Manifold Alignment Based on Sparse Local Structures of More Corresponding Pairs

Xiaojie Li, Jian Cheng Lv, Zhang Yi
Machine Intelligence Laboratory, College of Computer Science, Sichuan University,
Chengdu 610065, P. R. China
lvjiancheng@scu.edu.cn

Abstract

Manifold alignment is to extract the shared latent semantic structure from multiple manifolds. The joint adjacency matrix plays a key role in manifold alignment. To construct the matrix, it is crucial to get more corresponding pairs. This paper proposes an approach to obtain more and reliable corresponding pairs in terms of local structure correspondence. The sparse reconstruction weight matrix of each manifold is established to preserve the local geometry of the original data set. The sparse correspondence matrices are constructed using the sparse local structures of corresponding pairs across manifolds. Further more, a new energy function for manifold alignment is proposed to simultaneously match the corresponding instances and preserve the local geometry of each manifold. The shared low dimensional embedding, which provides better descriptions for the intrinsic geometry and relations between different manifolds, can be obtained by solving the optimization problem with closed-form solution. Experiments demonstrate the effectiveness of the proposed algorithm.

1 Introduction

Manifold alignment is very useful in many fields, such as machine learning and data mining [Wang and Mahadevan, 2008; 2009a; 2009b; Zhai *et al.*, 2010; Pei *et al.*, 2012]. It is simultaneously a solution to the problem of alignment and a framework for discovering a unifying representation of multiple datasets [Wang *et al.*, 2011]. The problem of aligning multiple data sets is to extract the shared latent semantic structure. Due to different data sets might be represented by different features, it is difficult to find the correspondence in their original high-dimensional space. Thus, extraction of the shared intrinsic structure from multiple high-dimensional data sets becomes more critical.

In order to extract the intrinsic structure from a given data set, many linear and nonlinear dimensionality reduction approaches have been proposed in the literature such as SMCE [Elhamifar and Vidal, 2011], PCA [Jolliffe, 2002; Lv *et al.*, 2010a; 2010b], MDS [Shepard, 1980], ISOMAP [Tenenbaum *et al.*, 2000], LE [Belkin and Niyogi, 2003], LLE [Roweis and Saul, 2000]. However, this paper focuses on extracting the shared latent semantic structure from multiple manifolds, i.e., manifold alignment.

The current manifold alignment can be classified into supervised and unsupervised algorithms. The supervised algorithms require a small set of corresponding pairs for initial alignments, such as Ham's algorithm [Ham *et al.*, 2004] and Lafon's algorithm [Lafon *et al.*, 2006]. Wang et al. [Wang *et al.*, 2011] proposed an important framework for unsupervised alignment. Based on Wang's framework, there is a growing body of work on unsupervised alignment, such as [Hoa *et al.*, 2012; Pei *et al.*, 2012]. Unfortunately, the adjacency matrix they constructed does not correctly reflect the corresponding pairs across manifolds so that the local structure of shared latent manifold can not be preserved well.

In this paper, to preserve the local geometry of the original data set, the sparse reconstruction weight matrix of each manifold is established by SMCE method. To get more and reliable corresponding pairs, we propose a new method based on the assumption: local structures of a corresponding pair should be corresponding. The sparse correspondence matrices, which play a key role in joining the different manifolds, are constructed by the sparse local structures of corresponding pairs. Clearly, the alignment accuracy can be improved by providing more and reliable corresponding pairs. The adjacency matrix we constructed can effectively preserve the shared intrinsic structure from multiple data sets. Further more, a new energy function for manifold alignment is proposed to simultaneously match the corresponding instances and preserve the local geometry of each manifold. The shared low dimensional embedding, which provides better descriptions for the intrinsic geometry and relations between different manifolds, can be obtained by solving the optimization problem with closed-form solution. Extensive experiments have been performed to demonstrate the superior performance of proposed approach.

The rest of this paper is organized as follows. In Section 2, the preliminaries are presented. Section 3 shows how to construct the joint adjacency matrix. The proposed algorithm is given in Section 4. Simulations and discussions are given in Section 5. Finally, Section 6 provides some concluding remarks.

2 Preliminaries

2.1 Mathematical Formulation

Manifold alignment is to extract the shared latent semantic structure from multiple data sets. Suppose two data sets $\mathbf{X}^{(1)}$ and $\mathbf{X}^{(2)}$ parameterized by the same intrinsic structure:

$$\mathbf{X}^{(1)} = [\mathbf{x}_1^{(1)}, \mathbf{x}_2^{(1)}, \cdots, \mathbf{x}_m^{(1)}] \in R^{p^{(1)} \times m},$$
$$\mathbf{X}^{(2)} = [\mathbf{x}_1^{(2)}, \mathbf{x}_2^{(2)}, \cdots, \mathbf{x}_n^{(2)}] \in R^{p^{(2)} \times n},$$

where $\mathbf{X}^{(1)}$ is a $p^{(1)} \times m$ data matrix with m samples and $p^{(1)}$ features or dimensional spaces ($p^{(1)} \gg 1$); $\mathbf{X}^{(2)}$ is a $p^{(2)} \times n$ data matrix with n samples and $p^{(2)}$ features or dimensional spaces ($p^{(2)} \gg 1$).

Denote $\mathbf{W}$ as the joint adjacency matrix

$$\mathbf{W} = \begin{pmatrix} \mathbf{W}^{(1)} & \mathbf{W}^{(1,2)} \\ \mathbf{W}^{(2,1)} & \mathbf{W}^{(2)} \end{pmatrix},$$

where $\mathbf{W}^{(1)} \in R^{(m \times m)}, \mathbf{W}^{(2)} \in R^{(n \times n)}, \mathbf{W}^{(1,2)} \in R^{(m \times n)}$, and $\mathbf{W}^{(2,1)} \in R^{(n \times m)}$. It includes both parts: the reconstruction weight matrix of each manifold $\mathbf{W}^{(1)}, \mathbf{W}^{(2)}$ and the correspondence matrix $\mathbf{W}^{(1,2)}, \mathbf{W}^{(2,1)}$, which play a key role in aligning both data sets $\mathbf{X}^{(1)}$ and $\mathbf{X}^{(2)}$. Generally, $\mathbf{W}^{(1)}$ and $\mathbf{W}^{(2)}$ are determined according to the neighborhood of each data point [Wang *et al.*, 2011; Hoa *et al.*, 2012]. They are as follows:

$$\begin{cases} W^{(1)}_{(i,j)} \neq 0 & \text{if } \mathbf{x}_i^{(1)} \text{ and } \mathbf{x}_j^{(1)} \text{ data points are} \\ & \text{neighbors } (i, j = 1, 2, \cdots, m; i \neq j), \\ W^{(1)}_{(i,j)} = 0 & \text{otherwise.} \end{cases}$$

$$\begin{cases} W^{(2)}_{(i,j)} \neq 0 & \text{if } \mathbf{x}_i^{(2)} \text{ and } \mathbf{x}_j^{(2)} \text{ data points are} \\ & \text{neighbors } (i, j = 1, 2, \cdots, n; i \neq j), \\ W^{(2)}_{(i,j)} = 0 & \text{otherwise.} \end{cases}$$

$\mathbf{W}^{(1,2)}$ and $\mathbf{W}^{(2,1)}$ are constructed in terms of the corresponding pairs. A corresponding pair, denoted by $(\mathbf{x}_i^{(1)} \leftrightarrow \mathbf{x}_j^{(2)})$, means $\mathbf{x}_i^{(1)}$ and $\mathbf{x}_j^{(2)}$ have shared the same latent features. Some different methods are proposed to find the corresponding pairs.

Then, the shared intrinsic structure $\mathbf{Y}$ can be calculated by solving the following problem:

$$C(\mathbf{Y}) = \sum_{i,j=1}^{m+n} \|\mathbf{y}_i - \mathbf{y}_j\|_2^2 W_{(i,j)}.$$

The new embedded coordinates of both datasets $\mathbf{X}^{(1)}$ and $\mathbf{X}^{(2)}$ are given by $\mathbf{Y}^{(1)}$ and $\mathbf{Y}^{(2)}$, respectively.

$$\begin{pmatrix} \mathbf{Y}^{(1)} \\ \mathbf{Y}^{(2)} \end{pmatrix} = \mathbf{Y} \in R^{(m+n) \times d},$$

where d indicates the latent dimensionality and the mutual embedding space is spanned by the range of $\mathbf{Y}$.

Clearly, if $(\mathbf{x}_i^{(1)} \leftrightarrow \mathbf{x}_j^{(2)})$, then manifold alignment should find a map such that $\mathbf{y}_i^{(1)}$ and $\mathbf{y}_j^{(2)}$ is closest to each other in terms of Euclidean distance and the local structures of both $\mathbf{x}_i^{(1)}$ and $\mathbf{x}_j^{(2)}$ are also preserved. $\mathbf{y}_i^{(1)} = \mathbf{y}_j^{(2)}$ if and only if the instances $\mathbf{x}_i^{(1)}$ and $\mathbf{x}_j^{(2)}$ are in exact correspondence [Wang *et al.*, 2011].

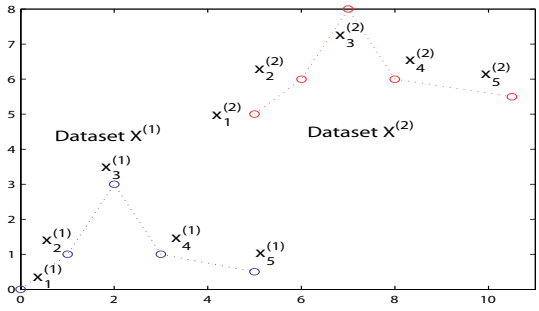

Figure 1: Two data sets

2.2 Correspondence Matrix

In unsupervised alignment, the most important step is to get some corresponding information across two datasets. Some structure similarity functions $f(\mathbf{x}_i^{(1)}, \mathbf{x}_j^{(2)})$ have been proposed to compute the similarity between $\mathbf{x}_i^{(1)} \in \mathbf{X}^{(1)}$ and $\mathbf{x}_j^{(2)} \in \mathbf{X}^{(2)}$, such as the parameterized curve representation method [Pei *et al.*, 2012] and the local pattern matching approach [Wang *et al.*, 2011]. Pei et al. introduced a parameterized curve representation to describe the local structure, where the points with similar parameterized curves are seen as corresponding pairs. Wang et al. employed a distance matrix of k nearest neighbors to characterize the local structure.

The method employed by Wang et al. has an important role in identifying the corresponding pairs. However , this method cannot preserve the local structures of corresponding pairs well. As shown in [Wang *et al.*, 2011], the supervised and unsupervised methods can be utilized for manifold alignment. The supervised algorithm (called WangS) is used to construct the correspondence matrices as:

$$\begin{cases} W^{(1,2)}_{(i,c)} = W^{(2,1)}_{(c,i)} = \eta & \text{if } (\mathbf{x}_i^{(1)} \leftrightarrow \mathbf{x}_c^{(2)}), \\ W^{(1,2)}_{(i,j)} = W^{(1,2)}_{(j,i)} = 0 & \text{otherwise,} \end{cases} \quad (1)$$

where $\eta > 0$. Clearly, $\mathbf{W}^{(1,2)}$ and $\mathbf{W}^{(2,1)}$ constructed by Eq.(1) may not represent the local structures of corresponding pairs.

In Wang's unsupervised algorithm (called WangU), $\mathbf{W}^{(1,2)}$ and $\mathbf{W}^{(2,1)}$ are constructed by:

$$W^{(1,2)}_{(i,j)} = W^{(2,1)}_{(j,i)} = f(\mathbf{x}_i^{(1)}, \mathbf{x}_j^{(2)}) \quad (2)$$
$$(i = 1, 2, \cdots, m;\ j = 1, 2, \cdots, n),$$

where $f(\mathbf{x}_i^{(1)}, \mathbf{x}_j^{(2)}) = e^{-dist(R_{\mathbf{x}_i^{(1)}}, R_{\mathbf{x}_j^{(2)}})/\theta^2}$, $R_{\mathbf{x}_i^{(1)}}$ is a distance matrix representing the local geometry of $\mathbf{x}_i^{(1)}$. However, a lager number of non-zero elements could violate the principles used to capture information about the corresponding structure [Elhamifar and Vidal, 2011]. Furthermore, the correspondence matrices didn't correctly represent the corresponding pairs across manifolds.

An example is as follows to illustrate the problem. In Figure 1, it is well known that the corresponding pairs should be

as follows:

$$(\mathbf{x}_1^{(1)} \leftrightarrow \mathbf{x}_1^{(2)}), (\mathbf{x}_2^{(1)} \leftrightarrow \mathbf{x}_2^{(2)}), (\mathbf{x}_3^{(1)} \leftrightarrow \mathbf{x}_3^{(2)}),$$
$$(\mathbf{x}_4^{(1)} \leftrightarrow \mathbf{x}_4^{(2)}), (\mathbf{x}_5^{(1)} \leftrightarrow \mathbf{x}_5^{(2)}).$$

However, by using WangU method,

$$\mathbf{W}^{(1,2)} = \begin{pmatrix} 2.6456 & 0 & 0 & 0 & 2.0428 \\ 0 & 2.6456 & 0 & 0.7276 & 0 \\ 0 & 0 & 2.6456 & 2.30777 & 0 \\ 0 & 2.2150 & 0 & 0.6619 & 0 \\ 2.3208 & 0 & 0.7450 & 0.7507 & 2.3086 \end{pmatrix}.$$

Due to $(W_{(1,1)}^{(1,2)} > W_{(1,j)}^{(1,2)}, W_{(2,2)}^{(1,2)} > W_{(2,j)}^{(1,2)}, W_{(3,3)}^{(1,2)} > W_{(3,j)}^{(1,2)})$, it means that

$$(\mathbf{x}_1^{(1)} \leftrightarrow \mathbf{x}_1^{(2)}), (\mathbf{x}_2^{(1)} \leftrightarrow \mathbf{x}_2^{(2)}), (\mathbf{x}_3^{(1)} \leftrightarrow \mathbf{x}_3^{(2)}).$$

Clearly, $W_{(4,4)}^{(1,2)} < W_{(4,2)}^{(1,2)}, W_{(5,5)}^{(1,2)} < W_{(5,1)}^{(1,2)}$, it means that

$$(\mathbf{x}_4^{(1)} \leftrightarrow \mathbf{x}_2^{(2)}), (\mathbf{x}_5^{(1)} \leftrightarrow \mathbf{x}_1^{(2)}).$$

Thus, $\mathbf{W}^{(1,2)}$ and $\mathbf{W}^{(2,1)}$ constructed directly by Eq.(2) cannot preserve the manifold structure very well. In the following section, an approach is proposed to overcome this problem.

3 Sparse Local structural Correspondence

3.1 Constructing the Sparse Reconstruction Weight Matrix of Each Data Set

In order to obtain the intrinsic structure of a given data set, dimensionality reduction methods play an important role in manifold learning, such as SMCE. It automatically finds a small neighborhood around each data point without fixing the neighborhood size k, and simultaneously obtains their weights. However, Sparse Manifold Clustering and Embedding (SMCE)[Elhamifar and Vidal, 2011] is used for clustering and dimensionality reduction not for manifold alignment. Inspired by SMCE, the sparse reconstruction weight matrix of each data set is constructed, i.e.,$\mathbf{W}^{(1)}$, $\mathbf{W}^{(2)}$, which characterizes the geometric property of the original data set.

3.2 Constructing the Sparse Correspondence Matrices between Data Sets

To construct the correspondence matrices $\mathbf{W}^{(1,2)}$, $\mathbf{W}^{(2,1)}$, it is crucial to get more corresponding pairs. In this section, we will introduce a new method to obtain more and reliable corresponding pairs in terms of local structure correspondence. Then, $\mathbf{W}^{(1,2)}$ and $\mathbf{W}^{(2,1)}$ are constructed using the sparse local structures of corresponding pairs across manifolds.

Given $\mathbf{x}_i^{(1)} (i = 1, 2, \cdots, m)$, a corresponding pair between two manifolds is chosen by

$$\{(\mathbf{x}_i^{(1)} \leftrightarrow \mathbf{x}_c^{(2)}) \mid c = \underset{j}{arg\,max}\ f(\mathbf{x}_i^{(1)}, \mathbf{x}_j^{(2)}), \\ and\ f(\mathbf{x}_i^{(1)}, \mathbf{x}_c^{(2)}) > \delta, j = 1, 2, \cdots, n\},$$

where δ is the threshold, which depends on the manifolds. In our experiments, let

$$\delta = min\{\mathcal{H}(\mathbf{W}^{(1,2)})\}, \tag{3}$$

where $\mathcal{H}(\mathbf{W}^{(1,2)})$ sorts the set

$$\{\text{the maximum of every row of } \mathbf{W}^{(1,2)}\}$$

and gets the top some values. A set of corresponding pairs S initialized to the empty set is obtained by

$$S = S \cup \{(\mathbf{x}_i^{(1)} \leftrightarrow \mathbf{x}_c^{(2)})\}, i = 1, 2, \cdots, m. \tag{4}$$

Clearly, S is just a part of all corresponding pairs. Based on the following assumption, a method is proposed to find the more and reliable corresponding pairs. Denote

$$N_{\mathbf{x}_i}^{(1)} = \{\mathbf{x}_{i,1}^{(1)}, \cdots, \mathbf{x}_{i,k}^{(1)}\},$$

where $N_{\mathbf{x}_i}^{(1)}$ denotes the k nearest neighbors of $\mathbf{x}_i^{(1)}$.

Assumption 1 *Suppose two data sets parameterized by the same intrinsic structure. If $(\mathbf{x}_i^{(1)} \leftrightarrow \mathbf{x}_c^{(2)})$, then the elements of $N_{\mathbf{x}_i}^{(1)}$ should be corresponding to that of $N_{\mathbf{x}_c}^{(2)}$ respectively, i.e, $(N_{\mathbf{x}_i}^{(1)} \leftrightarrow N_{\mathbf{x}_c}^{(2)})$. Conversely, if $(N_{\mathbf{x}_i}^{(1)} \leftrightarrow N_{\mathbf{x}_c}^{(2)})$, then $(\mathbf{x}_i^{(1)} \leftrightarrow \mathbf{x}_c^{(2)})$.*

Definition 1 *For sets A, B, denote the set $\{A \leftrightarrow B\}$ that the elements of A and B can be arbitrarily combined into the complete corresponding set.*

Definition 2 *$\mathbf{x}_i^{(1)} \nrightarrow S$ means $\mathbf{x}_i^{(1)}$ is not in any corresponding pair in S.*

Suppose

$$\mathbf{x}_l^{(1)} \nrightarrow S, l = 1, 2, \cdots, K_1,$$
$$\mathbf{x}_l^{(2)} \nrightarrow S, l = 1, 2, \cdots, K_2.$$

A corresponding pair can be found by

$$(\mathbf{x}_l^{(1)} \leftrightarrow \mathbf{x}_{lc}^{(2)}) \mid lc = \underset{j}{arg\,max}\ |g(N_{\mathbf{x}_l}^{(1)}, N_{\mathbf{x}_j}^{(2)})|, \\ for\ j = 1, 2, \cdots, K_2, \tag{5}$$

where

$$g(N_{\mathbf{x}_l}^{(1)}, N_{\mathbf{x}_j}^{(2)}) = \{N_{\mathbf{x}_l}^{(1)} \leftrightarrow N_{\mathbf{x}_j}^{(2)}\} \cap S.$$

$$S = \{S \cup (\mathbf{x}_l^{(1)} \leftrightarrow \mathbf{x}_{lc}^{(2)})\}.$$

Repeat this process until there is no reliable element added to S. Furthermore, the sparse correspondence matrices are constructed by

$$\mathbf{W}_{(i)}^{(1,2)} = \mathbf{W}_{(c)}^{(2)} \text{ and } \mathbf{W}_{(c)}^{(2,1)} = \mathbf{W}_{(i)}^{(1)}, \\ \text{if } (\mathbf{x}_i^{(1)}, \mathbf{x}_c^{(2)}) \in S. \tag{6}$$

where $\mathbf{W}^{(1,2)}$ and $\mathbf{W}^{(2,1)}$ are constructed by the sparse local structures of corresponding pairs, which play a key role in aligning different manifolds.

More and reliable corresponding pairs can be obtained by using our method. See the example in Figure 1. Let

$$\delta = min\{4.4284, 4.4284, 4.4284\} = 4.4284,$$

then by Eq.(4),

$$S = \{(\mathbf{x}_1^{(1)} \leftrightarrow \mathbf{x}_1^{(2)}), (\mathbf{x}_2^{(1)} \leftrightarrow \mathbf{x}_2^{(2)}), (\mathbf{x}_3^{(1)} \leftrightarrow \mathbf{x}_3^{(2)})\}.$$

By definition 2, $\{\mathbf{x}_4^{(1)}, \mathbf{x}_5^{(1)}, \mathbf{x}_4^{(2)}, \mathbf{x}_5^{(2)}\} \nrightarrow S$.
From section 3.1, it holds that

$$N_{\mathbf{x}_4}^{(1)} = \{\mathbf{x}_3^{(1)}, \mathbf{x}_5^{(1)}\}, N_{\mathbf{x}_4}^{(2)} = \{\mathbf{x}_3^{(2)}, \mathbf{x}_5^{(2)}\}.$$

By definition 1 $\{N^{(1)}_{\mathbf{x}^{(1)}_4} \leftrightarrow N^{(2)}_{\mathbf{x}^{(1)}_4}\} = \{(\mathbf{x}^{(1)}_3 \leftrightarrow \mathbf{x}^{(2)}_3), (\mathbf{x}^{(1)}_3 \leftrightarrow \mathbf{x}^{(2)}_5), (\mathbf{x}^{(1)}_5 \leftrightarrow \mathbf{x}^{(2)}_3), (\mathbf{x}^{(1)}_5 \leftrightarrow \mathbf{x}^{(2)}_5)\}$. Then,

$$\{N^{(1)}_{\mathbf{x}^{(1)}_4} \leftrightarrow N^{(2)}_{\mathbf{x}^{(1)}_4}\} \cap S = \{(\mathbf{x}^{(1)}_3 \leftrightarrow \mathbf{x}^{(2)}_3)\},$$

thus, $(\mathbf{x}^{(1)}_4 \leftrightarrow \mathbf{x}^{(2)}_4)$ can be obtained from the sparse local structure correspondence. Similarly, $(\mathbf{x}^{(1)}_5 \leftrightarrow \mathbf{x}^{(2)}_5)$ can be obtained. Then, construct $\mathbf{W}^{(1,2)}$ and $\mathbf{W}^{(2,1)}$ by Eq.(6).

Remark 1 *Obviously, $\mathbf{W}^{(2,1)}$ and $\mathbf{W}^{(1,2)}$ are sparse matrices, since $\mathbf{W}^{(1)}$ and $\mathbf{W}^{(2)}$ are sparse. The sparse matrix $\mathbf{W}$ characterizes geometric properties and reflects the joint manifold constructed from the two original manifolds and the corresponding pairs.*

Remark 2 *Based on simple geometric intuitions, nearby points in the high dimensional space remain nearby and similarly co-located with respect to one another in the low dimensional space. That means that the neighbors of $\mathbf{x}^{(1)}_i$ should be the neighbors of $\mathbf{x}^{(2)}_c$, and the neighbors of $\mathbf{x}^{(2)}_c$ should be the neighbors of $\mathbf{x}^{(1)}_i$ after alignment.*

In the following section, the SLSA algorithm will be used to learn low-dimensional embeddings based on the joint adjacency matrix $\mathbf{W}$.

4 Sparse Local Structural Alignment (SLSA) Algorithm

4.1 Loss Function

To extract the shared intrinsic structure from multiple manifolds in terms of the joint adjacency matrix constructed, a loss function $C(\mathbf{Y})$ is designed to simultaneously match the corresponding pairs and preserve the local structure of each manifold. Suppose a corresponding pair $(\mathbf{x}^{(1)}_i \leftrightarrow \mathbf{x}^{(2)}_c)$, $C(\mathbf{y}^{(1)}_i)$ includes two parts:

$$C(\mathbf{y}^{(1)}_i) \begin{cases} \|\mathbf{y}^{(1)}_i - \sum\limits_{\mathbf{y}^{(1)}_j \in N^{(1)}_{\mathbf{y}_i}} \mathbf{y}^{(1)}_j W^{(1)}_{(i,j)}\|^2_2, \\ \|\mathbf{y}^{(1)}_i - \mathbf{y}^{(2)}_c\|^2_2 = \|\mathbf{y}^{(1)}_i - \sum\limits_{\mathbf{y}^{(2)}_j \in N^{(2)}_{\mathbf{y}_c}} \mathbf{y}^{(2)}_j W^{(1,2)}_{(i,j)}\|^2_2, \end{cases} \tag{7}$$

where $N^{(1)}_{\mathbf{y}_i} = \{$ k nearest neighbors of $\mathbf{y}^{(1)}_i\}$, $N^{(2)}_{\mathbf{y}_c} = \{$ k nearest neighbors of $\mathbf{y}^{(2)}_c\}$. The first part is to compute the reconstruction residual with fixed weights $\mathbf{W}^{(1)}_{(i)}$, which is used to preserve the local geometric structure of each original data set [Roweis and Saul, 2000]. The second part is to compute the distance between $\mathbf{y}^{(1)}_i$ and $\mathbf{y}^{(2)}_c$. That means that $\mathbf{y}^{(1)}_i$ is close to $\mathbf{y}^{(2)}_c$ by minimizing this term, i.e., the instances $\mathbf{x}^{(1)}_i$ and $\mathbf{x}^{(2)}_c$ have been aligned in the mutual embedding space spanned by the range of $\mathbf{Y}$. Then, the loss function is as follows:

$$C(\mathbf{Y}) = \sum_{i=1}^{m+n} \|\mathbf{y}_i - \sum_{j=1}^{2k} \mathbf{y}_j W_{(i,j)}\|^2_2, \tag{8}$$

where the sum is taken over all instances from both data sets. $\mathbf{y}_i$ is reconstructed by its $2k$ joint nearest neighbors from two manifolds. The shared intrinsic structure $\mathbf{Y}$ can be obtained by minimizing the loss function $C(\mathbf{Y})$.

4.2 Optimization Solution

To optimize the embedding loss function $C(\mathbf{Y})$, we can rewrite it as the quadratic form:

$$\begin{aligned} C(\mathbf{Y}) &= \sum_i \|\mathbf{y}_{(i)} - \sum_{j=1} \mathbf{y}_{(j)} W_{(i,j)}\|^2_2 \\ &= \sum_i \|\mathbf{I}_{(i)}\mathbf{Y} - \mathbf{W}_{(i)}\mathbf{Y}\|^2_2 \\ &= \sum_i \|(\mathbf{I}_{(i)} - \mathbf{W}_{(i)})\mathbf{Y}\|^2_2 \\ &= \sum_i tr(\mathbf{Y}^T(\mathbf{I}_{(i)} - \mathbf{W}_{(i)})^T(\mathbf{I}_{(i)} - \mathbf{W}_{(i)})\mathbf{Y}) \\ &= tr(\mathbf{Y}^T(\mathbf{I} - \mathbf{W})^T(\mathbf{I} - \mathbf{W})\mathbf{Y}) \\ &= tr(\mathbf{Y}^T\mathbf{M}\mathbf{Y}), \end{aligned}$$

where $\mathbf{I}_{(i)}$ denotes the ith row of identity matrix and $\mathbf{I} \in R^{(m+n)\times(m+n)}$; $\mathbf{M} = (\mathbf{I} - \mathbf{W})^T(\mathbf{I} - \mathbf{W}) \in R^{(m+n)\times(m+n)}$ is sparse, symmetric and semipositive definite, $M_{(i,j)}$=$\beta_{(i,j)} - W_{(i,j)} - W_{(j,i)} + \sum_{2k} W_{(2k,i)}W_{(2k,j)}$ $(i, j = 1, 2, \cdots, m+n)$, and

$$\beta_{(i,j)} = \left\{ \begin{array}{ll} 1 & i = j \\ 0 & i \neq j \end{array} \right\}.$$

To avoid the trivial solution of minimizing $C(\mathbf{Y})$, the following constraint is required:

$$\mathbf{Y}^T\mathbf{D}\mathbf{Y} = \mathbf{I}, \tag{9}$$

where $\mathbf{I} \in R^{d\times d}$, $\mathbf{D} \in R^{(m+n)\times(m+n)}$ is a full rank diagonal matrix formed by $\mathbf{D}_{(i,i)} = \sum_{j=1}^{m+n} W_{(i,j)}$.

The Lagrange multiplier method is adopted to solve the optimization problem in Eq.(8):

$$\underset{\mathbf{y}}{arg\ min}\ L(\mathbf{Y}) = \underset{\mathbf{y}}{arg\ min}\ \{tr(\mathbf{Y}^T\mathbf{M}\mathbf{Y}) + \lambda(\mathbf{I} - \mathbf{Y}^T\mathbf{D}\mathbf{Y})\}. \tag{10}$$

Then,

$$\left\{ \begin{array}{l} \frac{\partial L(\mathbf{Y})}{\partial \mathbf{Y}} = 0 \\ \frac{\partial L(\mathbf{Y})}{\partial \lambda} = 0 \end{array} \right. \Rightarrow \left\{ \begin{array}{l} \mathbf{M}\mathbf{Y} = \lambda\mathbf{D}\mathbf{Y}, \\ \mathbf{I} = \mathbf{Y}^T\mathbf{D}\mathbf{Y}. \end{array} \right.$$

Clearly, it holds that $\mathbf{D^{-1}M}\mathbf{Y} = \lambda\mathbf{Y}$. The embedded coordinates $\mathbf{Y}$ are formed by eigenvectors of $\mathbf{D^{-1}M}$ corresponding to the d smallest non-zero eigenvalues.

4.3 The SLSA Algorithm

The whole SLSA method is summarized in Algorithm 1 and the following theorem will prove the algorithm is valid to simultaneously match the corresponding pairs and preserve the local structure of each manifold.

Theorem 1 *Given $\mathbf{x}^{(1)}_i (i = 1, 2, \cdots, m)$, if $(\mathbf{x}^{(1)}_i \leftrightarrow \mathbf{x}^{(2)}_c) \in S$, then there exists a small constant ϵ, such that $\|\mathbf{y}^{(1)}_i - \mathbf{y}^{(2)}_c\|_2 \leq 2\epsilon$.*

Proof: Since $(\mathbf{x}^{(1)}_i \leftrightarrow \mathbf{x}^{(2)}_c) \in S$, we have:

$$\begin{aligned} \mathbf{W}_{(i)} &= (\mathbf{W}^{(1)}_{(i)}\ \mathbf{W}^{(1,2)}_{(i)}), \\ \mathbf{W}_{(c)} &= (\mathbf{W}^{(2,1)}_{(c)}\ \mathbf{W}^{(2)}_{(c)}), \end{aligned}$$

where $\mathbf{W}_{(i)}$ denotes the ith row of $\mathbf{W}$. By Eq.(6), it follows that:

$$\begin{aligned} \mathbf{W}^{(2,1)}_{(c)} &= \mathbf{W}^{(1)}_{(i)}, \\ \mathbf{W}^{(1,2)}_{(i)} &= \mathbf{W}^{(2)}_{(c)}. \end{aligned}$$

Algorithm 1 The SLSA Algorithm

Input: Two data sets $\mathbf{X}^{(1)} \in R^{p^{(1)} \times m}$, $\mathbf{X}^{(2)} \in R^{p^{(2)} \times n}$.
Output: The new unified embedded coordinates $\mathbf{Y}$.
Step 1: Initialize $\mathbf{W} \leftarrow \mathbf{0}$, $S \leftarrow \emptyset$.
Step 2: Inspired by SMCE, construct $\mathbf{W}^{(1)}$, $\mathbf{W}^{(2)}$.
Step 3: Compute the corresponding pairs set S.
  for each entry $W^{(1,2)}_{(i,j)}$ in $\mathbf{W}^{(1,2)}$ **do**
    if
    $c = \underset{1<=j<=k}{arg\ max}\, W^{(1,2)}_{(i,j)}$, and $W^{(1,2)}_{(i,c)} >= \delta$
    then $S = \{S \cup (\mathbf{x}^{(1)}_i \leftrightarrow \mathbf{x}^{(2)}_c)\}$;
    end if
  end for
  for each point $\{\mathbf{x}^{(1)}_l \in \mathbf{X}^{(1)}\} \nrightarrow S$
    if
    $lc = \underset{j}{arg\ max}\ |g(N^{(1)}_{\mathbf{x}_l}, N^{(2)}_{\mathbf{x}_j})|$ **then**
    $S = \{S \cup (\mathbf{x}^{(1)}_l \leftrightarrow \mathbf{x}^{(2)}_{lc})\}$;
    end if
  end for
Step 4: Based on S, construct the correspondence matrices $\mathbf{W}^{(2,1)}$ and $\mathbf{W}^{(1,2)}$ by Eq.(6).
Step 5: Compute the embedded coordinates by Eq.(10).

Then, by minimizing of Eq.(7), there exist two vectors ϵ_1 and ϵ_2, it follows that:

$$\begin{cases} \mathbf{y}^{(1)}_i = \sum_{l=1}^{k} W^{(1)}_{(i,l)} \mathbf{y}^{(1)}_l + \sum_{l=1}^{k} W^{(1,2)}_{(i,l)} \mathbf{y}^{(2)}_l + \epsilon_1 \\ \quad = \sum_{l=1}^{k} W^{(1)}_{(i,l)} \mathbf{y}^{(1)}_l + \sum_{l=1}^{k} W^{(2)}_{(c,l)} \mathbf{y}^{(2)}_l + \epsilon_1 \\ \quad = \sum_{l=1}^{2k} W_{(i,l)} \mathbf{y}_l + \epsilon_1, \\ \mathbf{y}^{(2)}_c = \sum_{l=1}^{k} W^{(2,1)}_{(c,l)} \mathbf{y}^{(1)}_l + \sum_{l=1}^{k} W^{(2)}_{(c,l)} \mathbf{y}^{(2)}_l + \epsilon_2 \\ \quad = \sum_{l=1}^{k} W^{(1)}_{(i,l)} \mathbf{y}^{(1)}_l + \sum_{l=1}^{k} W^{(2)}_{(c,l)} \mathbf{y}^{(2)}_l + \epsilon_2 \\ \quad = \sum_{l=1}^{2k} W_{(c,l)} \mathbf{y}_l + \epsilon_2, \end{cases}$$

where $\|\epsilon_1\|_2 \leq \epsilon$ and $\|\epsilon_2\|_2 \leq \epsilon$. Thus,

$$\|\mathbf{y}^{(1)}_i - \mathbf{y}^{(2)}_c\|_2 = \|\epsilon_1 - \epsilon_2\|_2 \leq 2\epsilon.$$

From Eq.(8), it is easy to get that $\mathbf{y}^{(1)}_i$ is much closer to $\mathbf{y}^{(2)}_c$ than other points in $\mathbf{Y}^{(2)}$. The proof is complete.

Furthermore, it is easy to see that if $(\mathbf{x}^{(1)}_i \leftrightarrow \mathbf{x}^{(2)}_c)$, then $\mathbf{y}^{(1)}_i$ should share the same nearest neighbors of $\mathbf{y}^{(2)}_c$ in $\mathbf{Y}^{(2)}$, and vice versa.

5 Simulations and Discussions

In this section, two databases are used in our experiments: the COIL-20 database [Samneer *et al.*, 1996] and the Glutaredoxin protein structure database PDB-1G7O [J. and Bystroff, 2008]. The performance of the proposed algorithm (SLSA) is compared with that of the related alignment algorithm, such as Wang's method: WangS and WangU [Wang *et al.*, 2011]. In our experiments, the different sets of corresponding pairs with different cardinalities are selected randomly. Under the ideal condition, the low-dimensional coordinates of corresponding pairs should be overlap [Wang *et al.*, 2011; Pei *et al.*, 2012].

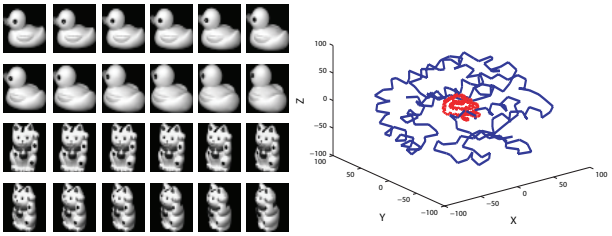

Figure 2: Sample images of duck and Maneki-neko from the COIL-20 database (left). The 1 and 21 protein models from the PDB-1G7O database (right).

5.1 Databases

The Columbia Object Image Library (COIL-20) is a gray-scale image database of 20 objects. With a fixed camera, the objects were rotated through 360 degrees and 72 images per object were taken at pose intervals of 5 degrees. In our experiments, each image is resized to 32×32 and becomes a point in 1024-dimensional space. Some images are shown in Figure 2 (left).

Proteins are flexible molecules, which may appear in different conformations. Due to that proteins are structurally similar in each family and a little similarity in different family, thus analysis of protein structure is expected, which can lead to a deep insight into protein function. More importantly, it has many applications in the studies of protein evolution and classification. We perform the alignment experiments with the 1 and 21 protein models. Each model has 215 amino acids and is represented in $3D$ spaces (Figure 2 (right)).

5.2 The Role of Alignment: ($\mathbf{W}^{(1)}$ and $\mathbf{W}^{(2)}$) or ($\mathbf{W}^{(1,2)}$ and $\mathbf{W}^{(2,1)}$)?

Set $\mathbf{W}^{(1,2)} = \mathbf{0}$ and $\mathbf{W}^{(2,1)} = \mathbf{0}$. Eq.(10) can be viewed as dimensionality reduction on two data sets respectively. In Figure 3, the left figure illustrates the alignment result by our method. the right figure illustrates the alignment result by Wang's method. The alignment accuracies of both methods are equal to zero. Therefore, a conclusion is arrived at that the correspondence matrices $\mathbf{W}^{(1,2)}$ and $\mathbf{W}^{(2,1)}$ play an important role in manifold alignment.

5.3 Experiments with Prior Correspondence

Based on prior corresponding pairs, WangS constructs $\mathbf{W}^{(1,2)}$ and $\mathbf{W}^{(2,1)}$ by Eq.(1), our method by Eq.(6). Under the different sets of corresponding pairs with different cardinalities, Figure 4 and Figure 5 illustrate the best and average alignment accuracies of both methods on two databases respectively. It can be seen that SLSA has better performance

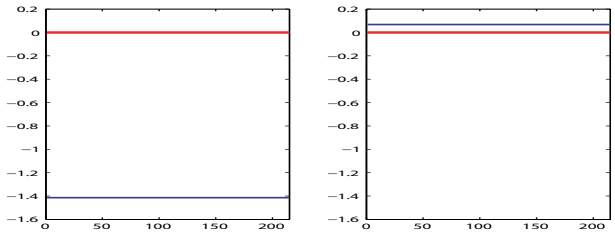

Figure 3: The $1D$ alignment results of the protein models by SLSA and Wang's method without corresponding pairs.

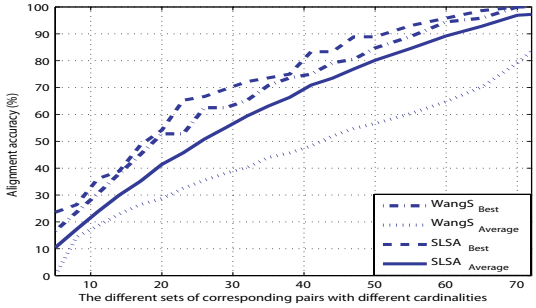

Figure 4: The alignment accuracies of SLSA and WangS methods on COIL-20 databases with $d = 2$ and $k = 3$.

than that of WangS method. Under the same conditions of prior corresponding pairs, sparse reconstruction weight matrices and loss function, different correspondence matrices constructed have different alignment results. Figure 6 illustrates the alignment results of the proposed algorithm with different correspondence matrices constructed by Eq.(1) and Eq.(6). It can be see that the manifold structure is not preserved well in the left Figure.

5.4 Experiments without Prior Correspondence

When there is no prior corresponding pairs, we need to use the function f to find them. WangU is done by local pattern matching approach (LPM) to find correspondence information [Wang *et al.*, 2011]. They construct $\mathbf{W}^{(1,2)}$ and $\mathbf{W}^{(2,1)}$ by Eq.(2). Based on the set of corresponding pairs found,

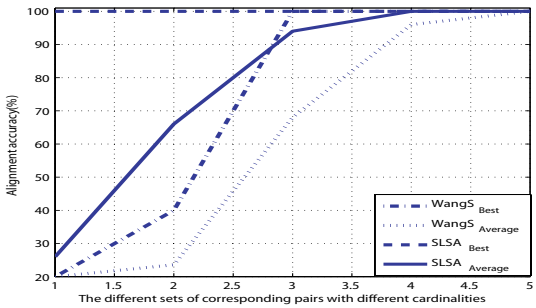

Figure 5: The alignment accuracies of SLSA and WangS methods on the synthetic datasets in Figure 1 with $d = 1$ and $k = 2$.

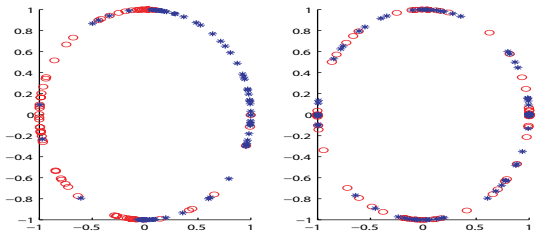

Figure 6: The $2D$ alignment results of the proposed algorithm with correspondence matrices constructed by Eq.(1) (left) and Eq.(6) (right) on the COIL-20 database.

Table 1: Alignment accuracy (%) with $k = 2$

Method	COIL-20	Synthetic Datasets
#Corresponding Pairs	1	3
WangU	1.39	80
SLSA	1.39	100

some more corresponding pairs can be got by section 3.2. As shown in Table 1, LPM with 2 nearest neighbors to characterize local geometry just found only one pairwise correspondence in COIL-20 database. No more corresponding pairs can be found. However, in the synthetic datasets, Two more corresponding pairs can be found based on the three pairwise correspondences found. Our method construct the correspondence matrices by Eq.(6). It can be seen that SLSA has better performance than that of WangU method.

6 Conclusions

This paper proposed a sparse local structural alignment (SLSA) algorithm for manifold alignment. A sparse optimization problem is formulated with a closed-form solution which extract the shared intrinsic structure from manifolds. We proposed an approach to obtain more and reliable corresponding pairs in terms of local structure correspondence. Then, the joint adjacency matrix is constructed using the sparse local structures of corresponding pairs. Experimental results demonstrate our method can get better performance than other related alignment methods. Furthermore, our method could be easily extended to align multiple data sets.

Acknowledgments

This work was supported by Specialized Research Fund for the Doctoral Program of Higher Education under Grant 2010081110053 and Program for New Century Excellent Talents in University under Grant NCET-10-0568, partially supported by National Program on Key Basic Research Project (973 Program) under Grant 2011CB302201.

References

[Belkin and Niyogi, 2003] Mikhail Belkin and Partha Niyogi. Laplacian eigenmaps for dimensionality re-

duction and data representation. *Neural Comput.*, 15(6):1373–1396, Jun. 2003.

[Elhamifar and Vidal, 2011] Ehsan Elhamifar and Rene Vidal. Sparse manifold clustering and embedding. In *Advances in Neural Information Processing Systems 24*, pages 55–63. 2011.

[Ham *et al.*, 2004] Jihun Ham, Daniel D. Lee, and Lawrence K. Saul. Semisupervised alignment of manifolds. pages 120–127, 2004.

[Hoa *et al.*, 2012] Trong Vu Hoa, Clifton Carey, and Sridhar Mahadevan. Manifold warping: Manifold alignment over time. pages 1–7, 2012.

[J. and Bystroff, 2008] Zaki M. J. and Christopher Bystroff. *Protein Structure Prediction Second Edition*. Humana Press, Humana, 2008.

[Jolliffe, 2002] I. T Jolliffe. *Principal Components Analysis*. Springer Second Edition, Springer, 2002.

[Lafon *et al.*, 2006] Stephane Lafon, Yosi Keller, and Ronald R. Coifman. Data fusion and multicue data matching by diffusion maps. *IEEE tansactions on pattern analysis and machine intelligence*, 28(11):1784–1797, Nov. 2006.

[Lv *et al.*, 2010a] Jian Cheng Lv, Kok Kiong Tan, Zhang Yi, and Sunan Huang. A family of fuzzy learning algorithms for robust principal component analysis neural networks. *Fuzzy Systems, IEEE Transactions on*, 18(1):217 –226, Feb. 2010.

[Lv *et al.*, 2010b] Jian Cheng Lv, Zhang Yi, and Jiliu Zhou. *Subspace Learning of Neural Networks*. CRC press, CRC, 2010.

[Pei *et al.*, 2012] Yuru Pei, Fengchun Huang, Fuhao Shi, and Hongbin Zha. Unsupervised image matching based on manifold alignment. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 34(8):1658 –1664, Aug. 2012.

[Roweis and Saul, 2000] S.T. Roweis and L.K. Saul. Nonlinear dimensionality reduction by local linear embedding. *Science*, 290(5500):2323–2326, Dec. 2000.

[Samneer *et al.*, 1996] A. N. Samneer, Shree K. Nayar, and Hiroshi Murase. Coumbia object image library(coil-20). *Department of Computer Science, Columbia University*, pages 1–4, Dec. 1996.

[Shepard, 1980] R. N. Shepard. Multidimensional scaling, tree-fitting, and clustering. *Science, New Series*, 210(4468):390–398, Oct. 1980.

[Tenenbaum *et al.*, 2000] J.B Tenenbaum, V.de Silva, and J.C Langford. A global geometric framework for nonlinear dimensionality reduction. *Science*, 290(5500):2319–2323, Dec. 2000.

[Wang and Mahadevan, 2008] Chang Wang and Sridhar Mahadevan. Manifold alignment using procrustes analysis. In *International Conference on Machine Learning*, pages 1120–1127, 2008.

[Wang and Mahadevan, 2009a] Chang Wang and Sridhar Mahadevan. Manifold alignment without correspondence. In *Proceedings of the 21st international jont conference on Artifical intelligence*, pages 1273–1278, San Francisco, 2009. Morgan Kaufmann Publishers Inc.

[Wang and Mahadevan, 2009b] Chang Wang and Sridhar Mahadevan. Multiscale analysis of document corpora based on diffusion models. In *Proceedings of the 21st international jont conference on Artifical intelligence*, pages 1592–1597, San Francisco, 2009. Morgan Kaufmann Publishers Inc.

[Wang *et al.*, 2011] Chang Wang, Peter Krafft, and Sridhar Mahadevan. Manifold alignment. *IEEE Trans. on Siganl Processsing*, 51(3):839–852, 2011.

[Zhai *et al.*, 2010] Deming Zhai, Hong Chang, Li Bo, Shan Shiguang, Chen Xilin, and Gao Wen. Manifold alignment via corresponding projections. *BMVC*, pages 1–11, 2010.