Laplacian Regularized Low-Rank Representation and Its Applications

Ming Yin, Junbin Gao and Zhouchen Lin

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 face recognition, image clustering, motion segmentation and feature extraction. In the real world, it is more often for data to reside on low-dimensional submanifolds embedded in a high-dimensional ambient space. However, the LRR method does not take into account the intrinsic geometrical structures within the data, thus the locality and similarity information within data may be missing in the learning process. To improve LRR in this regard, we propose a general Laplacian regularized low-rank representation model, i.e., a Non-negative Sparse Hyper-Laplacian regularized LRR model (NSHLRR). In our model, the high order Laplacian regularizer can be easily imposed on the traditional LRR to efficiently represent the pairwise, even high order Laplacian, Manifold Structure, Laplacian Matrix, Regularization.

Index Terms— Low-Rank Representation, Graph, Hyper-Laplacian, Manifold Structure, Laplacian Matrix, Regularization.

I. INTRODUCTION

Low-Rank Representation (LRR) [28], [30], 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], [47], image de-striping [32], image clustering [49], [54] and video background/foreground separation [11]. 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 the observed data all lie near some low-dimensional subspace, the data matrix stacked from all the vectorized observations should be approximately of low rank [42]. The conventional Principal Component Analysis (PCA) [22] is one of ways to recover the best low-rank representation in terms of $\ell_2$ errors. It is well known that PCA, applied in computer vision, is fragile to outliers or sparse/extreme errors introduced by occlusion and disguise [24]. 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 [16] and robust PCA [10]. These robust PCA schemes show that $\ell_1$-type noise models can offer better robustness than the conventional $\ell_2$-type. 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 turn to be non-convex and optimization algorithms to the problems get stuck at local minima resulting in performance depression.

The LRR method [11] [28] [29] [30] focuses on low-rank data representation based on the hypothesis that the data is approximately jointly spanned by several low-dimensional subspaces and it also takes care of largely contaminated outliers by incorporating $\ell_1$ noise models, thus LRR can accurately recover the row space of the original data and detect outliers under mild conditions. In general, the resulting problem becomes a convex optimization problem aiming at the optimal solution of minimizing a combination of the nuclear norm and the $\ell_1$-norm in polynomial time [10]. In order to handle the cases where the number of observation data is insufficient or data themselves too badly corrupted, Liu and Yan [30] further proposed a latent low-rank representation approach. In the latent LRR, the 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 is the row space information recovered and column space information of input data matrix is not sufficiently exploited for learning subspace structures. To address this issue, Yin et al. [47] proposed a novel approach, termed double LRR, which simultaneously learns the row space and column space information embedded in the 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 the face images are sampled from a non-linear low-dimensional manifold which is embedded in a high-dimensional ambient space [21]. Therefore, in
this case, the LRR method may fail to discover the intrinsic geometrical and discriminating structures of the data space [6] [32] [50], which is essential to the actual applications.

To preserve the local geometrical structures embedded in a high-dimensional space, numerous researchers have considered the manifold learning methods, such as the locally linear embedding (LLE) [35], ISOMAP [38], Locality Preserving Projection (LPP) [21], Neighborhood Preserving Embedding (NPE) [20] and Laplacian Eigenmap (LE) [2]. All these algorithms are motivated by the idea of the so-called local invariance [13], which aims to estimate geometrical and topological properties of the submanifold from random points (scattered data) lying on this unknown submanifold.

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 under a new basis are close to each other too [19]. In recent years, this observation inspires Zheng et al. [50] to propose a graph regularized sparse coding to learn the sparse representations that explicitly take into account the local manifold structures of the data. Similarly, Gao et al. [17] also proposed two Laplacian regularized sparse coding, 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 [23] of data representation in order to address the failure in representing geometric structures in data. To exploit the intrinsic geometry of the probability distribution, He et al. [19] 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 [49].

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. [32] proposed a novel graph-regularized LRR destriping approach by incorporating the LRR technique. Their formulation shares the similar idea as our proposed model, but with the purpose of removing the striping noise of hyperspectral images, however, more specifically, we consider the sparsity and non-negative constraints in our model in favoring data clustering and classification. Furthermore, we provide a convergence analysis of the algorithm in this paper.

Motivated by the above works, in this paper, we propose a non-negative sparse hyper-Laplacian regularized low-rank representation model, termed NSHLLRR, for image representation. Note that here we use a hypergraph, instead of normal graph, to describe similarity structures among the data instances in order to introduce a general manifold regularization to the LRR model. Generally speaking, although the data points reside on a non-linear submanifold, it is reasonable to assume that the local neighbors are linearly related. Thus, we can characterize the local geometry of the data by the linear coefficients that reconstruct each data point from its neighbors. Furthermore, the obtained optimal representation by our proposed model is close to block-diagonal. Only in this case can the non-linear manifolds be clustered successfully.

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

1) We extend the LRR model by introducing a regularized term based on manifold structures of data, and propose a novel general Laplacian regularized LRR. More specifically, as the data manifold is usually unknown, the local geometrical structures for data are often modeled by a nearest neighbour graph and then the graph structures are incorporated into an optimization problem aiming for the lowest-rank matrix representation.

2) Instead of considering pairwise graphs, 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 is effectively comparable or even better than many the state-of-the-art methods in the semi-supervised image classification and clustering tasks.

3) We constrain the representation coefficients to be positive to facilitate learning local manifold structures. Recently, many literatures [23] [54] have shown the non-negative constraint leads to a parts-based representation for data which can render the model more physical interpretation. Since we use locally linear manifold to approximate the nonlinear 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. This is distinct from the model proposed in [32].

The remainder of this paper is organized as follows. In Section II we give a brief review on related works. Section III is dedicated to introducing a novel Low-Rank Representation model with Laplacian regularization, i.e., NSHLLRR. In Section IV we present an algorithm for solving the optimization problem via the linearized alternating direction method with adaptive penalty (LADMAP) [27]. Section V presents the experimental results for image clustering and semi-supervised image classification tasks. The results for applying the proposed model on hypergraph-based applications are provided in Section VI and Section VII. Finally, Section VIII concludes our work.

II. RELATED WORKS

In recent years, the success of low-rank matrix and graph representations has prompted researchers to extend these representations by jointly considering two approaches. Before we introduce our proposed model, in this section, we review the recent novel methods such as LRR [23] [29] [50] and graph based analysis [2] [45] [50].

A. Low-Rank Matrix Recovery Formulation

The LRR model [23] is focused on the assumption that data may be approximately sampled from a union of several low-dimensional subspaces. Given a set of data samples, each of which can be represented as a linear combination of the bases in a dictionary. LRR aims at finding the lowest rank
representation of all data jointly. It has been demonstrated that LRR is quite effective in exploring low-dimensional subspace structures embedded in data.

Consider the case where data $Y$ are drawn from a union of multiple subspaces given by $\bigcup_{i=1}^{k} S_i$ where $S_1$, $S_2$, ..., $S_k$ are low-dimensional subspaces. The LRR model for the given data $Y$ is defined as the following rank minimization problem

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

where the columns of $A$ are a set of known bases or dictionary items and $E$ denotes the error components and $\lambda$ denotes 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 intractability of $\ell_0$-minimization, a convex relaxation version of the optimization problem is proposed

$$\min_{Z,E} \|Z\|_* + \lambda \|E\|_1 \quad \text{s.t. } Y = AZ + E$$  \hspace{1cm} (2)

where $\|Z\|_*$ is the so-called nuclear norm, defined as the sum of all singular values of $Z$, which is the convex envelope of the rank function, and $\|E\|_1$ is the $\ell_1$ norm, the convex envelope of the $\ell_0$ norm. In fact, it has been proved in [28] 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 the data $Y$ is also in low rank. Denote $X = AZ$. Instead of solving the optimization problem with respect to the representation coefficient matrix $Z$, we may solve the clean data $X$ directly. This links LRR with the state-of-the-art model Robust PCA (RPCA) [10] for which the optimization problem becomes

$$\min_{X,E} \|X\|_* + \lambda \|E\|_1 \quad \text{s.t. } Y = X + E$$  \hspace{1cm} (3)

The above optimization problem (3) is broadly used in computer vision community [41] [43]. RPCA formulation only aims to recover the low rank clean data from the given noised data. Compared to model (5), the LRR model is able to reveal the implicit data membership. To see this, let $X = [x_1, x_2, ..., x_n]$ be the matrix whose columns are clean data samples drawn from independent subspaces $\{S_i\}$. Given a basis $A = [a_1, a_2, ..., a_m]$ linearly spanning the whole data space, also called a dictionary, we can represent the data as $X = AZ$, where $Z = [z_1, z_2, ..., z_n]$ is the coefficient matrix with each $z_i$ being the representation of $x_i$. Under an appropriately designed dictionary $A$, the optimal solution of LRR $Z^*$ can actually reveal some underlying affinity between data points so as to uncover the ground truth structures of data. In [28], 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 $y_i$ and $y_j$.

A number of methods have been proposed for solving low-rank matrix problems, such as singular value thresholding [8], accelerated proximal gradient (APG) [20], and augmented Lagrange Multiplier Method (ALM) [25]. Noted that Lin et al. [27] proposed a fast method for minimizing nuclear norm problem, termed linearized alternating direction method with adaptive penalty (LADMAP), which uses less auxiliary variables and no matrix inversions and hence converges faster than usual alternating direction method (ADM) [41]. Most recently, in order to handle the multi-variable cases, Lin et al. [31] proposed a new iterative procedure, called LADMAPSAP, and proved its convergence.

### B. Graph Based Manifold Analysis

Without loss of generality, suppose $Y = [y_1, y_2, ..., y_n]$ are sampled from an underlying submanifold $\mathcal{M}$. It is reasonable to assume that the neighborhood relationship is linear [3]. Thus, the local geometry of these data points can be efficiently 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 graph $G$ with $n$ vertices, where each data point denotes a vertex. Meanwhile, we define a symmetric weight matrix $W \in \mathbb{R}^{n \times n}$, in which $W_{ij}$ has 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 $V = \{v_1, v_2, ..., v_n\}$ is the vertex set with each node $v_i$ corresponding to a data point $y_i$, 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.

$$W_{ij} = \begin{cases} 1 & \text{if } y_i \in N_k(y_j) \text{ or } y_j \in N_k(y_i) \\ 0 & \text{otherwise} \end{cases}$$  \hspace{1cm} (4)

where $N_k(y_i)$ denotes the set of $k$ nearest neighbors of $y_i$.

The graph embedding is based on the natural assumption that if two data points $y_i$ and $y_j$ 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 the graph embedding aims at describing each vertex of the graph by a low-dimensional vector that preserves affinity between the vertex pairs, where similarity is measured by the edge weight. In mathematics, this relationship defined by manifold assumption is formulated by,

$$\min \sum_{ij} \|z_i - z_j\|^2 W_{ij}$$  \hspace{1cm} (5)

where $z_i$ and $z_j$ are the mappings of $y_i$ and $y_j$ under some transformation, respectively. This rule plays a critical role in developing various kinds of algorithms including dimensionality reduction algorithms [45], clustering algorithms [5] and semi-supervised learning algorithms [24]. The degree matrix is defined as $D$, which is a diagonal matrix, and its $ij$th entry $D_{ij}$ corresponds to the summation of all the similarities related to $y_i$, i.e., $D_{ij} = \sum_j W_{ij}$. Then, the graph Laplacian matrix $L$ is defined by,

$$L = D - W.$$  

It is easy to prove that the graph embedding (5) can be written into

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

In many practical 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 another normal graph that is easier to handle \[17, 34, 52\]. If this complex relationship is simplified into pairwise one, it is inevitable to lose some useful information for targeted learning tasks. To efficiently represent the complex relationship among the given data, the concept of hypergraph is employed in manifold learning tasks \[52\], such as clustering, classification and embedding. In fact, a hypergraph is an extension of a normal graph in which an edge can link up more than two vertices.

For the convenience of subsequent presentation, we here introduce some basic notations on hypergraphs. Given a hypergraph \(G = (V, E)\), \(V\) denotes the vertex set and \(E\) represents the hyperedge set in which each \(e\) is a subset of vertices. The weight corresponding to a hyperedge \(e\) is denoted by \(W(e)\), usually a positive number. The degree of a vertex \(v\) \(d_v \in D\) and the diagonal degree matrix \(D\) as incidence matrix \(H\) as

\[
H = \sum_{e \in E} W(e)h(v, e),
\]

where \(h(v, e) = 1\) if \(v \in e\) and \(0\) otherwise. Denote the degree of a hyperedge \(e\) by \(d(e) = \sum_{v \in V} h(v, e)\) and the diagonal degree matrix \(D_E\) consists of diagonal elements \(d(e)\). Similarly denote \(D_V\) the diagonal matrix whose elements correspond to the degree of each vertex. Then the unnormalized hyper-Laplacian matrix \[52\] can be defined as,

\[
L^h = D_V - HW_ED_E^{-1}H^T
\]

where \(W_E\) is the diagonal matrix of the edge weight \(W(e)\).

\section{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 the 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 geometrical structures between the data points. Different from the graph regularized sparse representation \[17, 50\], our proposed model can effectively represent itself without relying on a dictionary in a large size.

\subsection{Non-negative sparse hyper-Laplacian regularized LRR}

If we use the data \(Y\) itself as the basis \(A\) in \[2\], there are two explanations to the coefficient matrix \(Z\). Firstly, the element \(z_{ij}\) of \(Z\) reflects the “similarity” between the data pair \(y_i\) and \(y_j\), hence \(Z\) is sometimes called affinity matrix. Secondly, we can explain each column \(z_i\) of \(Z\) as a new representation of the data \(y_i\) in terms of other samples in \(Y\), in other words, \(z_i\) can be used as a representative for \(y_i\).

To introduce richer information over \(Z\), people consider imposing some helpful regularization on the coefficient matrix, such as sparsity \[15, 43\] and a non-negative constraint \[23\]. 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.

\[
\begin{align*}
\min_{Z, E} & \|Z\|_\ast + \lambda \|Z\|_1 + \gamma \|E\|_1, \\
\text{s.t.} & \quad Y = YZ + E, \quad Z \geq 0
\end{align*}
\]

In fact, in order to deal with grossly corrupted data, Zhuang \textit{et al.} \[54\] proposed a novel 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 the 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 basis, the distance between \(z_i\) and \(z_j\) is actually one of dissimilarity measures for the original data points \(y_i\) and \(y_j\). Thus we can add a similarity matching regularization term \(\sum_{i,j} ||z_i - z_j||^2W_{ij}\) into the objective function \[7\].

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

According to the hypergraph definition, we enforce the LRR coefficients corresponding to the 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 \(e\) by \(W(e)/d(e)\).

Then, the Non-negative Sparse Hyper-Laplacian regularized LRR model, termed NSLRR, is formulated as,

\[
\begin{align*}
\min_{Z, E} & \|Z\|_\ast + \lambda \|Z\|_1 + \beta \sum_{i,j} \langle e \in E \rangle \|z_i - z_j\|^2 \frac{W(e)}{d(e)} + \gamma \|E\|_1, \\
\text{s.t.} & \quad Y = YZ + E, \quad Z \geq 0
\end{align*}
\]

The non-negative constraint on \(Z\) aims to guarantee that the coefficients are meaningful and better embody the dependency among the data points.

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., \(d(e) = 2\). In fact, a hypergraph is a general 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 have only two elements, \[8\] can be rewritten as a Non-negative Sparse Laplacian regularized LRR model, termed NSLLRR, as defined below,

\[
\begin{align*}
\min_{Z, E} & \|Z\|_\ast + \lambda \|Z\|_1 + \beta \sum_{ij} \|z_i - z_j\|^2 W_{ij} + \gamma \|E\|_1, \\
\text{s.t.} & \quad Y = YZ + E, \quad Z \geq 0
\end{align*}
\]
Therefore, we are actually considering a general version of graph regularized LRR model. The advantage of this generalized model is explained as follows. By introducing the hypergraph regularization term, we can effectively preserve the affinity among these samples belonging to the same hyperedge so as to keeping the manifold structures of data. In particular, it is more powerful to explore the multiple-wise relations among more data points.

Finally we can see that the matrix form of hyper-Laplacian regularized LRR model is achieved by some algebraic manipulations.

\[
\min_{Z, E} \|Z\|_* + \lambda \|Z\|_1 + \beta tr(ZL^hZ^T) + \gamma \|E\|_1,
\text{s.t. } Y = YZ + E, \ Z \geq 0
\]  

(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 as

\[
\min_{Z, E} \|Z\|_* + \lambda \|Z\|_1 + \beta tr(ZLZ^T) + \gamma \|E\|_1,
\text{s.t. } Y = YZ + E, \ Z \geq 0
\]  

(11)

where \(L\) is the Laplacian matrix for the graph built on \(W\) and other parameters are similar to the case in 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 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 any vertex pair and each hyperedge just includes two vertices. On the other hand, the NSLLRR [11] model can also be changed into a NSHLRR model [10] model under some special definition of the weight \(W\) [17]. To this end, we can unify a manifold constraint into LRR framework for better exploiting the geometrical structures of the data space.

Finally, we would like to further talk about the model of the graph-regularized LRR (GLRR) proposed in [32]. Actually, that GLRR model is a special case of our NSHLRR model by just considering the pairwise relation between data points. The similar idea can be seen in another recent work [51], however the authors used the second explanation of \(Z\) that \(z_{ij}\) is the similarity between the data \(y_i\) and \(y_j\), hence the relevant regularized term is defined as \(\sum |z_{ij}|\|y_i - y_j\|^2\), for which it is easy to extend to the case for multiple relations.

### B. Laplacian Regularized LRR for Unsupervised Learning

According to the principle of LRR, \(Z\) is actually a new representation of each data point under the dataset itself. Specifically, \(Z\) reflects the affinity between the data pair \(y_i\) and \(y_j\), which can be used for data clustering [30]. In recent years, many similarity-based clustering algorithms have emerged, such as K-means and spectral clustering method [36], which do not require the assumption on probability distribution of the data. In this regard, our proposed method should greatly benefit image clustering as it is powerful in learning similarity facilitated by incorporating data neighborhood information.

### C. Laplacian Regularized LRR for Semi-supervised Learning

In addition, NSLLRR not only can perform well in unsupervised learning, but also can be adopted to semi-supervised learning efficiently. In pattern recognition community, semi-supervised learning has been attracting considerable attention over the past decades [46] [54] because the labeled samples are hard or expensive to acquire and whereas unlabeled ones are easy and inexpensive. Recently, the graph-based learning methods have been widely used to develop the high performance algorithms for semi-supervised learning task [13] [38] [40] [54].

In general, most of graph-based semi-supervised learning algorithms use a common assumption of a well-posedness of the constructed graph. However, this supposition is not always true since the graph parameters require many manual settings for constructing graphs. On the other hand, the graph construction process plays a key role in the performance of algorithms. Given that our proposed model well preserves the locality and the similarity of data, we empirically expect our proposed method also has the potential to convey more discriminative information compared with traditional graph-based methods.

In this section, we present a novel semi-supervised learning framework incorporating the above proposed Laplacian regularized LRR model. It is well known that the objective of graph based semi-supervised learning can be formulated as follows [45].

\[
\min \sum_{ij} W_{ij} \|f_i - f_j\|^2
\]

where \(f_i\) and \(f_j\) denote the probabilities of \(y_i\) and \(y_j\) belonging to different classes, respectively.

Under semi-supervised learning framework, we denote a label matrix \(A = [A_l; A_u] \in \mathbb{R}^{V \times c}\) where \(A_l\) represents the label information for the labeled data and while \(A_u\) is the information for the unlabeled data, and \(c\) denotes the number of the classes. If a sample \(y_i\) is associated with label \(k \in \{1, ..., c\}\), then \(A_{ik} = 1\), and otherwise, \(A_{ik} = 0\). 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, in mathematics, the objective of graph based semi-supervised learning is readily transferred into,

\[
\min \sum_{ij} W_{ij} \|f_i - f_j\|^2 = \text{tr} \left( F^T L_w F \right)
\]  

(12)

where \(L_w\) is the graph Laplacian matrix inferred by a weighted graph \(G\). The graph is constructed via using relationship of the labeled and unlabeled data points where edge weights encode the similarity between samples. To achieve this target, our work focuses on how to efficiently learn the graph underlying the LRR structures of data where labeled and unlabeled data are represented as vertices in a graph. Specifically in this
paper \( L_w \) is computed by the symmetrized affinity matrix 
\[
\frac{1}{2} (|Z| + |Z^T|)
\]
where \( Z \) is learned by the Laplacian regularized LRR method.

In order to efficiently predict the labels of the unlabeled samples, we here adopt Gaussian fields rather than random fields over the label set as a basis for semi-supervised classification. That is, we form a Gaussian field to assign a probability distribution of \( f_l \). Then, the minimum energy function \((12)\) is harmonic; namely, it should simultaneously satisfy \( L_w F_w = 0 \) on unlabeled data samples and \( F_l = A_l \) on the labeled ones. Please refer to \([53]\) for more details.

Based on the above analysis, we finally learn a Gaussian Harmonic Function (GHF) \([53]\), which combines Gaussian fields and harmonic function, to realize the label propagation, illustrated as following.

\[
\min_{F \in \mathbb{R}^{1 \times c}} \text{tr} (F^T L_w F), \quad \text{s.t. } L_w F_w = 0, \quad F_l = A_l
\]

In particular, the learning objective is to recover the continuous classification function with the graph and the given labels by optimizing different predefined energy functions. Then, the propagated label information is measured with the error rate for further evaluating the performance of graph based semi-supervised learning methods.

IV. LADMAP FOR SOLVING LAPLACIAN REGULARIZED LRR

In recent years, a lot of algorithms have been proposed for solving the optimization problem of recovering a low-rank matrix from the data \([26, 27, 37, 59]\) and \([42]\). In particular, the alternating direction method (ADM) has drawn considerable attention \([41]\). This is achieved by minimizing the objective function with respect to the variables in a Gauss-Seidel manner. In this paper, we also apply ADM to solve our proposed optimization problems. Unfortunately, directly applying ADM to solve \((10)\) results in an intermediate problem as follows

\[
\min_Z \| Z \|_* + \frac{\lambda}{2} \| C(Z) - D \|_F^2 \tag{13}
\]

where \( C \) is a linear mapping over the matrix \( Z \). When \( C \) is the identity mapping, we can exactly solve \((13)\) by a closed form solution.

It is worthy pointing out that Lu et al. \([32]\) proposed a similar formulation in their work and they provided an algorithm for the corresponding optimization problem by the augmented Lagrange multiplier method (ALM), however, the issue of handling linear mapping over \( Z \) was missing, no mention on how to tackle the non-identity quadratic term for variable “\( F \)” in their paper \([32]\). Here, we would like to present a detailed procedure on how to solve this optimization problem.

To propose an efficient algorithm solving \((10)\), we take the strategy of linearizing the objective function as done in \([27]\) where the approach is named as the linearized version of ADM with Adaptive Penalty (LADMAP).

We introduce an auxiliary variable \( J \) in order to make our objective function \((10)\) separable. Thus the optimization problem can be rewritten as follows,

\[
\min_{Z, J, E} \| Z \|_* + \lambda \| J \|_1 + \beta \text{tr}(Z^T L Z^T) + \gamma \| E \|_1, \quad \text{s.t. } Y = Y Z + E, \quad Z = J, \quad J \geq 0
\] \tag{14}

To remove two linear contraints in \((14)\), we can introducing two Lagrange multipliers \( Y_1 \) and \( Y_2 \), hence the optimization problem can be transferred to minimize the following unconstrained problem,

\[
\min_{J, E, Z} \| Z \|_* + \lambda \| J \|_1 + \beta \text{tr}(Z^T L Z^T) + \gamma \| E \|_1
\]

\[
+ \langle Y_1, Y - Y Z - E \rangle + \langle Y_2, Z - J \rangle
\]

\[
+ \frac{\mu}{2} (\| Y - Y Z - E \|_F^2 + \| Z - J \|_F^2)
\] \tag{15}

where \( \mu \) is actually a penalty parameter, which is adjusted by using adaptive updating strategy as suggested in \([27]\), and \( \| \cdot \|_F \) is the matrix Frobenius norm defined as \( \| Y \|_F = \sum_{i=1}^m \sum_{j=1}^n |y_{ij}|^2 \). This problem can be readily optimized by alternately updating one variable while others fixed. Then, the multipliers are subsequently updated and the whole optimizing procedure is done in an iterative way till the convergence conditions are met.

A. Computation of \( Z \)

Solving \((15)\) w.r.t. \( Z \) is equivalent to optimizing the following objective,

\[
\mathcal{L}_1 = \| Z \|_* + \beta \text{tr}(Z^T L Z^T) + \frac{\mu}{2} \| Y - Y Z - E + \frac{1}{\mu} Y_1 \|_F^2
\]

\[
+ \frac{\mu}{2} \| Y - Y Z - E + \frac{1}{\mu} Y_2 \|_F^2
\] \tag{16}

Denote the quadratic term by

\[
q(Z, E, J, \mu, Y_1, Y_2) = \beta \text{tr}(Z^T L Z^T) + \frac{\mu}{2} \| Y - Y Z - E + \frac{1}{\mu} Y_1 \|_F^2
\]

Then, the subproblem \((16)\) can be approximately replaced at the current approximated solution \( Z_k \) by,

\[
\min_{Z} \| Z \|_* + \langle \nabla_{Z} q(Z_k), Z - Z_k \rangle + \frac{\eta_k \mu_k}{2} \| Z - Z_k \|^2 \tag{17}
\]

\((17)\) has a closed-form solution given by,

\[
Z_{k+1} = \Theta(\eta_k \mu_k) - 1 (Z_k - \nabla_{Z} q(Z_k) / \eta_k)
\] \tag{18}

where \( \Theta(\cdot) \) denotes the singular value thresholding operator (SVT) \([8]\).

B. Computation of \( J \)

Similarly, solving \((15)\) w.r.t. \( J \) is equivalent to optimizing the following objective, while other variables are fixed to their current values,

\[
\mathcal{L}_2 = \min_{J \geq 0} \lambda \| J \|_1 + \frac{\mu}{2} \| J - (Z + \frac{1}{\mu_k} Y_2) \|_F^2
\] \tag{19}
The subproblem (19) has the following solution,

\[ J_{k+1} = \max \left\{ S_{\frac{\lambda}{\mu_k}} \left( Z_{k+1} + \frac{1}{\mu_k} Y_k^k \right), 0 \right\} \]  \hspace{1cm} (20)

where \( S_{\frac{\lambda}{\mu_k}} (\cdot) \) represents the shrinkage operator \(^{23}\) defined by,

\[ S_{\frac{\lambda}{\mu_k}} (\cdot) = U \Sigma_{\frac{\lambda}{\mu_k}} V^T, \quad \Sigma_{\frac{\lambda}{\mu_k}} = \text{diag}(\max\{\sigma_i - \frac{\lambda}{\mu_k}, 0\}). \]

C. Computation of \( E \)

The subproblem for updating \( E \) can be recast as,

\[ L_\lambda = \min_{E} \gamma \| E \|_1 + \frac{\mu}{2} \left\| E - (Y - YZ + \frac{1}{\mu} Y_1) \right\|_F^2 \]  \hspace{1cm} (21)

The subproblem (21) has the solution defined by,

\[ E_{k+1} = S_{\frac{\lambda}{\mu_k}} \left( Y - YZ_{k+1} + \frac{1}{\mu_k} Y_k^k \right) \]  \hspace{1cm} (22)

The detailed algorithm procedure of solving the proposed Laplacian regularized LRR problem is described in Algorithm 1.

Note that the Laplacian regularized LRR problem can also be efficiently solved by Algorithm 1 only replacing the \( L^h \) with \( L \) instead.

D. Convergence and Complexity Analysis

As for the algorithm convergence, it is easy to be derived from \(^{27}\). Although Algorithm 1 is described in three major alternative steps for solving \( Z, J, \) and \( E, \) actually we can easily combine steps for \( J \) and \( E \) into one larger block step by simultaneously solving for \((J, E)\). Thus the convergence conclusion of two variable LADMAP in \(^{27}\) can be applied to our case. Finally the algorithm convergence is guaranteed.

The computational cost of our proposed algorithm is also defined in LADMAP algorithm \(^{27}\). Supposed we have \( n \) samples and each sample has \( d \) dimensions for data \( Y \) and let \( k \) denote the number of total iterations. And \( r \) is the lowest rank which we can find with our algorithm. In step 1 of Algorithm 1, each component of \( \nabla Z_k q \) can be computed in \( O(n^2) \) by using the skinny SVD decomposition \( Z_k \) from the last step, for example we can write \( Z_k L = U_k (\Sigma_k (V_k^T L)) \) and \( Z_k = ((Y U_k) \Sigma_k) V_k^T. \) Further it is easy to use Lanczos procedure to conduct the SVD for \( N_k = Z_k - \frac{1}{\mu_k} \nabla Z_k q, \) which only requires to compute matrix-vector multiplications \( N_k v \) and \( u^T N_k. \) Meanwhile, in steps 2 and 3, we employ soft thresholding to update the sparse matrix whose complexity is \( O(n^2) + O(d n) \). Then, the computational complexity of our algorithm is \( O(k n^2) \) if we assume \( d \leq n. \)

V. GRAPH REGULARIZED LRR FOR IMAGE CLUSTERING AND IMAGE CLASSIFICATION

In this section, in order to investigate the performance of our proposed Laplacian regularized LRR approach, we conducted comprehensive experiments on both unsupervised learning (image clustering) and semi-supervised learning (image classification) tasks. All of the experimental results have been obtained on an Intel Core i3 3.30 GHz WIN7 machine with 8GB memory.

Algorithm 1: LADMAP for solving non-negative sparse Laplacian regularized LRR

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

**Output:** \( Z^*, E^* \)

**Initialization:** \( Z_0 = E_0 = J_0 = Y^0 = Y_2^0 = 0, \lambda = 0.02, \beta = 1.0, \gamma = 5.0, \rho_0 = 2.5, \mu_0 = 10^{-6}, \mu_{\text{max}} = 10^{6}; \varepsilon_1 = 10^{-6}, \varepsilon_2 = 10^{-2}, \eta_1 = 1.25 \times \|Y\|_2^2, L^h. \)

While not converged \((k = 0, 1, \ldots)\) do

1) solving \( Z_{k+1} \) while fixing others:

\[ \min_{Z} \|Z\|_+ + \langle \nabla Z_k q, Z - Z_k \rangle + \frac{\eta_k \mu_k}{2} \| Z - Z_k \|_2^2 \]

where

\[ \nabla Z_k = \beta (Z_k L^h T + Z_k L^h) + \mu_k \left( Z_k - J_k + \frac{Y_k^k}{\mu_k} \right) \]

\[ Z_{k+1} = \Theta(\eta_{k+1}^{-1}) (Z_k - \nabla Z_k q/\eta_1) \cdot \]

2) updating \( J_{k+1} \) while fixing others:

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

3) updating \( E_{k+1} \) while fixing others:

\[ E_{k+1} = S_{\frac{\lambda}{\mu_k}} \left( Y - Y Z_{k+1} + \frac{1}{\mu_k} Y_k^k \right) \]

4) updating Lagrange multipliers \( Y_1 \) and \( Y_2 \):

\[ Y_1^{k+1} = Y_1^k + \mu_k (Y - Y Z_{k+1} - E_{k+1}) \]

\[ Y_2^{k+1} = Y_2^k + \mu_k (Z_{k+1} - J_{k+1}) \]

5) updating \( \mu_{k+1} \):

\[ \mu_{k+1} = \min \left\{ \mu_{\text{max}}, \rho_k \mu_k \right\} \]

where

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

6) checking convergence, if

\[ \| Y - Y Z_{k+1} - E_{k+1} \| / \| Y \| < \varepsilon_1 \text{ or} \]

\[ \| J_{k+1} - J_k \|, \| E_{k+1} - E_k \| < \varepsilon_2 \]

End while

A. The Synthetic Example

Firstly, we want to demonstrate the original motivation of this paper by using a synthetic toy problem for experiment. We consider a set of data points constructed in two moons pattern, as shown in Fig.(a). There are two natural clusters, i.e., the two half moons intersecting together. By using the conventional LRR method, we can achieve the result shown in Fig.(b) where the algorithm failed to distinguish the two moons. The accuracy rate of cluster is only 57%. However, by using our proposed Laplacian regularized LRR, i.e., NSLLRR,
we can better separate the mixed data, illustrated in Fig. 1(c), with a high accuracy rate of 96.5%. Clearly the local structural information coded in the Laplacian regularizer assists in learning clusters in the way that the close data points are categorized into the same clusters.

**B. Unsupervised Learning: Image Clustering**

Data clustering is a classification task of assigning samples to different groups and the spectral clustering approaches [36] are most popular. To quantitatively and effectively evaluate the clustering results, we adopt two quantity metrics, the accuracy (AC) and the normalized mutual information (NMI) [7], in our experiments. Given a data point $x_i$, let $L$ and $\hat{L}$ be the ground truth label and the cluster label provided by the cluster approaches, respectively, then the AC measure is defined by

$$AC = \frac{1}{n} \sum_{i=1}^{n} \delta(\hat{L}(i), \text{Map}(L,L)(i))$$

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

The other metric is the normalized mutual information...
between two index sets $K$ and $K'$, defined as,
\[
\text{NMI}(K, K') = \frac{\text{MI}(K, K')}{\max(H(K), H(K'))}
\]
where $H(K)$ and $H(K')$ denote the entropy 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 the two sets of clusters are identical and the value 0 means that two are independent. Different from AC, NMI is invariant with the permutation of labels, namely, does not require the matching processing in advance.

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

1) K-means clustering algorithm (K-means);
2) Normalized cut (Ncut) \[36\];
3) Principle component analysis (PCA);
4) Traditional LRR \[29\];
5) Our proposed Laplacian regularized LRR.

K-means method usually serves as a benchmark for image clustering task. 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 describing the relation among data points. However, the clustering information, obtained by above methods except K-means, cannot be directly used as final clustering results since they lack the clear cluster structures. Therefore, here K-means is adopted to the final label assignment on the learned lower dimensional representation space. The NSLLRR can be regarded as a non-linear manifold clustering algorithm, so it would be useful to compare it with some existing graph-based clustering methods. Here, we compare it with the classical method, locally linear embedding (LLE) \[40\], to demonstrate the clustering effectiveness of our proposed method.

In this experiment, the real world CMU-PIE\[1\] face image database is utilized to assess the performance of our proposed approach. The CMU-PIE is a popular face database, which is widely used in many kinds of learning tasks, and includes 68 subjects with 41,368 face images in total. In this dataset, the size of each sample is 32 $\times$ 32 and each subject is acquired under 13 different poses, 43 different illumination conditions and with 4 different expressions. We only select a part of images by fixing the pose and expression so that, for each subject, we have 21 images under different lighting conditions. For computing convenience, we first apply PCA to reduce the data dimensionality into 64 by keeping most information. The clustering experiments are conducted with varying the 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 I. The bold numbers denotes the best result with certain cluster numbers. We can see that our proposed method outperforms other algorithms, especially when the number of clusters is large. It also shows that the K-means is generally worse than other methods. As can be seen, the results with 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 the use of the K-means algorithm over the learned lower dimensional representation in the clustering step. The similar phenomenon occurs for the LLE method and LRR too.

There are several regularized parameters affecting on the performance of NSLLRR. In the following, we study the influence of penalty parameters $\lambda$, $\gamma$ and $\beta$ on the NMI by setting them to different values. We vary some parameter while keeping other fixed. The NMI results with different settings on CMU-PIE dataset are given in Fig. 2. From the results, we can conclude that our proposed Laplacian regularized LRR outperforms other algorithms over a large range of parameter values. We can also note that the consistent result is insensitive to the varying $\beta$ values. Moreover, the results on $\lambda$ and $\gamma$ show that both sparsity property and low rankness property are critical for data clustering. Thus, we set $\lambda = 0.2$, $\gamma = 5$ and $\beta = 5$ in our experiments.

### C. 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\[2\]. Specifically, PIE is face database and USPS is handwritten digit database. CMU-PIE includes many face images under different illumination conditions and with different expressions as described in the last section. The USPS is composed of 9298 handwritten digit images of size 16 $\times$ 16. Each image is represented by a 256-dimensional vector. Some samples of these datasets are shown in Fig. 3.

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

1) $k$-nearest-neighbor (kNN) based classification method \[38\]: We adopt Euclidean distance as the similarity measure on which the classification processing is carried out for further comparing with our model. In this paper, we set the numbers of nearest neighbors to be 5.
2) LLE-graph based classification method \[40\]: By constructing a LLE-graph with the numbers of nearest neighbors of 8, we can classify the face images efficiently.
3) $l_1$-graph based classification method \[13\]: Cheng et al. proposed to construct robust and datum-adaptive graph
<table>
<thead>
<tr>
<th>Cluster</th>
<th>Accuracy</th>
<th>NMI</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>K-means</td>
<td>PCA</td>
</tr>
<tr>
<td>4</td>
<td>48.60</td>
<td>52.30</td>
</tr>
<tr>
<td>12</td>
<td>46.81</td>
<td>46.81</td>
</tr>
<tr>
<td>20</td>
<td>41.60</td>
<td>57.14</td>
</tr>
<tr>
<td>28</td>
<td>38.40</td>
<td>36.40</td>
</tr>
<tr>
<td>36</td>
<td>35.40</td>
<td>34.60</td>
</tr>
<tr>
<td>44</td>
<td>33.20</td>
<td>33.70</td>
</tr>
<tr>
<td>52</td>
<td>33.20</td>
<td>33.70</td>
</tr>
<tr>
<td>60</td>
<td>33.10</td>
<td>33.20</td>
</tr>
<tr>
<td>68</td>
<td>31.70</td>
<td>32.90</td>
</tr>
</tbody>
</table>

(a) CMU-PIE samples.  
(b) USPS samples.

Fig. 3. Samples of test database.

by utilizing the overall contextual information instead of only pairwise Euclidean distance as conventionally.

4) LRR-graph based classification method: Following [28], we construct the LRR-graph by solving the lowest-rank coefficient matrix and then symmetrizing it as $\ell_1$-graph.

The parameters of LRR are the same as those in [28].

5) Non-negative low-rank and sparse (NNLRS)-graph based classification method: Following [54], 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.

Similar to the LRR-graph based method, our proposed approach is also effective in seeking the low-rank weight matrix so that we can conveniently carry out the classification experiments over the above databases using the learned graph. We denote our approach as NSLLRR-graph based classification method in our experiments.

The purpose of semi-supervised learning task is to reveal more unlabeled information with limited known labeled data. Therefore, we select the percentage of labeled samples ranges from 10% to 60%, instead of the higher percentage of ranging from 50% to 80%. This can help promote the effect of semi-supervised learning task. Empirically, we found that $\lambda = 0.08$, $\gamma = 2$ and $\beta = 5$ are good choices for these parameters. Therefore, we did not tune the parameter settings in these experiments individually. The classification results, for CMU-PIE and USPS, are reported in Table II and Table III respectively. The bold numbers in each table denotes the best performance under different labeling percentages. From these results, we can see that our proposed NSLLRR almost consistently outperforms other methods, especially for CMU-PIE database. This suggests our proposed method is effective for semi-supervised learning task. As well, this shows the manifold regularizer actually enhances the robustness of the LRR graph-based method.

VI. HYPERGRAPH REGULARIZED LRR FOR SEMI-SUPERVISED LEARNING

In the real world, there usually exists a co-occurrence relation that involves more than pairwise [34], which is more effectively represented by a hypergraph. In this section, we consider some categorical data including some instances with one or more attributes, i.e., Zoo dataset, which comes from the UCI Machine Learning Depository.

The Zoo dataset covers 100 animals with 17 Boolean-valued attributes, where the attributes contain hair, feathers, eggs, milk, legs, tail, etc. Moreover, these animals data have been manually grouped into 7 different class beforehand by the provider. Note that we found there were 2 repeated instances of “frog” in dataset, so we removed the repeated items manually in our experiment. That is, the test dataset contains 100 instances in our experiment. In order to evaluate our proposed hypergraph regularized LRR, we conduct a semi-supervised classification task on the test Zoo dataset, where we define a hypergraph on Zoo dataset with 36 hyper-edges.

As referred in [34], we construct a hypergraph by taking instances as vertices and creating a hyperedge for each value of the attributes of Zoo. For instance, attribute 2 (hair) is

3http://archive.ics.uci.edu/ml/datasets/Zoo
TABLE II
CLASSIFICATION ERROR RATES (%) FOR CMU-PIE BASED ON VARIOUS GRAPHS UNDER DIFFERENT PERCENTAGES OF LABELED SAMPLES.

<table>
<thead>
<tr>
<th>Labeling percentage#</th>
<th>kNN [38]</th>
<th>LLE-graph</th>
<th>$\ell_1$ graph</th>
<th>LRR-graph [28]</th>
<th>NNLRS-graph [54]</th>
<th>NSLLRR-graph</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>34.84</td>
<td>33.06</td>
<td>22.88</td>
<td>47.30</td>
<td><strong>11.11</strong></td>
<td>35.66</td>
</tr>
<tr>
<td>20</td>
<td>37.46</td>
<td>35.05</td>
<td>22.94</td>
<td>21.60</td>
<td>22.81</td>
<td><strong>9.28</strong></td>
</tr>
<tr>
<td>30</td>
<td>35.3</td>
<td>32.52</td>
<td>22.33</td>
<td>11.80</td>
<td>17.86</td>
<td><strong>5.83</strong></td>
</tr>
<tr>
<td>40</td>
<td>35.81</td>
<td>32.51</td>
<td>23.14</td>
<td>7.80</td>
<td>16.25</td>
<td><strong>3.36</strong></td>
</tr>
<tr>
<td>50</td>
<td>34.39</td>
<td>31.41</td>
<td>23.01</td>
<td>4.30</td>
<td>19.25</td>
<td><strong>2.10</strong></td>
</tr>
<tr>
<td>60</td>
<td>35.63</td>
<td>32.76</td>
<td>25.76</td>
<td>2.50</td>
<td>21.56</td>
<td><strong>1.62</strong></td>
</tr>
</tbody>
</table>

TABLE III
CLASSIFICATION ERROR RATES (%) FOR USPS BASED ON VARIOUS GRAPHS UNDER DIFFERENT PERCENTAGES OF LABELED SAMPLES.

<table>
<thead>
<tr>
<th>Labeling percentage#</th>
<th>kNN [38]</th>
<th>LLE-graph</th>
<th>$\ell_1$ graph</th>
<th>LRR-graph [28]</th>
<th>NNLRS-graph [54]</th>
<th>NSLLRR-graph</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>11.97</td>
<td>17.10</td>
<td>43.27</td>
<td>67.00</td>
<td>11.57</td>
<td><strong>10.55</strong></td>
</tr>
<tr>
<td>20</td>
<td>12.31</td>
<td>22.92</td>
<td>41.27</td>
<td>48.30</td>
<td>9.30</td>
<td><strong>5.90</strong></td>
</tr>
<tr>
<td>30</td>
<td>5.88</td>
<td>21.26</td>
<td>38.31</td>
<td>33.60</td>
<td>4.95</td>
<td><strong>4.80</strong></td>
</tr>
<tr>
<td>40</td>
<td>7.87</td>
<td>19.21</td>
<td>34.86</td>
<td>25.80</td>
<td>7.44</td>
<td><strong>4.01</strong></td>
</tr>
<tr>
<td>50</td>
<td>17.19</td>
<td>18.41</td>
<td>29.42</td>
<td>15.50</td>
<td>11.27</td>
<td><strong>3.20</strong></td>
</tr>
<tr>
<td>60</td>
<td>11.04</td>
<td>14.80</td>
<td>23.36</td>
<td>10.00</td>
<td>6.09</td>
<td><strong>2.05</strong></td>
</tr>
</tbody>
</table>

In our experiment, the labels of some vertices in hypergraph are supposed to be known. For fair comparison, the labeled vertices are randomly chosen to ensure that all the labels appear. For the labeling percentage, we apply similar partition as the case for normal graph, i.e., from 10% to 60%. We run 10 times for each algorithm with randomly labeled data sets and then average the results of error rates corresponding to each labeled percentage. We empirically set the parameters as follows $\lambda = 0.01$, $\gamma = 0.1$ and $\beta = 0.001$. To also assess the capacity of learning the lowest rank representation, we compared our proposed method NSHLRR with the selected methods LRR-graph and normal graph based NSLLRR. Moreover, we would like to compare with $k$NN-graph based method which is usually used as a benchmark in this field. As like aforementioned experiments, we select 5 of the nearest neighbors for each data point.

The classification results are reported in Table IV and Fig. 4 shows the plot of classification error rate versus the percentage of labeled points for Zoo dataset. As can be seen, the hypergraph regularized NSHLRR method consistently outperforms the LRR and the benchmark $k$NN graph based method. Among these methods, $k$NN graph based method produces the worst classification performance. NSHLRR is significantly better than all the other methods except NSLLRR in the cases of 30% and 40% labeled points. This result shows that a hypergraph is more sensible to modeling data with more complex relations than a normal graph approach since the complicated interactions in data will be well taken into account by a hypergraph. The experiment has demonstrated that our proposed Laplacian regularized LRR model not only can represent the globally linear structures, but also preserve the local geometrical manifold structures in data by incorporating a hypergraph regularizer and have capability of learning high order relations (i.e., three-way or higher) in the semi-supervised settings.

TABLE IV
CLASSIFICATION ERROR RATES (%) FOR ZOO BASED ON VARIOUS METHODS UNDER DIFFERENT PERCENTAGES OF LABELED SAMPLES.

<table>
<thead>
<tr>
<th>Labeling percentage#</th>
<th>$k$NN-graph [28]</th>
<th>LRR-graph [28]</th>
<th>normal NSLLRR</th>
<th>hyper NSHLRR</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>68.00</td>
<td>43.01</td>
<td>41.00</td>
<td><strong>27.00</strong></td>
</tr>
<tr>
<td>20</td>
<td>56.00</td>
<td>36.00</td>
<td>17.00</td>
<td><strong>13.00</strong></td>
</tr>
<tr>
<td>30</td>
<td>50.00</td>
<td>26.00</td>
<td><strong>11.00</strong></td>
<td>15.00</td>
</tr>
<tr>
<td>40</td>
<td>38.00</td>
<td>16.00</td>
<td><strong>10.00</strong></td>
<td>12.00</td>
</tr>
<tr>
<td>50</td>
<td>33.00</td>
<td>11.00</td>
<td><strong>11.00</strong></td>
<td>9.00</td>
</tr>
<tr>
<td>60</td>
<td>25.00</td>
<td>10.00</td>
<td>7.00</td>
<td><strong>7.00</strong></td>
</tr>
</tbody>
</table>

Fig. 4. Classification error rate versus the labeled percentage on Zoo dataset based on LRR, normal graph NSLLRR and NSHLRR.
VII. HYPERGRAPH EMBEDDING

In this section, we consider the task of hypergraph embedding that projects the vertices of \( G \) into a low dimensional Euclidean space where the vertices are best characterized by the similarity relationship between the vertex pairs in the hypergraph \([34]\). In graph-spectral approach, the leading eigenvectors of the graph adjacency matrix is often exploited to implement graph embedding \([33]\). Furthermore, by applying our proposed methods, the graph adjacency structure and graph weight matrix are readily derived from the previous learned \( Z^* \). Thus, the embedding of vertex \( v \) in a \( k \)-dimensional space is just the row vector at the leading eigenvectors of the graph weight matrix \([52]\).

To evaluate the effectiveness of our proposed methods on hypergraph embedding, we first use both hypergraph-based NSHLRR and normal graph-based NSLLRR to learn the representation \( Z \) of high dimensional data respectively. Then, we compute the embeddings in light of the learned graph weight matrix. In this experiment, we still apply Zoo dataset as that in the previous section \([VI]\) and project its vertices into 3-dimensional space. The embedding results are visualized in Fig. 5. As can be seen, the embeddings with the graph weight matrix learned by normal graph-based NSLLRR and hypergraph-based NSHLRR are different. Although the two methods can generate a separation between different classes in a 2-dimensional space, it is clear that hypergraph-based NSHLRR can handle better than that of normal graph-based NSLLRR. Since some samples (e.g., tortoise and tuatara) have special attributions, the embedding with hypergraph-based NSHLRR fails to give them clear categorization. However, on the whole, the visual result in Fig. 5 can demonstrate the efficacy of our proposed methods.

VIII. CONCLUSION

In this paper, we propose a general Laplacian regularized low-rank representation method, in which we explicitly considers the manifold structures of the data space by introducing Laplacian regularization term. Two different graph regularizers are introduced into the conventional LRR objective function, i.e., normal graph and hypergraph Laplacian regularizers. Our proposed method not only can represent the global low-rank structures, but can capture the intrinsic geometrical information in the data. It is interesting to consider applying our proposed methodology to a broader range of practical problems, such as clustering and semi-supervised learning tasks. The extensive experimental results on semi-supervised image clustering and classification show the efficacy of our proposed method.

ACKNOWLEDGMENTS

The research project is supported by the Australian Research Council (ARC) through the grant DP130100364. Ming Yin’s work is supported by the Foundation of Key Laboratory of Autonomous Systems and Networked Control, Ministry of Education, P.R. China (No. 2013A 06).

REFERENCES

[6] Deng Cai, Xiaofei He, Jiawei Han, and Thomas S. Huang. Graph regularized non-negative matrix factorization for data representation. IEEE Transactions on Pattern Analysis and Machine Intelligence, 33(8):1548–1560, 2011.
[7] Deng Cai, Xiaofei He, Xuanhui Wang, Hujun Bao, and Jiawei Han. Locality preserving nonnegative matrix factorization. In Proc. 2009 Int. Joint Conference on Artificial Intelligence (IJCAI’09), 2009.
IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE


[19] Xiaofei He, Deng Cai, Yuanlong Shao, Hujun Bao, and Jiawei Han. Laplacian regularized gaussian mixture model for data clustering. IEEE Transactions on Knowledge and Data Engineering, 23(9):1406–1418, 2011.


[34] Li Pu and Boi Faltings. Hypergraph Learning with Hyperedge Expansion. In ECML-PKD, 2012.


Ming Yin received the Ph.D. degree in information and communication engineering from Huazhong University of Science and Technology (HUST), Wuhan, China, in 2006. He worked as a visiting scholar at the School of Computing and Mathematics, Charles Sturt University, Bathurst, Australia, from Jan. 2012 to Dec. 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 graduated from Huazhong University of Science and Technology (HUST), China in 1982 with BSc. degree in Computational Mathematics and obtained PhD from Dalian University of Technology, China in 1991. He is a Professor in Computing Science in the School of Computing and Mathematics at Charles Sturt University, Australia. He was a senior lecturer, a lecturer in Computer Science from 2001 to 2005 at University of New England, Australia. From 1982 to 2001 he was an associate lecturer, lecturer, associate professor and professor in Department of Mathematics at HUST. His main research interests include machine learning, data mining, Bayesian learning and inference, and image analysis.

Zhouchen Lin received the Ph.D. degree in applied mathematics from Peking University in 2000. He is currently a Professor at Key Laboratory of Machine Perception (MOE), School of Electronics Engineering and Computer Science, Peking University. He is also a Chair Professor at Northeast Normal University. Before March 2012, he was a Lead Researcher at Visual Computing Group, Microsoft Research Asia. He was a guest professor at Shanghai Jiaotong University, Beijing Jiao Tong University and Southeast University. He was also a guest researcher at Institute of Computing Technology, Chinese Academy of Sciences. His research interests include computer vision, image processing, computer graphics, machine learning, pattern recognition, and numerical computation and optimization. He is an area chair of CVPR 2014, an associate editor of International Journal of Computer Vision, and a Senior member of the IEEE.