Journal of Information & Computational Science 8: 5 (2011) 725–737
Available at http://www.joics.com

Mixture of Manifolds Clustering via Low Rank Embedding ⋆

Risheng Liu [a,b], Ruru Hao [a], Zhixun Su [a,*]

[a] *School of Mathematical Sciences, Dalian University of Technology, Dalian 116024, China*
[b] *Robotics Institute, Carnegie Mellon University, Pittsburgh 15213, USA*

Abstract

Locally Linear Embedding (LLE) is an effective method for both single manifold embedding and multiple manifolds clustering. However, LLE may fail to model data drawn from mixture of manifolds because a point and its Nearest Neighbors (NNs) cannot guaranteed to be on the same manifold, especially when the point is at the junction of different manifolds. In this paper, we propose a new method, named Low Rank Embedding (LRE), to model the geometric structure of mixture of manifolds. LRE defines the neighborhood structure as linear combination of other samples with an additional minimum rank constraint on the points that belong to the same manifold. This way, LRE removes the dependency with the number and metric of NNs, and it makes a more accurate selection of the neighbors belonging to the same manifold. Both theoretical and experimental results show that LRE is a promising tool for embedding and clustering mixture of manifolds.

Keywords: Low Rank; Mixture of Manifolds; Clustering and Embedding

1 Introduction

High-dimensional data analysis has become increasingly important in a wide range of scientific fields, such as machine learning, pattern recognition and computer vision. Techniques for analyzing and modeling high-dimensional data have included both dimensionality reduction and clustering.

1.1 Previous Work

Dimensionality reduction is to transfer data in high-dimensional sample space to a low-dimensional feature space. Manifold learning, as one of the most important nonlinear dimensionality reduction techniques, assumes that some geometric properties of *single* manifold in the original sample

⋆Project supported by the National Nature Science Foundation of China (No. U0935004).
*Corresponding author.
Email addresses: rsliu@andrew.cmu.edu (Risheng Liu), hao@dlut.edu.cn (Ruru Hao), zxsu@dlut.edu.cn (Zhixun Su).

May 2011

space should be preserved in the feature space, e.g., ISOMAP [1] preserves geodesic distance and LLE [2] preserves the locally linear relationships.

In practice, however, the given data set usually can not be well described by one single data structure (see Fig. 1 for an example). A more reasonable model is to consider data set as lying in different groups and the membership of the data points to groups might be unknown, leading to the problem of clustering. Traditional methods, such as EM and K-means, have been very successful in clustering low-dimensional data with simple group structure, such as Gaussian distributed. However, when data becomes high dimensional, its structure can be very complex. Typically, data is often distributed on mixture of manifolds. The traditional methods cannot work well for data with such a complex structure.

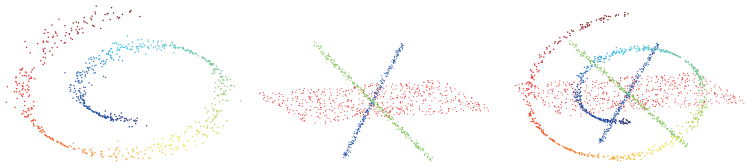


Fig. 1: Examples of different data structures. (a) Single manifold. (b) Mixture of subspaces. (c) Mixture of manifolds. Informally, we refer to mixture of manifolds as a union of mixed manifolds of different dimensionalities.

By using linear manifold (subspace) to approximate the group structure, a number of approaches, known as subspace clustering, have been proposed for this kind of data. According to the mechanisms of modeling the data, existing work can be roughly divided into four categories: algebraic, statistical, spectral clustering based, and embedding based. Generalized Principal Component Analysis (GPCA) [3] is an algebraic method for clustering data lying on subspaces. The idea behind GPCA is that one can fit the data with polynomials. Statistical approaches, such as Mixtures of Probabilistic PCA (MPPCA) [4] and Random Sample Consensus (RANSAC) [5], model data as a set of independent samples drawn from a mixture of probabilistic distributions, i.e., adopting a mixture of Gaussian model. Then the problem of clustering is converted to a model estimation problem. Spectral clustering [6] based methods first construct an affinity matrix to model the similarity among the data set. Then the clusters of the data set can be obtained by applying spectral clustering to the affinity matrix. One of the main challenges in applying spectral clustering is how to define a good affinity matrix. Factorization methods [7] build the affinity matrix by factorizing the data matrix. Recently proposed Sparse Subspace Clustering (SSC) [8] borrows the sparsity principle from compressive sensing and calculates the sparsest representation of the data set as the affinity matrix for clustering. Low Rank Representation (LRR) [9] is very similar to SSC, except that it aims to find a low rank representation instead of a sparse representation.

Different from the above mentioned methods, embedding based clustering techniques combine nonlinear dimensionality reduction with clustering technique. The major advantage of these methods is that they are applicable to nonlinear manifolds. [10] is an EM-like extension of ISOMAP for clustering manifolds. This method is very sensitive to initialization. The work of [11] first uses LLE [2] to learn a low-dimensional representation and clusters data in this embedding space. Locally Linear Manifold Clustering (LLMC) [12] studies the mathematical properties of the LLE embedding for separated manifold clustering. LLE based methods have

to assume that a point and its K nearest neighbors (K-NNs) are from the same manifold. This clearly cannot be guaranteed, especially when the point is at the junction of different manifolds. Moreover, choosing the correct number K of nearest neighbors is also a challenge.

1.2 Our Contributions

In this paper, we propose a new method, called Low Rank Embedding (LRE), to recover the intrinsic structure of mixture of manifolds modeled data. Our idea is from two sources: the success of linear reconstruction for single manifold learning and the viewpoint from compressive sensing. We extend the linear reconstruction of LLE to mixture of manifolds and introduce a low rank constraint for computing the reconstruction coefficients. Based on our new criterion, the separable structure of mixture of manifolds can be transferred to a low-dimensional space and we can successfully cluster data into different groups in the low-dimensional space by the traditional clustering technique (e.g., K-means). The LRE algorithm has several advantages over previous methods. For example, the LLE based methods cannot guarantee that the K-NNs are on the same manifold, and if K is not appropriately chosen a data point may not be an exact linear combination of its K-NNs because the dimensions of the manifolds may be different. The low rank criterion in LRE can overcome this drawback as it is a structured way to obtain linear reconstruction. It simultaneously obtains the neighbors and the reconstruction coefficients by solving a rank regularized linear reconstruction problem. As rank is a global property, the low rank criterion can better discover the global structure of mixture of manifolds. Namely, the neighbors are found based on the similarity computed from the data structure. In contrast, K-NNs are found based on the local pre-defined distance which cannot reveal the data structure. Compared to other compressive sensing based methods, such as SSC and LRR, LRE performs clustering in the transformed feature space via a nonlinear dimensionality reduction, while SSC and LRR perform clustering using the reconstruction coefficients directly. Moreover, LRE can handle mixture of nonlinear manifolds, while SSC and LRR were designed for mixture of subspaces. Our theoretic analysis also shows that LRE can not only remove the noise from the data set but also significantly reduce the error in the reconstruction coefficients matrix, which is another important advantage of LRE over SSC and LRR.

2 Mixture of Manifolds Clustering

In this section, we consider modeling the structure of data sampled from mixture of manifolds and provide details of LRE. A variety of matrix norms are used throughout this paper: $\|\mathbf{X}\|_*$ is the nuclear norm (the sum of singular values); $\|\mathbf{X}\|$ is the spectral norm (the largest singular value); $\|\mathbf{X}\|_F$ is the Frobenius norm (the l_2-norm of the vector of singular values).

2.1 Manifold Reconstruction: from Locally Linear to Low Rank

Traditional locally linear based manifold learning methods, such as LLE, assume that the point on a manifold can be well approximated by its K-NNs and the manifold structure can be preserved by the reconstruction coefficients. So each point $\mathbf{x}_j$ on a manifold can be approximated by the

linear combination of its nearest neighbors measured by distances: $\mathbf{x}_j \approx \sum_i \mathbf{R}_{ij}\mathbf{x}_i$. For a point set $\mathbf{X} = [\mathbf{x}_1, ..., \mathbf{x}_n]$, we can formulate the above linear combination in a matrix form: $\mathbf{X} \approx \mathbf{XR}$, where the j-th column of $\mathbf{R}$ is the reconstruction coefficients of $\mathbf{x}_j$. This locally linear reconstruction criterion can successfully recover the nonlinear structure of a single manifold. However, for mixture of manifolds (e.g., Fig. 2(a)), the distance-based neighbors of a point near the junctions do not all belong to the same manifold (e.g.,"A" and "B" in Fig. 2(b)). So the locally linear criterion may fail to model the structure of mixture of manifolds, especially for junctions. To overcome this drawback, in the following subsection, a rank based criterion is introduced for mixture of manifolds reconstruction.

2.2 Mixture of Manifolds Embedding via Rank Regularized Linear Reconstruction

From now on, let $\mathbf{X} = [\mathbf{X}_1, ..., \mathbf{X}_s]$ be a given set of data points in $\mathbb{R}^d$, which are drawn from a mixture of manifolds, namely a union of $s \geq 1$ mixed manifolds $\{\mathbf{X}_k\}_{k=1}^s$ of unknown dimensions $d_k = \dim(\mathbf{X}_k)$, $0 < d_k < d$, $k = 1, .., s$. LRE aims at learning a low-dimensional embedding $\mathbf{Y}$ in $\mathbb{R}^l$ ($\max_k(d_k) \leq l < d$), in which the similarity of each manifold part can be preserved, while the separable structure of mixture of manifolds can be enhanced.

We model each manifold data $\mathbf{X}_k$ based on the following considerations: *(a) Each point on the mixture of manifolds should only be reconstructed by a linear combination of points belonging to the same manifold (an extension of LLE for mixture of manifolds). (b) All the combination coefficients of points belonging to the same manifold should have similar underlying structure (this consideration will be illustrated by an example in Fig. 2).* Translating statement (a) into mathematics, we have

$$\mathbf{X}_k \approx \mathbf{X}_k\mathbf{R}_k,$$

where $\mathbf{R}_k$ is the reconstruction coefficients matrix. As discussed in the previous subsection, the locally linear criterion is not valid for modeling mixture of manifolds due to the intersection between manifolds. By the statement (b), we propose a new constraint for $\mathbf{R}_k$. As all the reconstruction coefficients of $\mathbf{X}_k$ (columns of $\mathbf{R}_k$) have similar underlying structure, low rankness may be a more appropriate criterion for constraining $\mathbf{R}_k$. The above analysis immediately suggests that we should seek the lowest rank reconstruction matrix $\mathbf{R}_k$ that could minimize the linear reconstruction error. Therefore, the optimization problem for each manifold is:

$$\min_{\mathbf{R}_k} \frac{1}{2}\|\mathbf{X}_k - \mathbf{X}_k\mathbf{R}_k\|_F^2 + \lambda \cdot \mathrm{rank}(\mathbf{R}_k), \tag{1}$$

where $\lambda > 0$ is used to balance the effects of the two parts. The above optimization is difficult to solve due to the discrete nature of the rank function. As suggested by the matrix completion method [13], the following convex relaxation provides a good surrogate for problem (1):

$$\min_{\mathbf{R}_k} \mathcal{J}_{\mathbf{X}_k}(\mathbf{R}_k) = \min_{\mathbf{R}_k} \frac{1}{2}\|\mathbf{X}_k - \mathbf{X}_k\mathbf{R}_k\|_F^2 + \lambda\|\mathbf{R}_k\|_*. \tag{2}$$

The low rank coefficients matrix $\mathbf{R}_k$ that reconstruct $\mathbf{X}_k$ in the d-dimensional space should reconstruct its embedding coordinates in the l-dimensional space. Hence $\mathbf{R}_k$ are kept fixed and

the embedded subspace $\mathbf{Y}_k$ is sought by minimizing the following cost function:

$$\min_{\mathbf{Y}_k} \mathcal{J}_{\mathbf{R}_k}(\mathbf{Y}_k) = \min_{\mathbf{Y}_k} \frac{1}{2}\|\mathbf{Y}_k - \mathbf{Y}_k\mathbf{R}_k\|_F^2. \tag{3}$$

As we aim at deriving the global reconstruction of the whole data set, we add all the reconstruction errors together to have:

$$\mathcal{J}_{\mathbf{X}}(\mathbf{R}) = \sum_{k=1}^{s} \mathcal{J}_{\mathbf{X}_k}(\mathbf{R}_k) = \tfrac{1}{2}\|\mathbf{X} - \mathbf{X}\mathbf{R}\|_F^2 + \lambda\|\mathbf{R}\|_*, \tag{4}$$

where

$$\mathbf{R} = \begin{pmatrix} \mathbf{R}_1 & & 0 \\ & \ddots & \\ 0 & & \mathbf{R}_s \end{pmatrix},$$

and the low-dimensional embedding can be learnt by minimizing the following cost function:

$$\mathcal{J}_{\mathbf{R}}(\mathbf{Y}) = \sum_{k=1}^{s} \mathcal{J}_{\mathbf{R}_k}(\mathbf{Y}_k) = \tfrac{1}{2}\|\mathbf{Y} - \mathbf{Y}\mathbf{R}\|_F^2. \tag{5}$$

In this way, we obtain the following two optimization problems for mixture of manifolds embedding:

$$\min_{\mathbf{R}} \frac{1}{2}\|\mathbf{X} - \mathbf{X}\mathbf{R}\|_F^2 + \lambda\|\mathbf{R}\|_*, \tag{6}$$

and

$$\min_{\mathbf{Y}} \frac{1}{2}\|\mathbf{Y} - \mathbf{Y}\mathbf{R}\|_F^2 \ s.t.\ \mathbf{Y}^T\mathbf{Y} = \mathbf{I}, \tag{7}$$

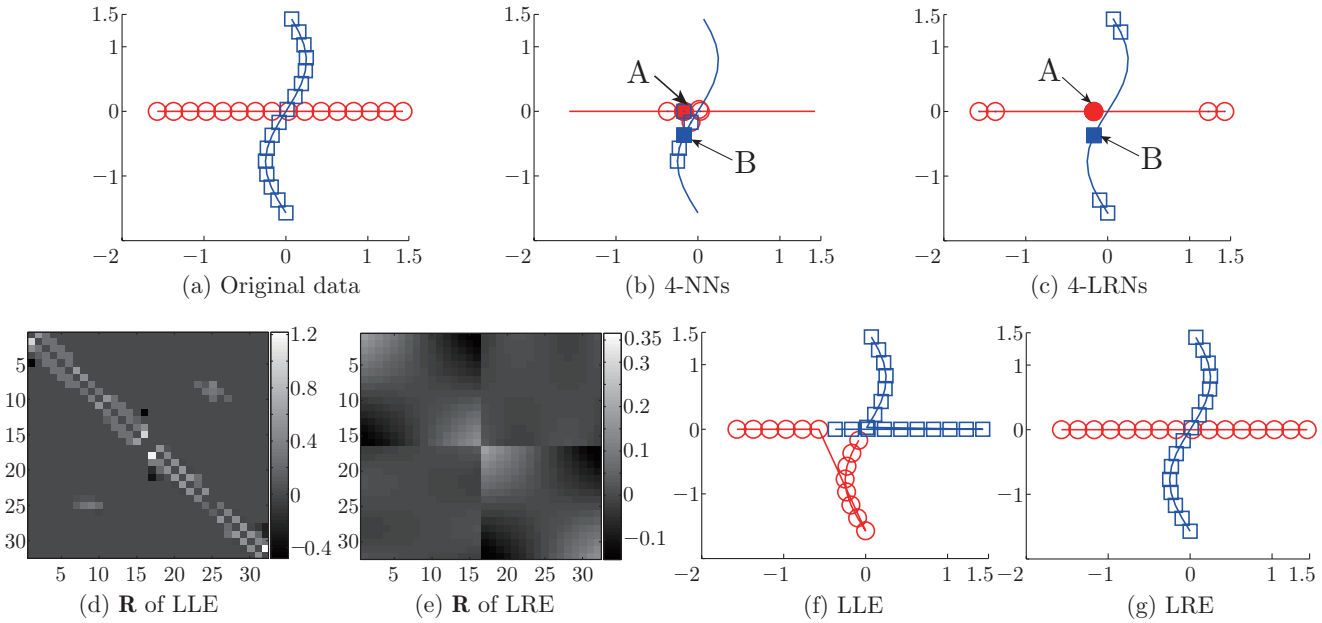
where constraint $\mathbf{Y}^T\mathbf{Y} = \mathbf{I}$ is imposed to make the problem well-posed.

In Fig. 2, we refer to an example to illustrate the benefits of LRE to model the structure of the mixture of manifolds versus LLE. One can see that the distance-based neighbors of points near the junctions may belong to different manifolds (Fig. 2(b)), while our low rank neighbors (LRNs)[1] all belong to its own manifold (Fig. 2(c)). We also compare the $\mathbf{R}$'s calculated by LLE and LRE. As shown in Fig. 2(d), each column of $\mathbf{R}$ calculated by LLE only have 4 nonzero elements and the rank is 32. However, it can be seen in Fig. 2(e) that $\mathbf{R}$ calculated by LRE has a block structure and its rank is 2. This structure successfully reveals the memberships of the samples: the reconstruction coefficients of points belong to the same manifold are nonzero and have similar underlying structure (linearly dependent), and the coefficients of points belong to different manifolds are zero (linearly independent). All above results verify that our model is more effective for modeling the structure of mixture of manifolds.

2.3 Mixture of Manifolds Clustering in Low-dimensional Space via K-means

After solving problems (6) and (7), we apply K-means to the low-dimensional embedding $\mathbf{Y}$ for data clustering.

[1]The K-LRNs of the j-th point in Fig. 2 are the points associated with the K largest absolute values of the j-th column of $\mathbf{R}$.



(a) Original data (b) 4-NNs (c) 4-LRNs

(d) **R** of LLE (e) **R** of LRE (f) LLE (g) LRE

Fig. 2: Illustration of the LRE and LLE for modeling mixture of manifolds. (a) The original data with true labels, where 32 points are drawn from one line (circles) and one curve (square). (b) The four nearest-neighbors of point "A" (point "A" is the solid circle and its neighbors are hollow circles) and "B" (point "B" is the solid square and its neighbors are hollow squares), respectively. (c) The four low-rank neighbors of points "A" and "B", respectively. (d) The **R** calculated by LLE, whose rank is 32. (e) The **R** calculated by LRE, whose rank is 2. (f) LLE based clustering result, whose accuracy is 56.25%. (g) LRE based clustering result, whose accuracy is 100%.

We have found that the separable structure of the embedding can be enhanced if we convert each column of $\mathbf{R}$ to convex combination coefficients before computing the low-dimensional embedding $\mathbf{Y}$. Namely, the $\mathbf{R}$ in (7) is replaced by $\tilde{\mathbf{R}}$ with $\tilde{\mathbf{R}}_{ij} = |\mathbf{R}_{ij}| / \sum_i |\mathbf{R}_{ij}|$. This is because such a postprocessing can make the features in the low-dimensional space more globular. So clustering by K-means can be more effective.

Algorithm 1 summarizes the whole algorithm for mixture of manifolds clustering.

Algorithm 1 LRE-based Mixture of Manifolds Clustering Algorithm

Input: Data matrix $\mathbf{X}$, the number of manifolds s.

Step 1: Obtain the reconstruction coefficients matrix $\mathbf{R}$ by solving (6). Then convert the columns of $\mathbf{R}$ into convex linear combination coefficients.

Step 2: Obtain the low-dimensional embedding $\mathbf{Y}$ by solving (7).

Step 3: Use K-means to cluster the columns of $\mathbf{Y}$ into s groups.

3 Theoretical Analysis and Numerical Solution

In this section, we state the theoretical analysis on problem (6) and present an efficient algorithm to solve our LRE model.

3.1 Theoretical Analysis

Different from traditional methods which only consider the noise in the data matrix $\mathbf{X}$, our theoretical analysis shows that the nuclear norm regularization in LRE also reduces the errors in the reconstruction coefficients matrix $\mathbf{R}$.

We assume that the data matrix $\mathbf{X}$ has the restricted isometry property (RIP) [14] with an isometry constant $\delta_r \in [0, 1]$. Namely, the following inequalities

$$(1 - \delta_r)\|\mathbf{B}\|_F^2 \leq \|\mathbf{XB}\|_F^2 \leq (1 + \delta_r)\|\mathbf{B}\|_F^2 \tag{8}$$

hold for all matrices $\mathbf{B} \in \mathbb{R}^{n\times n}$ of rank at most r. Note that we may scale $\mathbf{X}$ appropriately so that is satisfies the RIP.

Now we present the error bound in estimating the reconstruction coefficients matrix $\mathbf{R}$.

Theorem 1 *Assume that rank*$(\mathbf{R}) \leq r$ *and let* $\hat{\mathbf{R}}$ *be the solution to (6). If* $\delta_{4r} < (3\sqrt{2} - 1)/17$ *and* $\|\mathbf{X}^T\mathbf{E}\| \leq \lambda/2$*, then*

$$\|\hat{\mathbf{R}} - \mathbf{R}\|_F^2 \leq c_{\delta_{4r}} r\lambda^2, \tag{9}$$

where $\mathbf{E} = \mathbf{X} - \mathbf{XR}$ *and* $c_{\delta_{4r}}$ *is a constant depending on the isometry constant* δ_{4r} *only. In particular, if* $\mathbf{E}$ *is a Gaussian error with standard variance* σ *and* $\lambda = 16m\sigma$*, where* $m = \max(d, n)$*, we have*

$$\|\hat{\mathbf{R}} - \mathbf{R}\|_F^2 \leq c'_{\delta_{4r}} mr\sigma^2, \tag{10}$$

with probability at least $1 - 2e^{-cm}$ *for a constant* $c'_{\delta_{4r}}$ *depending on* δ_{4r} *only and a fixed constant* $c > 0$*.*

Remark[2]: Theorem 1 actually implies that the low rank structure helps to remove noise in $\mathbf{R}$. For example, if we measure each entry of $\mathbf{R}$ (a measurement ensemble with an isometry constant $\delta_r = 0$), but each measurement is corrupted by $\mathcal{N}(0, \sigma^2)$ noise, then taking the measurements as they are as the estimate of $\mathbf{R}$ would lead to an expected error equal to

$$\mathbb{E}\|\hat{\mathbf{R}} - \mathbf{R}\|^2 = m^2\sigma^2. \tag{11}$$

In comparison, the estimate (10) by nuclear norm regularization is by a factor of about r/m smaller than that in (11).

3.2 Numerical Solution

Now we show how to solve problem (6) and (7). With an auxiliary variable $\mathbf{Z}$, problem (6) is equivalent to

$$\min_{\mathbf{Z},\mathbf{R}} \frac{1}{2}\|\mathbf{X} - \mathbf{XR}\|_F^2 + \lambda\|\mathbf{Z}\|_*, \ s.t.\ \mathbf{R} = \mathbf{Z}. \tag{12}$$

It can be efficiently solved, with iterations, by the inexact augmented Lagrange multiplier (IALM, [15], also called the alternating direction method (ADM) in the literature) method. Namely,

[2]For more details about this remak, the reader is advised to see [14] and the references therein.

introducing the augmented Lagrange function $\mathbf{L}$ of problem (12):

$$\begin{aligned}\mathcal{L}_A(\mathbf{Z},\mathbf{R},\mathbf{L}) &= \tfrac{1}{2}\|\mathbf{X}-\mathbf{X}\mathbf{R}\|_F^2+\lambda\|\mathbf{Z}\|_*\\ &+\langle\mathbf{L},\mathbf{R}-\mathbf{Z}\rangle+\tfrac{\beta}{2}\|\mathbf{R}-\mathbf{Z}\|_F^2,\end{aligned} \tag{13}$$

$\mathbf{Z}$ and $\mathbf{R}$ are both updated by minimizing $\mathcal{L}_A$ with the rest variables fixed, and the Lagrange multiplier $\mathbf{L}$ is updated by adding the error in the constraints. More specifically, the updating schemes can be found to be:

$$\begin{cases}\mathbf{Z}^+ = \arg\min\limits_{\mathbf{Z}} \frac{\lambda}{\beta}\|\mathbf{Z}\|_* + \frac{1}{2}\|\mathbf{Z}-(\mathbf{R}+\frac{\mathbf{L}}{\beta})\|_F^2,\\ \mathbf{R}^+ = (\mathbf{X}^T\mathbf{X}+\beta\mathbf{I})^{-1}(\mathbf{X}^T\mathbf{X}-\mathbf{L}+\beta\mathbf{Z}^+),\\ \mathbf{L}^+ = \mathbf{L}+\beta(\mathbf{R}^+-\mathbf{Z}^+),\end{cases} \tag{14}$$

where the superscripts "+" denote the values are updated. The entire procedure is summarized in Algorithm 2. The solution for $\mathbf{Z}^+$ is by singular value thresholding [16].

For optimization problem (7), subject to constraint that makes the problem well-posed, it can be minimized by solving a sparse eigenvector problem, whose bottom l non-zero eigenvectors provide the solution.

Algorithm 2 Solving Problem (6) via the Inexact ALM [15]

Input: Data matrix $\mathbf{X}\in\mathbb{R}^{d\times n}$, parameter λ.

Initialize: Set $\mathbf{R}$, $\mathbf{Z}$, and $\mathbf{L}$ to zero matrices of size $n\times n$, $\beta>0$ and $\varepsilon=10^{-8}$.

while not converged **do**

Step 1: Update ($\mathbf{Z}$, $\mathbf{R}$, $\mathbf{L}$) by (14).

Step 2: Check the convergence conditions

$\|\mathbf{R}^+-\mathbf{R}\|_F/\|\mathbf{R}\|_F\le\varepsilon$ and $\|\mathbf{Z}^+-\mathbf{Z}\|_F/\|\mathbf{Z}\|_F\le\varepsilon$.

end while

Output: $\mathbf{R}$.

4 Experiments

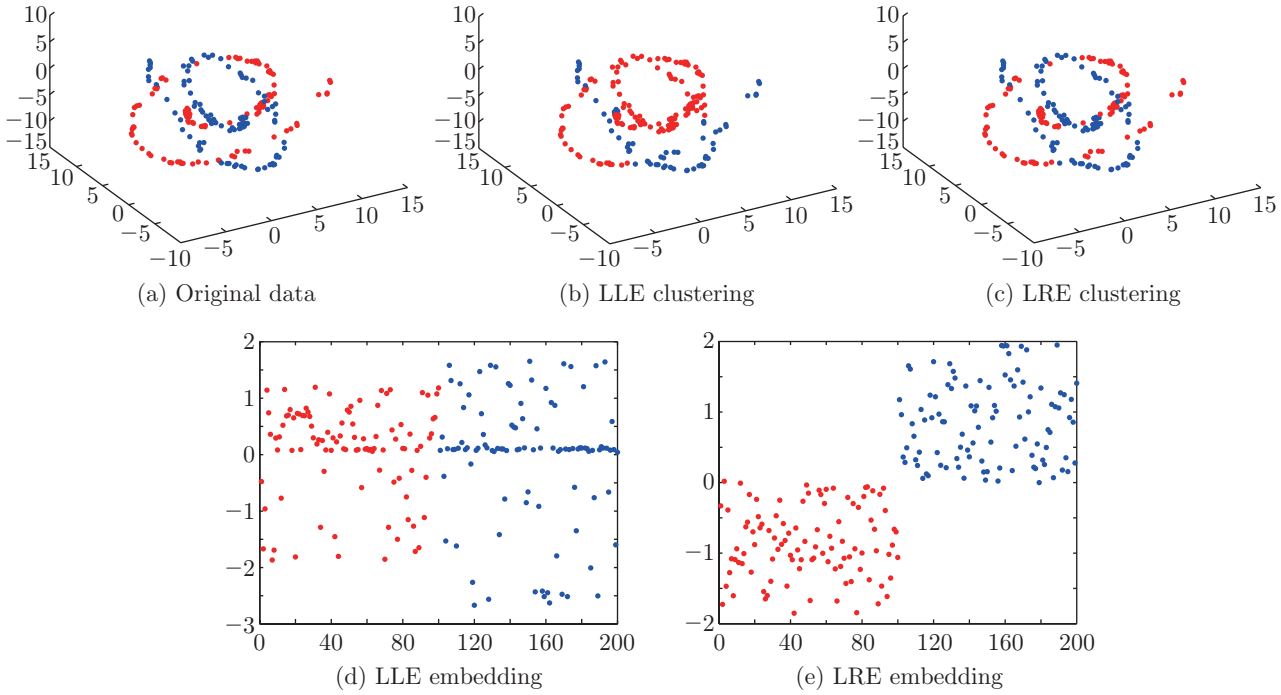
In this section, we report experimental results that validate our proposed algorithm. We first compare the performance of LRE and LLE for mixture of manifolds embedding and clustering. Then real-world motion segmentation experiments are done on the Hopkins155 database [17].

4.1 Mixture of Manifolds Embedding and Clustering

The following experiments mainly compare the capability of LRE and LLE in transferring separable structure of mixture of manifolds from the high-dimensional sample space to the low-dimensional feature space for clustering.

The first data we use is two mixed Swiss Rolls contaminated by randomly generated Gaussian noise ($\sigma = 0.1$), as shown in Fig. 3(a). Swiss Roll is widely used for experiments in manifold learning and it has been demonstrated that LLE can successfully recover the nonlinear structure of single Swiss Roll. However, in this subsection we will test the performance of LLE and LRE for clustering mixed Swiss Rolls.

As shown in Fig. 3(b), LLE can not separate the two Swiss Rolls and the clustering accuracy is only 60.0%. However, in Fig. 3(c), LRE can identify the two Swiss Rolls well and the clustering accuracy is dramaturgically higher (99.0%). We can clearly see in the last two subfigures in Fig. 3 that LRE embedding yields two groups of points that are easily identifiable while LLE embedding results in points that do not have clear clusters.

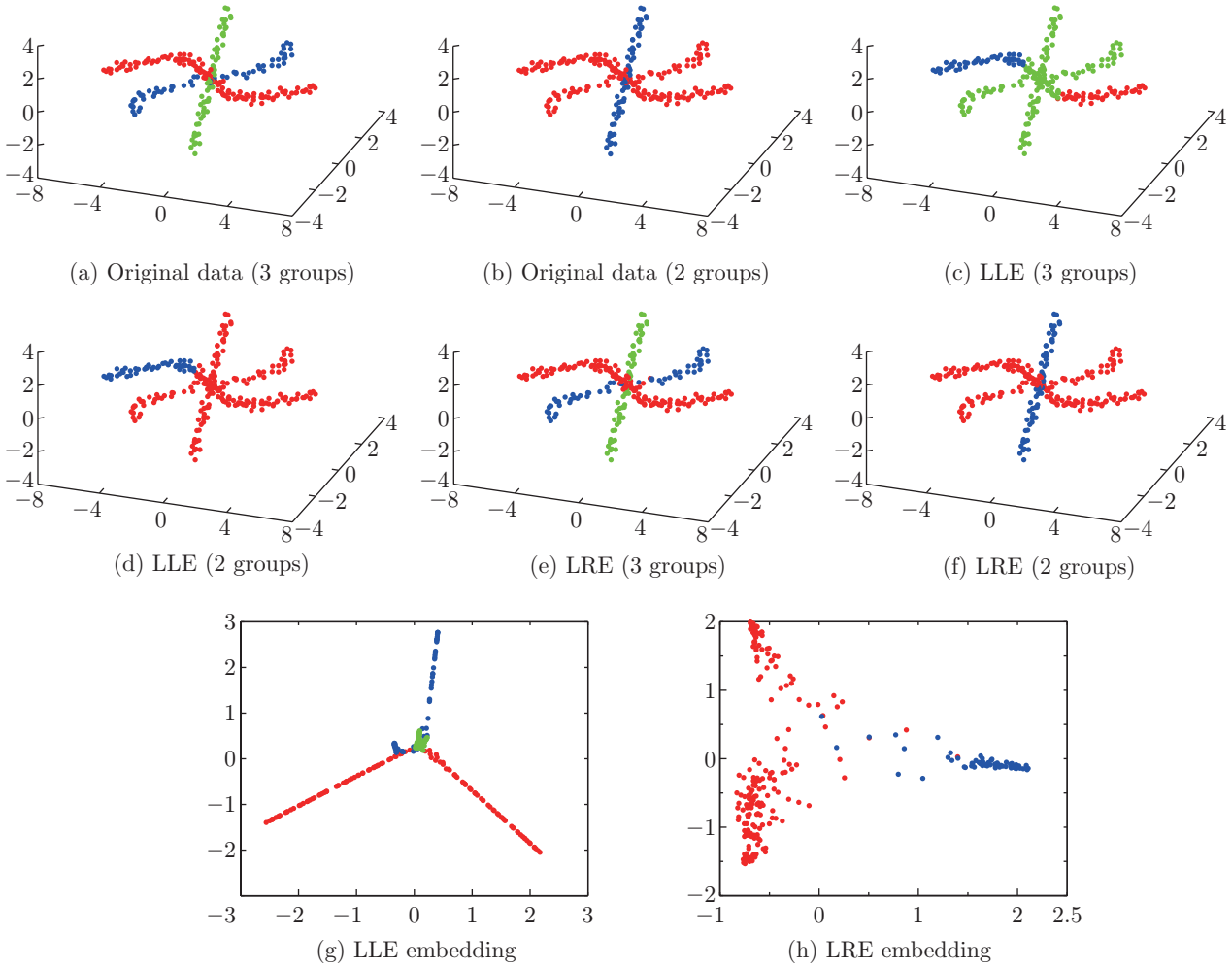


(a) Original data (b) LLE clustering (c) LRE clustering

(d) LLE embedding (e) LRE embedding

Fig. 3: Data sampled from two mixed Swiss Rolls (100 points each) contaminated with Gaussian noise ($\sigma = 0.1$). (a) The original data with true labels. (b) The clustering result based on LLE embedding ($K = 16$), whose clustering accuracy is 60.0%. (c) The clustering result based on LRE embedding ($\lambda = 1$), whose clustering accuracy is 99.0%. (d) and (e) are 1-dimensional embeddings of LLE and LRE, respectively. The horizontal axes are the indices of the points and the colors of the points are the true labels.

Another interesting mixture of manifolds example is one line mixed with two sine curves, where the two sine curves are on the same plane. Gaussian noise ($\sigma = 0.25$) is also added to this data. Actually, this data set can be clustered in two ways: one line and two sine curves (Fig. 4(a)) or one line and one plane containing the two sine curves (Fig. 4(b)). We can see that our LRE outperforms LLE in both cases. The embedding of LLE and LRE are also shown in Fig. 4. We can see that the embedding of LRE can be easily and correctly clustered, while that of LLE cannot.



(a) Original data (3 groups) (b) Original data (2 groups) (c) LLE (3 groups)

(d) LLE (2 groups) (e) LRE (3 groups) (f) LRE (2 groups)

(g) LLE embedding (h) LRE embedding

Fig. 4: Data sampled from mixed line and sine curves with Gaussian noise ($\sigma = 0.25$). This data set can be separated into either three groups: one line (63 points) and two curves (126 and 63 points, respectively) or two groups: one line and two curves. (a) and (b) are the original data with three-group labels and two-group labels, respectively. (c) and (d) are the clustering results of LLE with $K = 16$, whose clustering accuracy is 61.1% (three groups) and 75% (two groups), respectively. (e) and (f) are the clustering results of LRE with $\lambda = 1$, whose clustering accuracy are 96.8% (three groups) and 98.0% (two groups), respectively. (g) and (h) are the two-dimensional embeddings of LLE and LRE, respectively. The colors of the points are the true labels.

4.2 Motion Segmentation

The methods chosen for comparison in this experiment can be divided into four categories: (a) algebraic method: GPCA, (b) statistical method: RANSAC, (c) spectral clustering based method: Local Subspace Analysis (LSA) [18], SSC and LRR, and (d) embedding based method: LLMC and our proposed LRE. The Hopkins155 database consists of 155 sequences each of which has 39 to 550 data vectors drawn from two or three motions. To test the performance of our approach for real-world problems, we add Gaussian noise ($\sigma = 0.03$) to all the motion data. For computational efficiency, a PCA preprocessing step is used to project the data to be 5-dimensional. As shown in Table 1, both spectral clustering based methods (LSA, SSC and LRR) and embedding based methods (LLMC and LRE) perform better than algebraic (GPCA) and statistical (RANSAC)

methods. Furthermore, SSC and LRR also achieve better results than the traditional embedding based method (LLMC) in most cases (LLMC only beats SSC in clustering three motions). Finally, our proposed LRE outperforms all these state-of-the-art algorithms.

Table 1: Clustering errors (mean ± std-dev %) of several algorithms on the Hopkins155 motion segmentation database. "S" or "E" in the brackets means that the methods are based on spectral clustering or embedding, respectively.

	Two	Three	All
GPCA	10.8 ± 13.2	33.5 ± 11.8	16.0 ± 16.0
RANSAC	11.6 ± 14.8	29.3 ± 15.1	15.6 ± 16.6
LSA (S)	10.2 ± 12.6	26.1 ± 14.9	13.8 ± 14.7
SSC (S)	9.9 ± 14.7	19.2 ± 16.4	12.0 ± 15.6
LRR (S)	5.0 ± 9.2	15.9 ± 14.1	7.5 ± 11.4
LLMC (E)	11.7 ± 16.3	17.9 ± 15.2	13.1 ± 16.3
LRE (E)	**3.4 ± 7.1**	**10.9 ± 12.1**	**5.1 ± 9.0**

5 Conclusions

In this paper, we introduce a novel rank based reconstruction criterion for mixture of manifolds embedding and clustering. Both theoretical and experimental results demonstrate that our proposed method, LRE, can successfully transfer separable structure of mixture of manifolds from the high-dimensional sample space to the low-dimensional feature space.

Acknowledgement

This work is partially supported by the grants of the NSFC-Guangdong Joint Fund (No. U0935004), the NSFC Fund (No. 60873181) and the Fundamental Research Funds for the Central Universities. We would like to thank Zhouchen Lin, Feng Zhou and Fernando De la Torre for valuable discussions. The first author would also like to thank the support from China Scholarship Council.

Appendix

The proof of Theorem 1 is based on the following three lemmas[3]. Without loss of generality, we assume that all the matrices are square (namely, $m = d = n$) for notational simplicity.

Lemma 1 *Let* $\mathbf{A}$ *and* $\mathbf{B}$ *be matrices of the same dimensions. Then there exist matrices* $\mathbf{B}_1$ *and*

[3]See [14] for the proofs of these Lemmas.

$\mathbf{B}_2$ *such that*

$$\begin{cases} \mathbf{B} = \mathbf{B}_1 + \mathbf{B}_2, \\ rank(\mathbf{B}_1) \leq 2rank(\mathbf{A}), \\ \mathbf{A}\mathbf{B}_2^T = 0,\ \mathbf{A}^T\mathbf{B}_2 = 0, \\ \|\mathbf{A} + \mathbf{B}_2\|_* = \|\mathbf{A}\|_* + \|\mathbf{B}_2\|_* \\ \langle \mathbf{B}_1, \mathbf{B}_2 \rangle = 0. \end{cases}$$

Lemma 2 *Suppose that the reconstruction error* $\mathbf{E}$ *is a Gaussian distributed matrix with i.i.d.* $\mathcal{N}(0, \sigma^2)$ *entries. If* $c_0 > 4\sqrt{(1+\delta_1)\log 12}$*, we have*

$$\|\mathbf{X}^T\mathbf{E}\| \leq c_0\sqrt{m}\sigma$$

with a probability at least $1 - 2e^{-cm}$*, where* c *is a fixed numerical constant and* δ_1 *is the isometry constant of* A *at rank* 1*.*

Lemma 3 *If* $\hat{\mathbf{R}}$ *is the solution to (6) and* $\|\mathbf{X}^T\mathbf{E}\| \leq \lambda/2$ *holds, then we have*

$$\begin{cases} \|\mathbf{X}^T(\mathbf{X} - \mathbf{X}\hat{\mathbf{R}})\| \leq \lambda, \\ \|\hat{\mathbf{R}}\|_* \leq c_0\|\hat{\mathbf{R}} - \mathbf{R}\|_* + \|\mathbf{R}\|_*, \end{cases}$$

where c_0 *is a small constant.*

PROOF. (of Theorem 1) We split our proof into four steps.

Step 1: Set $\mathbf{H} = \hat{\mathbf{R}} - \mathbf{R}$. By the triangle inequality we have

$$\|\mathbf{X}^T\mathbf{X}\mathbf{H}\| \leq \|\mathbf{X}^T(\mathbf{X}\hat{\mathbf{R}} - \mathbf{X})\| + \|\mathbf{X}^T(\mathbf{X} - \mathbf{X}\mathbf{R})\| \leq 3\lambda/2, \tag{15}$$

since the first inequality in Lemma 3 holds.

Step 2: Decompose $\mathbf{H} = \mathbf{H}_0 + \mathbf{H}_c$, where $\mathrm{rank}(\mathbf{H}_0) \leq 2r$, $\mathbf{R}\mathbf{H}_c^T = 0$ and $\mathbf{R}^T\mathbf{H}_c = 0$ (see Lemma 1). We have $\|\mathbf{R} + \mathbf{H}\|_* \geq \|\mathbf{R}\|_* + \|\mathbf{H}_c\|_* - \|\mathbf{H}_0\|_*$. Since the second inequality in Lemma 3, $\|\mathbf{R} + \mathbf{H}\|_* = \|\hat{\mathbf{R}}\|_* \leq c_0\|\mathbf{H}\|_* + \|\mathbf{R}\|_*$, this gives $(1 - c_0)\|\mathbf{H}_c\|_* \leq (1 + c_0)\|\mathbf{H}_0\|_*$. Let $\mathbf{H}_c = \mathbf{U}\mathrm{diag}(\sigma)\mathbf{V}^T$ be the SVD of $\mathbf{H}_c$. Decompose $\mathbf{H}_c$ into a sum of matrices $\mathbf{H}_1, \mathbf{H}_2, ...,$, each whose rank is at most $2r$, as follows. For each j define the index set $I_j = \{2r(j-1)+1, ..., 2rj\}$, and let $\mathbf{H}_j := \mathbf{U}_{I_j}\mathrm{diag}(\sigma_{I_j})\mathbf{V}_{I_j}^T$; that is, $\mathbf{H}_1$ is the part of $\mathbf{H}_c$ corresponding to the $2r$ largest singular values, $\mathbf{H}_2$ is the part corresponding to the next $2r$ largest, and so on. By such construction, we have

$$\sum_{j \geq 2} \|\mathbf{H}_j\|_F \leq \frac{1 + c_0}{1 - c_0}\|\mathbf{H}_0\|_F. \tag{16}$$

Step 3: The RIP of $\mathbf{X}$ (8) gives $(1 - \delta_{4r})\|\mathbf{H}_0 + \mathbf{H}_1\|_F^2 \leq \|\mathbf{X}(\mathbf{H}_0 + \mathbf{H}_1)\|_F^2$ and observe that $\|\mathbf{X}(\mathbf{H}_0 + \mathbf{H}_1)\|_F^2 = \langle \mathbf{X}(\mathbf{H}_0 + \mathbf{H}_1), \mathbf{X}(\mathbf{H} - \sum_{j \geq 2}\mathbf{H}_j)\rangle$. It is easy to check that $\langle \mathbf{X}(\mathbf{H}_0 + \mathbf{H}_1), \mathbf{X}\mathbf{H}\rangle \leq \|\mathbf{H}_0 + \mathbf{H}_1\|_F\sqrt{4r}\|\mathbf{X}^T\mathbf{X}\mathbf{H}\|$ and $\langle \mathbf{X}(\mathbf{H}_0 + \mathbf{H}_1), \mathbf{X}\mathbf{H}_j\rangle \leq \sqrt{2}\delta_{4r}\|\mathbf{H}_0 + \mathbf{H}_1\|_F\|\mathbf{H}_j\|_F$. So we have

$$\begin{aligned} &(1 - \delta_{4r})\|\mathbf{H}_0 + \mathbf{H}_1\|_F \\ &\leq \sqrt{4r}\|\mathbf{X}^T\mathbf{X}\mathbf{H}\| + \tfrac{\sqrt{2}\delta_{4r}(1+c_0)}{1-c_0}\|\mathbf{H}_0 + \mathbf{H}_1\|_F. \end{aligned} \tag{17}$$

Step 4: To conclude, we have that $\|\mathbf{H}_0 + \mathbf{H}_1\|_F \leq c_1\sqrt{4r}\|\mathbf{X}^T\mathbf{X}\mathbf{H}\|$, where $c_1 = \frac{1}{1-(1+\sqrt{2}\frac{1+c_0}{1-c_0})\delta_{4r}}$, provided that $c_1 > 0$[4]. Our claim (9) then follows from (15) together with $\|\mathbf{H}\|_F \leq \frac{2}{1-c_0}\|\mathbf{H}_0 + \mathbf{H}_1\|_F$. The claim (10) follows from (9) simply by plugging $\lambda/2 = 8m\sigma$ into Lemma 2.

References

[1] J. Tenenbaum, V. Silva, J. Langford, A global geometric framework for nonlinear dimensionality reduction, Science, 290 (5500), 2000, 2319-2323

[2] S. Roweis, L. Saul, Nonlinear dimensionality reduction by locally linear embedding, Science 290 (5500), 2000, 2323-2326

[3] R. Vidal, Y. Ma, S. Sastry, Generalized principal component analysis (GPCA), IEEE Trans. on PAMI 27 (12), 2005, 1945-1959

[4] M. Tipping, C. Bishop, Mixtures of probabilistic principal component analyzers, Neural Computation 11 (2), 1999, 443-482

[5] M. Fischler, R. Bolles, Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography, Communications of the ACM, 26 (1981), 381-395

[6] U. Luxburg, A tutorial on spectral clustering, Statistics and Computing 17 (4), 2007, 395-416

[7] J. Costeira, T. Kanade, A multiply factorization method for independently moving objects, International Journal of Computer Vision 29 (3) (1998), 159-179

[8] E. Elhamifar, R. Vidal, Sparse subspace clustering, in: CVPR, 2009

[9] G. Liu, Z. Lin, Y. Yu, Robust subspace segmentation by low-rank representation, in: ICML, 2010

[10] R. Souvenir, R. Pless, Manifold clustering, in: ICCV, 2005

[11] M. Polito, P. Perona, Grouping and dimensionality reduction by locally linear embedding, in: NIPS, 2001

[12] A. Goh, R. Vidal, Segmenting motions of different types by unsupervised manifold clustering, in: CVPR, 2007

[13] E. Candès, B. Recht, Exact matrix completion via convex optimization, Foundations of Computational Mathematics

[14] E. Candès, Y. Plan, Tight oracle bounds for low-rank matrix recovery from a minimal number of random measurements, IEEE Trans. on Information Theory

[15] Z. Lin, M. Chen, L. Wu, Y. Ma, The augmented Lagrange multiplier method for exact recovery of corrupted low-rank matrices, UIUC Technical Report UILU-ENG-09-2215

[16] J. Cai, E. Candès, Z. Shen, A singular value thresholding algorithm for matrix completion, SIAM J. on Optimization 20 (4), 2008, 1956-1982

[17] R. Tron, R. Vidal, A benchmark for the comparison of 3D motion segmentation algorithms, in: CVPR, 2007

[18] J. Yan, M. Pollefeys, A general framework for montion segmentation: Independent, articulated, rigid, non-rigid, degenerate and nondegenerate, in: ECCV, 2006

[4] In particular, we suppose $c_0 = 1/2$ [14]