Neighborhood preserving embedding," in *Proc. IEEE 10th Int. Conf. Comput. Vis.*, 2005, pp. 1208–1213.
- [22] X. He, S. Yan, Y. Hu, P. Niyogi, and H. J. Zhang, "Face recognition using Laplacianfaces," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 3, pp. 328–340, Mar. 2005.
- [23] S. Ji and J. Ye, "An accelerated gradient method for trace norm minimization," in *Proc. 26th Annu. Int. Conf. Mach. Learn.*, 2009, pp. 457–464.
- [24] I. Jolliffe, *Principal Component Analysis*, 2nd ed. New York, NY, USA: Springer, 2002.
- [25] H. Kim and H. Park, "Nonnegative matrix factorization based on alternating nonnegativity constrained least squares and active set method," *SIAM J. Matrix Anal. Appl.*, vol. 30, no. 2, pp. 713–730, 2008.
- [26] F. De la Torre and M. J. Black, "Robust principal component analysis for computer vision," in *Proc. IEEE 8th Int. Conf. Comput. Vis.*, 2001, vol. 1, pp. 362–369.
- [27] Z. Lin, M. Chen, and Y. Ma, "The augmented Lagrange multiplier method for exact recovery of corrupted low-rank matrices," UIUC, Champaign, IL, USA, Tech. Rep. UIUC-ENG-09-2215, 2009.
- [28] Z. Lin, A. Ganesh, J. Wright, L. Wu, M. Chen, and Y. Ma, "Fast convex optimization algorithms for exact recovery of a corrupted low-rank matrix," UIUC, Champaign, IL, USA, Tech. Rep. UIUC-ENG-09-2214, Jul. 2009.
- [29] Z. Lin, R. Liu, and Z. Su, "Linearized alternating direction method with adaptive penalty for low-rank representation," in *Proc. Adv. Neural Inf. Process. Syst.*, 2011, pp. 612–620.
- [30] G. Liu, Z. Lin, S. Yan, J. Sun, 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.
- [31] G. Liu, Z. Lin, and Y. Yu, "Robust subspace segmentation by low-rank representation," in *Proc. Int. Conf. Mach. Learn.*, 2010, pp. 663–670.
- [32] G. Liu, H. Xu, and S. Yan, "Exact subspace segmentation and outlier detection by low-rank representation," in *Proc. 15th Int. Conf. Artif. Intell. Stat.*, 2012, pp. 703–711.

- [33] G. Liu and S. Yan, "Latent low-rank representation for subspace segmentation and feature extraction," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2011, pp. 1615–1622.
- [34] R. Liu, Z. Lin, and Z. Su, "Linearized alternating direction method with parallel splitting and adaptive penalty for separable convex programs in machine learning," *J. Mach. Learning Res.*, vol. 29, pp. 1–16, 2013.
- [35] X. Lu, Y. Wang, and Y. Yuan, "Graph-regularized low-rank representation for destriping of hyperspectral images," *IEEE Trans. Geosci. Remote Sensing*, vol. 51, no. 7, pp. 4009–4018, Jul. 2013.
- [36] L. Pu and B. Faltings, "Hypergraph learning with hyperedge expansion," in *Proc. Eur. Conf. Mach. Learning Knowl. Discovery Databases*, 2012, pp. 410–425.
- [37] S. T. Roweis and L. K. Saul, "Nonlinear dimensionality reduction by locally linear embedding," *Science*, vol. 290, pp. 2323–2326, 2000.
- [38] 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.
- [39] M. Tao and X. Yuan, "Recovering low-rank and sparse components of matrices from incomplete and noisy observations," *SIAM J. Optimization*, vol. 21, no. 2, pp. 57–81, 2011.
- [40] J. B. Tenenbaum, V. d. Silva, and J. C. Langford, "A global geometric framework for nonlinear dimensionality reduction," *Science*, vol. 290, pp. 2319–2323, 2000.
- [41] K. Toh and S. Yun, "An accelerated proximal gradient algorithms for nuclear norm regularized least squares problems," *Pacific J. Optimization*, vol. 6, pp. 615–640, 2010.
- [42] J. Wang, F. Wang, C. Zhang, H. C. Shen, and L. Quan, "Linear neighborhood propagation and its applications," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 31, no. 9, pp. 1600–1615, Sep. 2009.
- [43] Z. Wen, D. Goldfarb, and W. Yin, "Alternating direction augmented Lagrangian methods for semidefinite programming," *Math. Program. Comput.*, vol. 2, pp. 203–230, 2010.
- [44] J. Wright, A. Ganesh, S. Rao, and Y. Ma, "Robust principal component analysis: Exact recovery of corrupted low-rank matrices by convex optimization," in *Proc. Adv. Neural Inf. Process. Syst.*, 2009, pp. 2080–2088.
- [45] J. Wright, Y. Ma, J. Mairal, G. Sapiro, T. Huang, and S. Yan, "Sparse representation for computer vision and pattern recognition," *Proc. IEEE*, vol. 98, no. 6, pp. 1031–1044, Jun. 2010.
- [46] L. Wu, A. Ganesh, B. Shi, Y. Matsushita, Y. Wang, and Y. Ma, "Convex optimization based low-rank matrix completion and recovery for photometric stereo and factor classification," submitted to *IEEE Trans. Pattern Anal. Mach. Intell.*, Aug. 2012.
- [47] S. Yan, D. Xu, B. Zhang, H.-J. Zhang, Q. Yang, and S. Lin, "Graph embedding and extensions: A general framework for dimensionality reduction," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 29, no. 1, pp. 40–51, Jan. 2007.
- [48] S. Yang, X. Wang, M. Wang, Y. Han, and L. Jiao, "Semi-supervised low-rank representation graph for pattern recognition," *IET Image Process.*, vol. 7, no. 2, pp. 131–136, Mar. 2013.
- [49] M. Yin, S. Cai, and J. Gao, "Robust face recognition via double low-rank matrix recovery for feature extraction," in *Proc. IEEE Int. Conf. Image Process.*, 2013, pp. 3770–3774.
- [50] Z. Zhang, X. Liang, A. Ganesh, and Y. Ma, "Tilt: Transform invariant low-rank textures," in *Proc. 10th Asian Conf. Comput. Vis.*, 2010, pp. 314–328.
- [51] 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.
- [52] M. Zheng, J. Bu, C. Chen, C. Wang, L. Zhang, G. Qiu, and D. Cai, "Graph regularized sparse coding for image representation," *IEEE Trans. Image Process.*, vol. 20, no. 5, pp. 1327–1336, May 2011.
- [53] Y. Zheng, X. Zhang, S. Yang, and L. Jiao, "Low-rank representation with local constraint for graph construction," *Neurocomputing*, vol. 122, pp. 398–405, 2013.
- [54] D. Zhou, J. Huang, and B. Scholkopf, "Learning with hypergraphs: Clustering, classification, and embedding," in *Proc. Adv. Neural Inf. Process. Syst.*, 2006, pp. 1601–1608.
- [55] X. Zhu, Z. Ghahramani, and J. Lafferty, "Semi-supervised learning using Gaussian fields and harmonic functions," in *Proc. Proc. Int. Conf. Mach. Learn.*, 2003, pp. 912–919.
- [56] L. Zhuang, H. Gao, Z. Lin, Y. Ma, X. Zhang, and N. Yu, "Non-negative low rank and sparse graph for semi-supervised learning," in *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, 2012, pp. 2328–2335.



**Ming Yin** received the PhD degree in information and communication engineering from Huazhong University of Science and Technology (HUST), Wuhan, China, in 2006. He was a visiting scholar at the School of Computing and mathematics, Charles Sturt University, Bathurst, Australia, from January 2012 to December 2012. He is currently an assistant professor with the School of Automation, Guangdong University of Technology, Guangzhou, China. His research interests include image/video coding, image deblurring, sparse representation, and unsupervised/semi-supervised data cluster/classification.



**Junbin Gao** received the BSc degree in computational mathematics from Huazhong University of Science and Technology (HUST), China, in 1982 and the PhD degree from Dalian University of Technology, China in 1991. He is a professor in computing science in the School of Computing and Mathematics, Charles Sturt University, Australia. He was a senior lecturer, a lecturer in computer science from 2001 to 2005 at the University of New England, Australia. From 1982 to 2001, he was an associate lecturer, lecturer, associate professor, and a professor in the Department of Mathematics at HUST. His main research interests include machine learning, data mining, Bayesian learning and inference, and image analysis.



**Zhouchen Lin** (M'00-SM'08) received the PhD degree in applied mathematics from Peking University in 2000. He is currently a professor with the Key Laboratory of Machine Perception, School of Electronics Engineering and Computer Science, Peking University. He is also a chair professor with Northeast Normal University. He was a guest professor with Shanghai Jiaotong University, Beijing Jiaotong University, and Southeast University. He was also a guest researcher with the Institute of Computing Technology, Chinese Academy of Sciences. His research interests include computer vision, image processing, machine learning, pattern recognition, and numerical optimization. He is an area chair of CVPR 2014, ICCV 2015, NIPS 2015, and AAAI 2016. He is an associate editor of the *IEEE Transactions on Pattern Analysis and Machine Intelligence* and the *International Journal of Computer Vision*. He is a senior member of the IEEE.

► For more information on this or any other computing topic, please visit our Digital Library at [www.computer.org/publications/dlib](http://www.computer.org/publications/dlib).