Manifold Learning and Its Applications: Papers from the AAAI Fall Symposium (FS-10-06)

# Alignments of Manifold Sections of Different Dimensions in Manifold Learning

**Qiang Ye** and **Weifeng Zhi**
Department of Mathematics
University of Kentucky
Lexington, KY 40506, USA

### Abstract

We consider an alignment algorithm for reconstructing global coordinates from local coordinates constructed for sections of manifolds. We show that, under certain conditions, the alignment algorithm can successfully recover global coordinates even when local neighborhoods have different dimensions. Our results generalize an earlier analysis to allow alignment of sections of different dimensions. We also apply our result to a semisupervised learning problem.

## Introduction

Manifold-based nonlinear dimensionality reduction has attracted significant research interests in recent years. Mathematically, it can be described as follows. Consider a $d$-dimensional parameterized manifold $\mathcal{M}$ embedded in $\mathbb{R}^m$ $(d < m)$ characterized by a nonlinear mapping, $f : C \subset \mathbb{R}^d \to \mathbb{R}^m$, where $C$ is a compact and connected subset of $\mathbb{R}^d$. Here $\mathbb{R}^m$ represents the high-dimensional data space and $\mathbb{R}^d$ represents the low-dimensional parameter space. Given a set of data points $x_1, \cdots, x_N \in \mathbb{R}^m$ with

$$x_i = f(\tau_i), \quad i = 1, \ldots, N, \tag{01}$$

where $\tau_i \in C$, the problem of dimensionality reduction is to recovery low dimensional coordinates (parameterization) $\tau_i$'s from the $x_i$'s. For the theoretical purpose, we consider noise-free date (01) and we follow Donoho and Grimes (Donoho and Grimes 2003) to assume that $f$ is a local isometry.

Since the publications of LLE (Roweis and Saul 2000) and Isomap (Tenenbaum, de Silva, and Langford 2000), several competitive algorithms have been proposed for nonlinear dimensionality reduction, which include Laplacian Eigenmap (Belkin and Niyogi 2002), Hessian Eigenmap (Donoho and Grimes 2003), and LTSA (Local Tangent Space Alignment) (Li, Li, and Ye 2004) among many others; see (Saul et al. 2006) for a thorough review. One idea underlying several of these methods is to reconstruct global coordinates $\tau_i$ from their local relations as defined by data points in a small neighborhood. For example, the LTSA method (Zha and Zhang 2005) constructs global coordinates through first constructing and then aligning local coordinates.



One common assumption for methods based on local relations is that the underlying manifolds for the local neighborhoods (or the sets of local points) all have the same dimension $d$. (Here, we say a set of data points (01) is of dimension $p$ if the corresponding set of coordinates $\tau_i$ spans, after being centered, a $p$-dimensional space.) However, there are many situations where such an assumption may not hold. For example, the data points may lie on several manifolds of different dimensions or they may be sampled from a $d$-dimensional manifold with lower dimensional branches/sections. Then the ability of dimensionality reduction algorithms to detect and work with change of dimension in the data set is very important.

In this paper, we consider the alignment algorithm for reconstructing global coordinates from local coordinates that is derived in the LTSA method. We shall show that, under certain conditions, the alignment algorithm can successfully recover global coordinates from local neighborhoods of different dimensions. Our main results generalize the analysis of Ye, Zha and Li (Ye, Zha, and Li 2007) to allow alignment of sections of different dimensions. We shall also consider an application to a semisupervised learning problem (Ham, Lee, and Saul 2004) where one wishes to find full association of two data sets that are partially associated.

This paper is organized as follows. We first review the alignment algorithm and present related notation in Section 2. We present an analysis of the alignments of sections of different dimensions in Section 3. We discuss a semisupervised learning problem in Section 4 and we give two image examples in section 5.

NOTATION. We use $e$ to denote a column vector of all ones of appropriate dimension determined from the context. $\mathsf{null}(\cdot)$ is the null space of a matrix, and $\mathsf{span}(\cdot)$ denote the subspace spanned by all the columns of the argument matrix. $M^\dagger$ denotes the pseudo inverse of $M$.

## Alignment Algorithm

Consider the data set (01). Let $\boldsymbol{X} = \{x_1, \cdots, x_N\}$ and let $\{\boldsymbol{X}_i, i = 1, \ldots, s\}$ be a collection of subsets of $\boldsymbol{X}$ with $\boldsymbol{X}_i = \{x_{i_1}, \ldots, x_{i_{k_i}}\}$ $(i_1 < i_2 < \cdots < i_{k_i})$. Assume that $\cup_i \boldsymbol{X}_i = \boldsymbol{X}$, in which case we say $\{\boldsymbol{X}_i, i = 1, \ldots, s\}$ is a covering of $\boldsymbol{X}$. In the context of LTSA, each $\boldsymbol{X}_i$ is a small local neighborhood so that a coordinate system on the local tangent space can be obtained. In general, we assume that $\boldsymbol{X}_i$ is any subset such that an isometric coordinate $\{s_1^{(i)}, \ldots, s_{k_i}^{(i)}\} \subset \mathbb{R}^d$ can be constructed, i.e. $\|s_p^{(i)} - s_q^{(i)}\|_2 = d_{\mathcal{M}}(x_{i_p}, x_{i_q})$ (for any $1 \le p, q \le k_i$) where $d_{\mathcal{M}}(\cdot\,,\,\cdot)$ is the geodesic distance along $\mathcal{M}$. In practice, only an approximate isometric coordinate can be computed.

It has been shown (Zha and Zhang 2005; Ye, Zha, and Li 2007) that the global coordinates $\tau_i$'s can be constructed from the local coordinates through an alignment process as follows. Set

$$S_i = \left[s_1^{(i)}, \ldots, s_{k_i}^{(i)}\right]. \tag{02}$$

and define $Q_i$ to be the orthogonal projection onto the orthogonal complement of $\mathsf{span}\{[e, S_i^T]\}$ in $\mathbb{R}^{k_i}$. Let $E_i = [e_{i_1}, \ldots, e_{i_{k_i}}] \in \mathbb{R}^{N \times k_i}$, where $e_i \in \mathbb{R}^N$ is the $i$-th column of $I_N$ (the $N \times N$ identity matrix). Let

$$\Psi_i = E_i Q_i E_i^T, \quad \Psi = \sum_{i=1}^{s} \Psi_i. \tag{03}$$

Note that $\Psi_i$ is the embedding of $Q_i$ into an $N \times N$ matrix such that the $(i_p, i_q)$th element of $\Psi_i$ is the $(p, q)$th element of $Q_i$. $\Psi$ is called the alignment matrix for $\{S_i\}$.

Under a condition called fully overlap for the covering $\{\boldsymbol{X}_i\}$, it is shown in Ye, Zha and Li (Ye, Zha, and Li 2007, Theorem 2.7) that

$$\mathsf{null}\{\Psi\} = \mathsf{span}\{[e, T^T]\}$$

where $T = [\tau_1, \cdots, \tau_N]$. Hence, the global coordinates $\tau_i$'s can be obtained from computing $\mathsf{null}\{\Psi\}$, up to an orthogonal transformation (a rigid motion). For the ease of references, we state the alignment process as follows.

**Algorithm 01** **Alignment Algorithm:**
Given $\boldsymbol{X} = \{x_1, \cdots, x_N\} \subset \mathbb{R}^n$.

1. Construct a fully overlapped covering $\{\boldsymbol{X}_i, i = 1, \ldots, s\}$ with $\boldsymbol{X}_i = \{x_{i_1}, \ldots, x_{i_{k_i}}\}$.
2. For each $\boldsymbol{X}_i$, construct its local coordinates $s_1^{(i)}, \ldots, s_{k_i}^{(i)}$.
3. Construct $\Psi$ from $S_i = [s_1^{(i)}, \ldots, s_{k_i}^{(i)}]$ as in (03)
4. compute $[e/\sqrt{N}, Q^T]$ as an orthonormal basis of the spectral subspace of $\Psi$ corresponding to the smallest $d+1$ eigenvalues, where $Q^T \in \mathbb{R}^{N \times d}$.
5. Recover $T$ as $T = WQ$, where $W = [S_1, \ldots, S_s][Q_1, \ldots, Q_s]^\dagger$ and $Q_i = QE_i$.

The fully overlapped condition guarantees sufficient intersection (overlap) among $\boldsymbol{X}_i$'s to allow alignments. In the case of two subsets $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$, it requires that the intersection $\boldsymbol{X}_1 \bigcap \boldsymbol{X}_2$ is of dimension $d$ (see Definition 01 below or (Ye, Zha, and Li 2007) for details). This immediately requires that all subsets $\boldsymbol{X}_i$ to have the same dimension $d$. However, the structure of the data set $\boldsymbol{X}$ may contain lower dimensional branches. In the next section, we generalize the analysis of (Ye, Zha, and Li 2007) to include such cases. Interestingly, the alignment algorithm still works as long as a generalized fully overlapped condition holds.

## Alignment of Sections of Different Dimensions

First we define the dimension of a data set or its coordinate set.

**Definition 01** *A data set $\boldsymbol{X}_0 = \{x_{i_1}, \ldots, x_{i_k}\}$ and the corresponding coordinate set $\boldsymbol{T}_0 = \{\tau_{i_1}, \ldots, \tau_{i_k}\}$ are said to be of dimension $p$ if*

$$\mathsf{rank}[\tau_{i_1} - \bar{\tau}, \tau_{i_2} - \bar{\tau}, \ldots, \tau_{i_k} - \bar{\tau}] = p \tag{04}$$

*where $\bar{\tau} = (\Sigma_{j=1}^k \tau_{i_j})/k$. We write $\dim(\boldsymbol{X}_0) = \dim(\boldsymbol{T}_0) = p$.*

It can be shown that (04) is equivalent to

$$\mathsf{rank}([e, T_0^T]) = 1 + p.$$

where $T_0 = [\tau_{i_1}, \ldots, \tau_{i_k}]$; see (Ye, Zha, and Li 2007).

As in (Ye, Zha, and Li 2007), our analysis begins with the construction of the alignment matrix based on $\tau_i$'s. Let $\boldsymbol{T} = \{\tau_1, \tau_2, \cdots, \tau_N\} \subset \mathbb{R}^d$ and let $\{\boldsymbol{T}_i, 1 \le i \le s\}$ be the collection corresponding to $\{\boldsymbol{X}_i, 1 \le i \le s\}$, i.e.

$$\boldsymbol{T}_i = \{\tau_{i_1}, \cdots, \tau_{i_{k_i}}\}, \quad i_1 < i_2 < \cdots < i_{k_i}. \tag{05}$$

Set

$$T = [\tau_1, \cdots, \tau_N] \in \mathbb{R}^{d \times N}, \quad T_i = [\tau_{i_1}, \cdots, \tau_{i_{k_i}}]. \tag{06}$$

Let $P_i$ be orthogonal projection onto the orthogonal complement of span$([e, T_i^T])$, i.e., $\mathsf{null}(P_i) = \mathsf{span}([e, T_i^T])$. Define

$$\Phi_i = E_i P_i E_i^T; \;\; \Phi = \sum_{i=1}^{s} \Phi_i. \tag{07}$$

$\Phi$ is the alignment matrix of the collection $\{T_i\}$. If $\boldsymbol{S}_i$ is isometric to $\boldsymbol{X}_i$ (and hence to $\boldsymbol{T}_i$), then it can be shown that $\Psi = \Phi$, see (Ye, Zha, and Li 2007).

First, we extend the definition of fully overlap to sets with different dimensions.

**Definition 02** *Let $\boldsymbol{T}_1$ and $\boldsymbol{T}_2$ be two subsets of $\boldsymbol{T} \subset \mathbb{R}^d$. We say $\boldsymbol{T}_1$ and $\boldsymbol{T}_2$ are fully overlapped if*

$$\min\{\dim(\boldsymbol{T}_1), \dim(\boldsymbol{T}_2)\} = \dim(\boldsymbol{T}_1 \cap \boldsymbol{T}_2).$$

**Definition 03** *This definition is recursive. Let $\boldsymbol{T}_i, 1 \le i \le s$, be $s$ subsets of $R^d$. The collection $\{\boldsymbol{T}_i, 1 \le i \le s\}$ is fully overlapped if it can be partitioned into two nonempty disjoint collections, say, $\{\boldsymbol{T}_i, i = 1, \ldots, p\}$ and $\{\boldsymbol{T}_i, i = p+1, \ldots, s\}$ each of which is a fully overlapped collection, and if the union sets of the two collections $\hat{\boldsymbol{T}}_1 := \cup_{i=1}^{p} \boldsymbol{T}_i$ and $\hat{\boldsymbol{T}}_2 := \cup_{i=p+1}^{s} \boldsymbol{T}_i$ are fully overlapped.*

**Definition 04** *The collection $\{\boldsymbol{T}_i, 1 \le i \le s\}$ is a covering of $\boldsymbol{T}$ if $\cup_{i=1}^{s} \boldsymbol{T}_i = \boldsymbol{T}$, and a fully overlapped covering if the collection is a covering and fully overlapped.*

We now show that this fully overlapped condition is sufficient to guarantee reconstruction of $T$ from $\Phi$ or $\Psi$. First, the following is a lemma from (Ye, Zha, and Li 2007).

**Lemma 01** *Let $\{\boldsymbol{T}_i, 1 \le i \le s\}$ be a covering of $\boldsymbol{T}$, and let $\Phi_i$ and $\Phi$ be defined as in (07). Then*

$$\begin{aligned} \mathsf{null}(\Phi_i) &= \{x \mid E_i^T x \in \mathsf{span}([e, T_i^T])\} \\ \mathsf{null}(\Phi) &= \bigcap_{i=1}^{s} \mathsf{null}(\Phi_i). \end{aligned}$$

**Theorem 01** *Let $\{\boldsymbol{T}_1, \boldsymbol{T}_2\}$ be a fully overlapped covering of $\boldsymbol{T}$ and let $\Phi_i$ and $\Phi$ be defined as in (07). We have* $\mathsf{null}\{\Phi\} = \mathsf{span}[e, T^T]$.

*Proof:* Without loss of generality, we assume that $d_2 := \dim(\boldsymbol{T}_2) < d_1 := \dim(\boldsymbol{T}_1)$. Then $\mathrm{rank}([e, T_2^T]) = d_2 + 1$. There is a nonsingular matrix $U$, such that $[e, T_1^T]U = [e, \widetilde{T}_1^T]$ and $[e, T_2^T]U = [e, \widetilde{T}_2^T]$ with the last $d - d_2$ columns of $\widetilde{T}_2^T$ being all zero. Suppose there are $k$ vectors in $\boldsymbol{T}_1 \cap \boldsymbol{T}_2$. Without loss of generality, we assume that the last $k$ columns of $T_1$ and the first $k$ columns of $T_2$ are the vectors in $\boldsymbol{T}_1 \cap \boldsymbol{T}_2$. Then we write

$$\widetilde{T}_1 = \begin{array}{c} {\scriptstyle d_2} \\ {\scriptstyle d-d_2} \end{array} \begin{pmatrix} \widetilde{T}_{11}^{(1)} & \widetilde{T}_{12} \\ \widetilde{T}_{11}^{(2)} & 0 \end{pmatrix}; \ \widetilde{T}_2 = \begin{array}{c} {\scriptstyle d_2} \\ {\scriptstyle d-d_2} \end{array} \begin{pmatrix} \widetilde{T}_{21} & \widetilde{T}_{22} \\ 0 & 0 \end{pmatrix}$$

where $\widetilde{T}_{12} = \widetilde{T}_{21}$. Next, let the columns of $Q$ form a basis of $\mathsf{null}(\Phi)$. We have $\mathsf{span}(Q) \subset \{x | E_i^T x \in \mathsf{span}([e, \widetilde{T}_i^T])\}$ for each $i$. Then we can find a matrix $W_i \in R^{(d+1)\times m}$, where $m = \dim(\mathsf{null}(\Phi))$, such that $E_i^T Q = [e, \widetilde{T}_i^T]W_i$. Let

$$W_i = \begin{array}{c} {\scriptstyle d_2+1} \\ {\scriptstyle d-d_2} \end{array} \begin{pmatrix} W_i^{(1)} \\ W_i^{(2)} \end{pmatrix}.$$

Comparing the common rows of $E_1^T Q$ and $E_2^T Q$, we have

$$[e, \begin{pmatrix} \widetilde{T}_{12}^T & 0 \end{pmatrix}]W_1 = [e, \begin{pmatrix} \widetilde{T}_{21}^T & 0 \end{pmatrix}]W_2.$$

From the first $d_2 + 1$ columns of last equation, we obtain that

$$[e, \widetilde{T}_{12}^T]W_1^{(1)} = [e, \widetilde{T}_{21}^T]W_2^{(1)}.$$

Since $\boldsymbol{T}_1$ and $\boldsymbol{T}_2$ are fully overlapped, we have that $[e, \widetilde{T}_{12}^T]$ has full column rank. From

$$[e, \widetilde{T}_{12}^T](W_1^{(1)} - W_2^{(1)}) = 0,$$

it follows

$$W_1^{(1)} - W_2^{(1)} = 0.$$

Noting that

$$\begin{aligned} [e, \widetilde{T}_2^T]W_2 &= \left[e, \begin{pmatrix} \widetilde{T}_{21}^T & 0 \\ \widetilde{T}_{22}^T & 0 \end{pmatrix}\right] \begin{pmatrix} W_2^{(1)} \\ W_2^{(2)} \end{pmatrix} \\ &= \left[e, \begin{pmatrix} \widetilde{T}_{21}^T & 0 \\ \widetilde{T}_{22}^T & 0 \end{pmatrix}\right] \begin{pmatrix} W_1^{(1)} \\ W_1^{(2)} \end{pmatrix} \\ &= [e, \widetilde{T}_2^T]W_1, \end{aligned}$$

we have

$$E_i^T Q = [e, T_i^T]UW_1.$$

So we can write $Q$ as

$$Q = [e, T^T]UW_1.$$

Thus

$$\mathsf{null}\{\Phi\} = \mathsf{span}[e, T^T].$$

**Theorem 02** *Let $\{\boldsymbol{T}_i, i = 1, \ldots s\}$ be a fully overlapped covering of $\boldsymbol{T}$ and let $\Phi_i$ and $\Phi$ be defined as in (07). Then* $\mathsf{null}\{\Phi\} = \mathsf{span}[e, T^T]$.

*Proof:* This is proved by virtually the same induction as in the proof of (Ye, Zha, and Li 2007, Theorem 2.6) using Theorem 01 and Definition 03. We omit the details.

In practice, when we have a neighborhood consisting of points lying on a lower dimensional branches, their coordinates are likely computed with large errors in the components that are supposed to be zero. Amazingly, with a slightly extra condition, this does not affect the result of the alignment process. Before we present an analysis, we first illustrate with an example.

**Example 01** *Let $a, b, c, d, f, g, u, v, w, x$ be distinct numbers and*

$$\boldsymbol{T} = \left\{ \begin{bmatrix} a \\ g \end{bmatrix}, \begin{bmatrix} b \\ 0 \end{bmatrix}, \begin{bmatrix} c \\ 0 \end{bmatrix}, \begin{bmatrix} d \\ 0 \end{bmatrix}, \begin{bmatrix} f \\ 0 \end{bmatrix} \right\}.$$

*Assume that we have two subsets*

$$\boldsymbol{T}_1 = \left\{ \begin{bmatrix} a \\ g \end{bmatrix}, \begin{bmatrix} b \\ 0 \end{bmatrix}, \begin{bmatrix} c \\ 0 \end{bmatrix}, \begin{bmatrix} d \\ 0 \end{bmatrix} \right\}$$

*and*

$$\boldsymbol{T}_2 = \left\{ \begin{bmatrix} b \\ 0 \end{bmatrix}, \begin{bmatrix} c \\ 0 \end{bmatrix}, \begin{bmatrix} d \\ 0 \end{bmatrix}, \begin{bmatrix} f \\ 0 \end{bmatrix} \right\}.$$

*Then $\dim(\boldsymbol{T}_1) = 2$ and $\dim(\boldsymbol{T}_2) = \dim(\boldsymbol{T}_1 \cap \boldsymbol{T}_2) = 1$. By Theorem 01, we can recover $T$ from $\Phi$ as constructed from $T_1$ and $T_2$.*

*In practice, however, we can only compute two coordinate sets $\boldsymbol{S}_1$ and $\boldsymbol{S}_2$ that are (approximately) isometric to $\boldsymbol{T}_1$ and $\boldsymbol{T}_2$. However, large errors could be present in the second components of $\boldsymbol{T}_2$. For example, when computing $\boldsymbol{T}_2^T$ from local first order approximation (Li, Li, and Ye 2004), they are computed as 2 smallest singular vectors and the second components derived from a singular vector corresponding to a tiny singular value may effectively be random. $\boldsymbol{T}_1$ can be computed accurately, however. Suppose the computed coordinates for the two sections are*

$$S_1 = \begin{bmatrix} a & b & c & d \\ g & 0 & 0 & 0 \end{bmatrix}; \ S_2 = \begin{bmatrix} b & c & d & f \\ u & v & w & x \end{bmatrix}.$$

*Now, constructing $\Psi$ from $S_i$ as in (03). Using Maple, we can compute $\Psi$ and verify that* $\mathsf{null}\{\Psi\} = \mathsf{span}\{[e, T^T]\}$. *Hence, even when the second components in $S_2$ are computed completely wrong, original $T$ can still be recovered from $\Psi$!*

The phenomenon explained in the example above is true in general as shown in the following theorem.

**Theorem 03** *Let $\{\boldsymbol{T}_1, \boldsymbol{T}_2\}$ be a fully overlapped covering of $\boldsymbol{T} \subset \mathbb{R}^d$ with $\dim(\boldsymbol{T}_1) = d_1$ and $\dim(\boldsymbol{T}_2) = \dim(\boldsymbol{T}_1 \cap \boldsymbol{T}_2) = d_0 < d_1$. Assume that the vectors in $\boldsymbol{T}_2$ have vanishing last $d - d_0$ components. Let $\boldsymbol{S}_1 = \boldsymbol{T}_1$ and*

$$\boldsymbol{S}_2 = \{ \begin{array}{c} {\scriptstyle d_0} \\ {\scriptstyle d-d_0} \end{array} \begin{pmatrix} \hat{\tau}_i \\ \hat{\rho}_i \end{pmatrix} : \begin{array}{c} {\scriptstyle d_0} \\ {\scriptstyle d-d_0} \end{array} \begin{pmatrix} \hat{\tau}_i \\ 0 \end{pmatrix} \in \boldsymbol{T}_2 \}.$$

*Let $\Psi_i$ and $\Psi$ be defined from $S_i$ as in (03). If the points of $\boldsymbol{S}_2$ that correspond to $\boldsymbol{T}_1 \cap \boldsymbol{T}_2$ form a $d$-dimensional set, i.e.*

$$\dim\left(\left\{ \begin{pmatrix} \hat{\tau}_i \\ \hat{\rho}_i \end{pmatrix} \in \boldsymbol{S}_2 : \begin{pmatrix} \hat{\tau}_i \\ 0 \end{pmatrix} \in \boldsymbol{T}_1 \cap \boldsymbol{T}_2 \right\}\right) = d \quad (08)$$

*then we have $null\{\Psi\} = span[e, T^T]$.*

*Proof:* Without loss of generality, we assume that the last $k$ columns of $T_1$ and the first $k$ columns of $T_2$ are the vectors in $\boldsymbol{T}_1 \cap \boldsymbol{T}_2$. Write

$$T_1 = \begin{array}{c} {\scriptstyle d_0} \\ {\scriptstyle d-d_0} \end{array} \begin{pmatrix} T_{11}^{(1)} & T_{12}^{(1)} \\ T_{11}^{(2)} & 0 \end{pmatrix} \quad T_2 = \begin{array}{c} {\scriptstyle d_0} \\ {\scriptstyle d-d_0} \end{array} \begin{pmatrix} T_{21}^{(1)} & T_{22}^{(1)} \\ 0 & 0 \end{pmatrix}$$

where $T_{12}^{(1)} = T_{21}^{(1)}$. Let $S_1$ and $S_2$ be the matrices whose columns are the vectors in $\boldsymbol{S}_1$ and $\boldsymbol{S}_2$ respectively, i.e. we write

$$S_1 = T_1 = \begin{array}{c} {\scriptstyle d_0} \\ {\scriptstyle d-d_0} \end{array} \begin{pmatrix} T_{11}^{(1)} & T_{12}^{(1)} \\ T_{11}^{(2)} & 0 \end{pmatrix} \quad S_2 = \begin{array}{c} {\scriptstyle d_0} \\ {\scriptstyle d-d_0} \end{array} \begin{pmatrix} T_{21}^{(1)} & T_{22}^{(1)} \\ T_{21}^{(2)} & T_{22}^{(2)} \end{pmatrix}.$$

Let $Q$ be such that its columns form a basis for $\mathsf{null}(\Psi)$. We have $\mathsf{span}(Q) \subset \{x | E_i^T x \in \mathsf{span}([e, S_i^T])\}$. Then we can find a matrix $W_i \in R^{(d+1)\times m}$, where $m = \dim(\mathsf{null}(\Phi))$, such that $E_i^T Q = [e, S_i^T] W_i$. Let

$$W_i = \begin{array}{c} {\scriptstyle d_0+1} \\ {\scriptstyle d-d_0} \end{array} \begin{pmatrix} W_i^{(1)} \\ W_i^{(2)} \end{pmatrix}.$$

Then we have

$$\left[e, \begin{pmatrix} T_{12}^{(1)^T} & 0 \end{pmatrix}\right] W_1 = \left[e, \begin{pmatrix} T_{21}^{(1)^T} & T_{21}^{(2)^T} \end{pmatrix}\right] W_2.$$

Equivalently,

$$[e, T_{12}^{(1)^T}] W_1^{(1)} = [e, T_{21}^{(1)^T}] W_2^{(1)} + T_{21}^{(2)^T} W_2^{(2)}.$$

Noting $T_{12}^{(1)^T} = T_{21}^{(1)^T}$, we have

$$[e, T_{21}^{(1)^T}](W_1^{(1)} - W_2^{(1)}) = T_{21}^{(2)^T} W_2^{(2)}.$$

Using (08), we see that $[e, T_{21}^{(1)^T}, T_{21}^{(2)^T}]$ has full column rank. It follows from

$$[e, T_{21}^{(1)^T}](W_1^{(1)} - W_2^{(1)}) - T_{21}^{(2)^T} W_2^{(2)} = 0$$

that

$$[e, T_{21}^{(1)^T}](W_1^{(1)} - W_2^{(1)}) = 0, \quad T_{21}^{(2)^T} W_2^{(2)} = 0.$$

This further implies that

$$W_1^{(1)} - W_2^{(1)} = 0, \quad W_2^{(2)} = 0.$$

Thus $W_1^{(1)} = W_2^{(1)}$. Then we can write

$$E_2^T Q = [e, S_2^T] W_2 = [e, T_2^T] W_1.$$

This together with $E_1^T Q = [e, T_1^T] W_1$ implies that

$$Q = [e, T^T] W_1.$$

Therefore,

$$\mathsf{null}\{\Psi\} = \mathsf{span}[e, T^T].$$

## Semisupervised Alignment of Manifolds

The results in the previous section show that the alignment algorithm is capable of aligning sections of different dimensions. Other than the alignment, it has application in other context. Here we consider a problem in semisupervised learning of manifolds that has been introduced in (Ham, Lee, and Saul 2004).

Assume that there are two data sets that admit a pairwise correspondence, some of which is known. The objective is to generate full association (correspondence) of the data sets from the partial association of samples. One approach to this problem is to first generate a common low-dimensional embedding for those two data sets. From the common embedding, we can associate samples between the two data sets.

For example, in applications of images, we have two sets of pictures of two different objects taken by a camera from various positions and angles and we wish to match images taken from the same positions/angles, provided matching of a sample is available.

Let $\boldsymbol{X}$ and $\boldsymbol{Y}$ be two data sets with two subsets $\boldsymbol{X}_l$ and $\boldsymbol{Y}_l$. Assume that $\boldsymbol{X}_l$ and $\boldsymbol{Y}_l$ are already in pairwise correspondence. $\boldsymbol{X}$ has $M$ samples points. $\boldsymbol{Y}$ has $N$ samples points. $\boldsymbol{X}_l$ and $\boldsymbol{Y}_l$ each has $k$ samples points. We first find the low-dimensional embedding for each data set $\boldsymbol{X}$ and $\boldsymbol{Y}$, which are $Z_1$ and $Z_2$. Assume that the vectors corresponding to $\boldsymbol{X}_l$ and $\boldsymbol{Y}_l$ are in the first $k$ columns, i.e. $Z_1 = [Z_1^{(l)}, Z_1^{(u)}] \in \mathbb{R}^{d\times M}$ and $Z_2 = [Z_2^{(l)}, Z_2^{(u)}] \in \mathbb{R}^{d\times N}$ where $Z_1^{(l)}$ and $Z_2^{(l)}$ are the low-dimensional parametrization of $\boldsymbol{X}_l$ and $\boldsymbol{Y}_l$ respectively. Given the association of the sample points in $\boldsymbol{X}_l$ and $\boldsymbol{Y}_l$, $Z_1^{(l)}$ and $Z_2^{(l)}$ have the same intrinsic parametrization, say, $Z_I^{(l)}$ or its transformation. To find the association between $Z_1^{(u)}$ and $Z_2^{(u)}$, we need to align $Z_1$ and $Z_2$ by determining two affine transformations $L_i$, $i = 1, 2$, such that $L_1 Z_1^{(l)} = L_2 Z_2^{(l)}$. Then $Z_{i,I} = L_i Z_i$, $i = 1, 2$, provide intrinsic parametrizations for $\boldsymbol{X}$ and $\boldsymbol{Y}$ from which a full association can be deduce.

This is a problem that can be solved by the alignment algorithm that is discussed in the previous section. Specifically, the alignment matrix constructed from $Z_i$ is the same as that constructed from $Z_{i,I}$. Let $P_i$ for $i = 1, 2$ be the alignment matrices for $Z_i$, and let $E_i$ be the selection matrices, such that $ZE_i = Z_i$. The embedding of $P_i$ into $\mathbb{R}^{M+N-l}$ is $\Phi_i$, where $\Phi_i = E_i P_i E_i^T$. Then we construct an alignment matrix from $Z_1$ and $Z_2$ as below:

$$\begin{aligned} \Phi &= \Phi_1 + \Phi_2, \\ &= E_1 P_1 E_1^T + E_2 P_2 E_2^T. \end{aligned}$$

$P_i$ can be constructed from the local coordinate patches of $Z_i$ as in (03). The null space of the alignment matrix $\Phi$ yields a joint embedding of those two data sets, which is the intrinsic low-dimensional representation for those data sets up to a rigid motion.

## Numerical Examples

In this section, we present two examples to show the alignment algorithm works well with sections having different dimensions. We consider two examples. The first example has

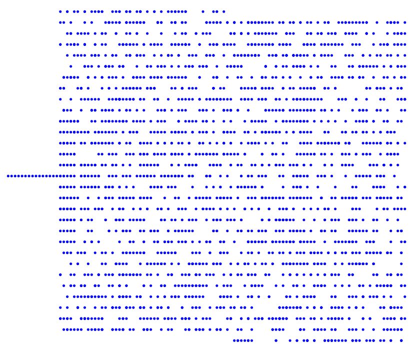

Figure 1: The generating coordinates of the data set.

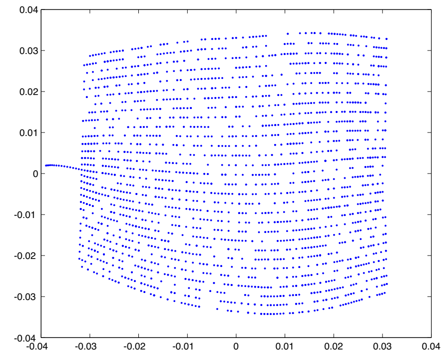


Figure 2: The reconstructed coordinates of the data set by LTSA with $k = 15$ and $d = 2$.

a set of face images generated from a 3D face model. One section of this image set has intrinsic dimensionality one. The other section of the image set has intrinsic dimensionality two. We try to find the low-dimensional parametrization for this image set.

**Example 02** *The data set consists of $N = 2715$ face images generated based on the 3D face model in (Blanz and Vetter 1999). The set contains $64 \times 64$ images of the same face, which are obtained by varying pan and tilt angles for the observer. For 2700 images, they vary from $-30$ to 45 degrees of pan angles and $-10$ to 10 degrees for tilt angles. For 15 images, they vary from $-45$ to $-30$ degrees of pan angles and have 0 degree of tilt angles. We are interested in finding the low-dimensional parametrization for these face images. The original coordinates of all those 2715 pictures are showed in Figure 1, where the x-axis is the pan angle and the y-axis is the tilt angle. From Figure 1, we can see that those 2700 images form a manifold with dimensionality $d = 2$, whereas the other 15 images with the same tilt angles form a branch with dimensionality $d = 1$.*

*We implement LTSA algorithm with fifteen neighbors of each $x_i$ ($k = 15$) and dimension two ($d = 2$) to recover the parameters of the images. We notice that for those 15 images with the same tilt angles, their local coordinates should have intrinsic dimensionality one with this example, but our algorithm will treat it as if it were dimension two, having the second components derived from a singular vector corresponding to a tiny singular value. The reconstructed coordinates of all these 2715 images after LTSA are showed in Figure 2. Though one set of these data points is of intrinsic dimensionality one and the other set of data points is of intrinsic dimensionality two, LTSA recover the parametrization correctly. The lower dimensional branch is clearly identified by the algorithm from 2.*

Our second example concerns two sets of face images generated from different face model. We are interested in finding the face images shoot from the same tilt and pan angles.

**Example 03** *We have two sets of pictures generated from two different 3D face models in (Blanz and Vetter 1999). The pictures of two different persons are taken from different pan and tilt angle. We are interested in matching the images with the same pan and tilt angles from different image sets. This problem can be solved by the semisupervised alignment of manifolds (Ham, Lee, and Saul 2004) that we discussed in the previous section.*

*The first data set $\boldsymbol{X}$ contains 100 pictures coming from face model A and all these pictures have the same tilt angle of 0 degree and pan angles varying from $-45$ to 45 degrees. The second data set $\boldsymbol{Y}$ contains 2700 pictures generated from face model B. These pictures have pan angles varying from $-45$ to 45 degrees and tilt angle varying from $-10$ to 10 degrees . The goal is to match the images with the same tilt angle and the same pan angle. First, 20 matching pairs of pictures in $\boldsymbol{X}$ and $\boldsymbol{Y}$ are manually chosen so that each pair of images are shoot from the same tilt angle and pan angle. These $l = 20$ pictures are labeled samples. We first compute the alignment matrices $P_1$ and $P_2$ of those two data set $\boldsymbol{X}$ and $\boldsymbol{Y}$ separately from the local coordinates with fifteen neighbors(k=15) and dimension 2(d=2). Next, we construct the alignment matrix $\Phi = E_1 P_1 E_1^T + E_2 P_2 E_2^T$. Then, we compute the joint embedding by the alignment matrix.*

*For the data set $\boldsymbol{X}$, we notice that the intrinsic dimensionality should be one, whereas we calculate the parametrization with dimension two which is necessary in order to carry out alignment. In the left plot of Figure 3, we show the computed local coordinates of data set $\boldsymbol{X}$ with fifteen neighbors ($k = 15$) and dimensions two ($d = 2$). We see that it shows a one-dimensional curve embedded in a 2 dimensional space. However, this detected after the alignment process. The right plot of Figure 3 shows the computed global coordinates of data sets $\boldsymbol{X}$ and $\boldsymbol{Y}$ by the alignment algorithm based on LTSA. The red circle line shows the points corresponding to those in $\boldsymbol{X}$. Again, the alignment algorithm works well with data sets of different dimensions.*

*We now discuss how to match the unlabeled samples between $\boldsymbol{X}$ and $\boldsymbol{Y}$. For any image $t$ from the data set $\boldsymbol{X}$ or $\boldsymbol{Y}$, we can find a low-dimensional parametrization $f(t)$ by the alignment algorithm. Given one unlabeled sample picture $x$ from the data set $\boldsymbol{X}$ as the input, which has a parameter*

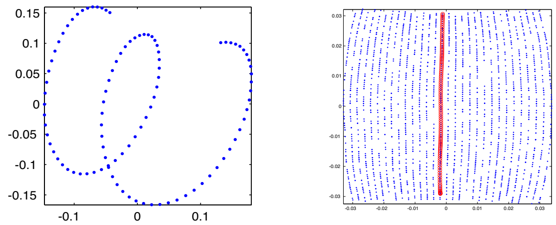


Figure 3: **Left** The computed local coordinates of data set $\boldsymbol{X}$ with $k = 15$ and $d = 2$. **Right** The computed global coordinates of data sets $\boldsymbol{X}$ (red circles) and $\boldsymbol{Y}$ (blue dots) after alignment algorithm based on LTSA with $k = 15$ and $d = 2$.

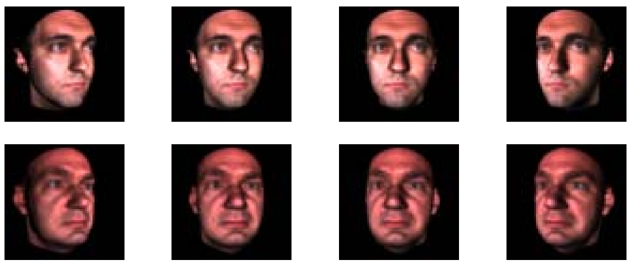

Figure 4: **Up** The four unlabeled pictures from data set $\boldsymbol{X}$ with different pan angles. **Down** The computed four pictures of data set $\boldsymbol{Y}$ matched to the four unlabeled pictures from data set $\boldsymbol{X}$.

*$f(x)$, we find an image $y^* \in \boldsymbol{Y}$ with parameter $f(y^*)$, such that*

$$y^* = \arg_{y \in \boldsymbol{Y}} \min \|f(x) - f(y)\|_2.$$

*We take four unlabeled sample pictures from $\boldsymbol{X}$ data as the input and show the best matching data for $\boldsymbol{Y}$ found and show the pictures in Figure 4. There is a clear match in the pan and tilt angle for the pairs.*

*Our examples confirm our theoretical results that the alignment algorithm can recover parametrization properly even when local neighborhoods/sections have different intrinsic dimension. This is a property not known for other manifold learning algorithms and would be an advantage of the alignment algorithm.*

## Acknowledgment

This work is supported in part by NSF under grant DMS-0915062.

## References

Belkin, M., and Niyogi, P. 2002. Laplacian eigenmaps for dimensionality reduction and data representation. *Neural Computation* 15:1373–1396.

Blanz, V., and Vetter, T. 1999. A morphable model for the synthesis of 3d faces. *SIGGRAPH* (187–194).

Donoho, D., and Grimes, C. 2003. Hessian eigenmaps: new tools for nonlinear dimensionality reduction. *Proceedings of National Academy of Science* (5591–5596).

Ham, J.; Lee, D.; and Saul, L. 2004. Semisupervised alignment of manifolds. In *AISTATS*.

Li, C.-K.; Li, R.-C.; and Ye, Q. 2004. Principal manifolds and nonlinear dimension reduction via local tangent space alignment. *SIAM J. Sci. Comput.* 26:313–338.

Roweis, S. T., and Saul, L. K. 2000. Nonlinear dimensionality reduction by locally linear embedding. *SCIENCE* 290:2323–2326.

Saul, L.; Weinberger, K.; Ham, J.; Sha, F.; and Lee, D. 2006. Spectral methods for dimensionality reduction. In *Semisupervised Learning*.

Tenenbaum, J. B.; de Silva, V.; and Langford, J. C. 2000. A global geometric framework for nonlinear dimensionality reduction. *SCIENCE* 290:2319–2326.

Ye, Q.; Zha, H.; and Li, R.-C. 2007. Analysis of an algnment algorithm for nonlinear dimensionality reduction. *BIT- Numerical Mathematics* (873–885).

Zha, H., and Zhang, Z. 2005. Spectral analysis of alignment in manifold learning. In *Proceedings of Acoustics, Speech, and Signal Processing*, 1069–1072.