

# ANALYSIS OF AN ALIGNMENT ALGORITHM FOR NONLINEAR DIMENSIONALITY REDUCTION

QIANG YE <sup>\*</sup>, HONGYUAN ZHA <sup>†</sup>, AND REN-CANG LI <sup>‡</sup>

**Abstract.** The goal of dimensionality reduction or manifold learning for a given set of high-dimensional data points is to find a low-dimensional parametrization for them. Usually it is easy to carry out this parametrization process within a small region to produce a collection of local coordinate systems. Alignment is the process to stitch those local systems together to produce a global coordinate system and is done through the computation of a partial eigendecomposition of a so-called alignment matrix. In this paper, we present an analysis of the alignment process giving conditions under which the null space of the alignment matrix recovers the global coordinate system up to an affine transformation. We also propose a post-processing step that can determine the global coordinate system up to a rigid motion. This in turn shows that Local Tangent Space Alignment method (LTSA) can recover locally isometric embedding up to a rigid motion.

**1. Introduction.** An important goal of exploratory data analysis and data mining is to discover compact and informative descriptions and summaries of high-dimensional data which can then be used for further processing such as data visualization, classification, outlier detection and feature selection [7]. One way to summarize the data is to group them into several clusters so that data objects within a cluster are more similar than those across clusters, this leads to the so-called cluster analysis [7, 8]. Another way to obtain a compact description of the data is through dimensionality reduction which is the main motivation for the analysis presented in this paper.

The problem of dimensionality reduction is concerned with determining a low-dimensional parametrization for a given set of high-dimensional data points lying on a manifold, which is also known as manifold learning, see [11, 12]. Mathematically, consider a *parameterized manifold* of dimension  $d$  defined by a mapping  $f : \Omega \subset \mathbb{R}^d \rightarrow \mathbb{R}^m$ , the  $m$ -dimensional real vector space, where  $d < m$  and  $\Omega$  is open in  $\mathbb{R}^d$ . Suppose we have a set of points  $x_1, \dots, x_N$  sampled from the manifold, i.e.,

$$(1.1) \quad x_i = f(\tau_i), \quad i = 1, \dots, N,$$

for some  $\tau_i \in \Omega$ . We are interested in recovering the coordinate vectors  $\tau_i$ 's and/or the mapping  $f(\cdot)$  from the  $x_i$ 's [11, 12]. We typically need to impose additional constraints on the mapping so that the parametrization is well determined and preserves certain properties of the original data. For example, if  $f$  is restricted to be an isometric mapping, then  $\tau_i$ 's are uniquely determined up to a rigid motion and capture the geometric structure of the data. With the dimension  $d$  generally being much smaller than  $m$ , the parametric vectors  $\tau_i$  offer a low-dimensional representation of  $x_i$  and could reveal underlying structure of the data points. In particular, the computed  $\tau_i$ 's can be further used for visualization and clustering [5, 7, 8].

---

<sup>\*</sup>Department of Mathematics, University of Kentucky, Lexington, KY 40506 ([qye@ms.uky.edu](mailto:qye@ms.uky.edu)). Supported in part by the NSF Grant DMS-0411502.

<sup>†</sup>Department of Computer Science and Engineering, The Pennsylvania State University, University Park, PA 16802 ([zha@cse.psu.edu](mailto:zha@cse.psu.edu)). Supported in part by NSF grants DMS-0311800 and DMS-0405681.

<sup>‡</sup>Department of Mathematics, University of Texas at Arlington, P.O. Box 19408, Arlington, TX 76019-0408 ([rccli@uta.edu](mailto:rccli@uta.edu)). Supported in part by the National Science Foundation CAREER award under Grant No. CCR-9875201 and by the National Science Foundation under Grant No. DMS-0510664.

Traditionally, dimensionality reduction has been confined to the linear case, i.e. when the data points  $x_i$  lie on a linear manifold. With  $f$  being a linear map, the coordinates  $\tau_i$  can be recovered, up to a linear transformation, through a singular value decomposition [7]. The nonlinear case, however, has attracted much attention recently. We refer to the pioneering works of Tenenbaum, de Silva and Lanford [12] and Roweis and Saul [11] for discussions of some algorithms and applications of nonlinear dimensionality reduction.

For data points that lie on a nonlinear manifold, several numerical procedures have been developed recently that are based on local first order linear approximations. One approach is to regard a small neighborhood (i.e. a patch) of the manifold as a linear one and construct a local coordinate system for a small neighborhood around each sample point. The local coordinate systems generally overlap with each other and can then be aligned (either implicitly or explicitly) into a global one. Examples of the local approximation methods include LLE (Local Linear Embedding) [11], manifold charting [2], Hessian LLE [4] and LTSA (Local Tangent Space Alignment) [14]. The idea of global alignment was also discussed in [13].

A crucial step of these local approximation methods is the process of global alignment and this is carried out by computing eigenvectors of some matrices. In the LTSA method, a matrix, which we call an alignment matrix, is constructed from the local coordinates and then its eigenvectors corresponding to the second to the  $(d + 1)$ -st smallest eigenvalues (with the smallest being 0 as determined by the structure of the matrix) are used as the global coordinates, see [14] for the details. We note that these numerical procedures are derived through first order local approximations and the coordinates obtained should preserve local geometry of the data points. However, what we obtain globally is not clear. Indeed, there is very little theoretical analysis on a precise relation between the eigenvectors of the alignment matrix and the coordinate systems to be recovered.

In this paper, we shall present and theoretically analyze the global alignment procedure that produces a global coordinate system from the coordinate systems for some local patches of the manifold. We shall consider isometric mappings for the theoretical analysis, although the numerical procedures are applicable in a more general setting. Our main result is that, if the coordinate system in each local patch is computed exactly, then the null space of the alignment matrix recovers the global coordinate  $\tau_i$ 's up to an affine transformation, provided the local patches "overlap" sufficiently. Furthermore, with most existing methods constructing a global coordinate system directly from a basis of the null space (or eigenvectors) which is only determined up to an affine transformation, we also propose a postprocessing step that reduces this uncertainty to a rigid motion.

The paper is organized as follows. In Section 2, we present our analysis of the alignment matrix. In particular, we derive conditions under which the global coordinates can be recovered from those of the local patches. In Section 3, we describe the alignment process with a postprocessing step and use a numerical example to illustrate its utility. As a by-product, we show that the LTSA proposed in [14] can recover isometric embedding up to a rigid motion. We end the paper with some concluding remarks in Section 4.

NOTATION. We use  $e$  to denote a column vector of all ones the dimension of which should be clear from the context.  $\text{null}(\cdot)$  is the null space of a matrix, and  $\text{span}(\cdot)$  denote the subspace spanned by all the columns of argument matrices. For an index set  $\mathbf{I} = [i_1, \dots, i_k]$ ,  $A(:, \mathbf{I})$  denotes the submatrix of  $A$  consisting of columns

of  $A$  with indices in  $I$ .  $A^T$  denotes the transpose of matrix  $A$ .

**2. Null Space of the Alignment Matrix.** The setting of our discussion is the following: we are interested in computing a set of parameter vectors  $\tau_i$ 's given a set of data points  $x_i$ 's. We assume that the data set is partitioned into several subsets called "local" patch<sup>1</sup> of the manifold. On each of the patches, we compute a local coordinate system from which we construct an alignment matrix and recover the global coordinates [13, 14]. We first introduce the alignment matrix in a more special context.

Consider  $\mathcal{S} = \{\tau_1, \dots, \tau_N\} \subset \mathbb{R}^d$  and let  $\{\mathcal{S}_i, 1 \leq i \leq s\}$  be a collection of subsets of  $\mathcal{S}$ . Write

$$(2.1) \quad \mathcal{S}_i = \{\tau_{i_1}, \dots, \tau_{i_{k_i}}\}, \quad i_1 < i_2 < \dots < i_{k_i},$$

and set

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

We say  $T_i$  is a *section* of  $T$ . Let

$$E_i = [e_{i_1}, \dots, e_{i_{k_i}}] \in \mathbb{R}^{N \times k_i}$$

where  $e_i \in \mathbb{R}^N$  is the  $i$ -th column of  $I_N$ , the  $N \times N$  identity matrix. Then  $TE_i = T_i$ . Let  $P_i$  be the orthogonal projection onto the orthogonal complement of  $\text{span}([e, T_i^T])$ , i.e.,

$$(2.2) \quad \text{null}(P_i) = \text{span}([e, T_i^T]).$$

Embed  $P_i$  into an  $N$ -by- $N$  matrix  $\Phi_i$  such that, the  $(i_p, i_q)$ th element of  $\Phi_i$  is the  $(p, q)$ th element of  $P_i$ , i.e.,

$$(2.3) \quad \Phi_i = E_i P_i E_i^T.$$

Define

$$(2.4) \quad \Phi = \sum_{i=1}^s \Phi_i.$$

We call  $\Phi$  the *alignment matrix* for the collection  $\{\mathcal{S}_i\}$ .

We note that  $P_i$  can be constructed from  $\text{span}([e, \Theta_i^T])$  for any  $\Theta_i^T$  that spans the same subspace as  $T_i^T$ . For example, if for each  $T_i$ , we have  $\Theta_i$  satisfying  $\Theta_i = V_i T_i$  for some invertible matrix  $V_i$ , then  $P_i$  and hence  $\Phi$  can be obtained from such  $\Theta_i$  without knowing  $T_i$ , see Theorem 2.7 later for details. Our goal is to recover  $T$  from  $\Phi$ , which can be constructed from  $\Theta_i$  rather than  $T$ , and we achieve this by proving  $\text{span}([e, T^T]) = \text{null}(\Phi)$ . We first present several definitions and some preliminary results.

**DEFINITION 2.1.** Let  $\mathcal{S}_x = \{x_1, \dots, x_m\}$  and  $\mathcal{S}_y = \{y_1, \dots, y_n\}$  be two subsets of  $\mathbb{R}^d$ . Denote by  $\mathcal{S}_z = \mathcal{S}_x \cap \mathcal{S}_y = \{z_1, \dots, z_k\}$  the set of column vectors that are in the intersection of  $\mathcal{S}_x$  and  $\mathcal{S}_y$ . We say the two sets  $\mathcal{S}_x$  and  $\mathcal{S}_y$  are fully overlapped if

$$(2.5) \quad \text{span}(z_1 - \bar{z}, z_2 - \bar{z}, \dots, z_k - \bar{z}) = \mathbb{R}^d,$$

---

<sup>1</sup>The points in the subset need not be local in space but in practical uses, they are typically taken from a small neighborhood.

where  $\bar{z} = \left( \sum_{i=1}^k z_i \right) / k$ .

Let  $Z = [z_1, \dots, z_k]$ . Condition (2.5) is equivalent to

$$[z_1 - \bar{z}, z_2 - \bar{z}, \dots, z_k - \bar{z}] = Z - \bar{z}e^T \equiv Z \left( I - \frac{1}{k}ee^T \right) \text{ having full row rank,}$$

We shall use later the fact that this is also equivalent to

$$(2.6) \quad [e, Z^T] \text{ having full column rank.}$$

This is because orthogonalizing the columns of  $Z^T$  against  $e$  gives  $Z^T - e\bar{z}^T = (Z - \bar{z}e^T)^T$  and therefore

$$\text{rank}([e, Z^T]) = 1 + \text{rank}(Z - \bar{z}e^T).$$

**DEFINITION 2.2.** *This definition is recursive. Let  $\mathcal{S}_i$ ,  $1 \leq i \leq s$ , be  $s$  subsets of  $\mathbb{R}^d$ . The collection  $\{\mathcal{S}_i, 1 \leq i \leq s\}$  is fully overlapped if it can be partitioned into two nonempty disjoint collections, say,  $\{\mathcal{S}_i, i = 1, \dots, p\}$  and  $\{\mathcal{S}_i, i = p+1, \dots, s\}$ , each of which is a fully overlapped collection, and if the union sets of the two collections  $\hat{\mathcal{S}}_1 \equiv \bigcup_{i=1}^p \mathcal{S}_i$  and  $\hat{\mathcal{S}}_2 \equiv \bigcup_{i=p+1}^s \mathcal{S}_i$  are fully overlapped.*

This definition of full overlapping is rather general. Practically, the situation that the graph<sup>2</sup> of  $\{\mathcal{S}_i, 1 \leq i \leq s\}$  is connected is perhaps more common. That the graph of  $\{\mathcal{S}_i, 1 \leq i \leq s\}$  is connected implies full overlapping.

**DEFINITION 2.3.** *The collection  $\{\mathcal{S}_i, 1 \leq i \leq s\}$  is a covering of  $\mathcal{S}$  if  $\bigcup_{i=1}^s \mathcal{S}_i = \mathcal{S}$ , and a fully overlapped covering if the collection is a covering and fully overlapped.*

We now proceed to prove our main theorem, i.e.  $\text{span}\{[e, T^T]\} = \text{null}\{\Phi\}$ . We do that by showing two way inclusions.

**LEMMA 2.4.** *Let  $\{\mathcal{S}_i, 1 \leq i \leq s\}$  be a covering of  $\mathcal{S}$ , and let  $\Phi_i$  and  $\Phi$  be defined as in (2.3) and (2.4), respectively. Then*

$$(2.7) \quad \text{null}(\Phi_i) = \{x \mid E_i^T x \in \text{span}([e, T_i^T])\}, \quad \text{null}(\Phi) = \bigcap_{i=1}^s \text{null}(\Phi_i),$$

and

$$(2.8) \quad \text{span}\{[e, T^T]\} \subset \text{null}\{\Phi\}.$$

*Proof.* It can be readily verified that  $\{x \mid E_i^T x \in \text{span}([e, T_i^T])\} \subseteq \text{null}(\Phi_i)$ . Now, for any  $x \in \text{null}(\Phi_i)$ ,  $E_i P_i E_i^T x = 0$  which implies  $x^T E_i P_i E_i^T x = 0$ , and thus  $P_i E_i^T x = 0$ , where we note that  $P_i$  is positive semi-definite. So  $E_i^T x \in \text{null}(P_i) = \text{span}([e, T_i^T])$  by (2.2). This proves the first equality in (2.7).

Clearly  $\bigcap_{i=1}^s \text{null}(\Phi_i) \subset \text{null}(\Phi)$ . If  $x \in \text{null}(\Phi)$ ,  $\sum_{i=1}^s x^T \Phi_i x = x^T \Phi x = 0$ . Since  $P_i$  and hence  $\Phi_i$  are positive semi-definite,  $\Phi_i x = 0$  or  $x \in \text{null}(\Phi_i)$  for  $i = 1, \dots, s$ . This proves the second equality in (2.7).

---

<sup>2</sup>By the graph of  $\{\mathcal{S}_i, 1 \leq i \leq s\}$  we mean the graph whose nodes represent the  $s$  subsets  $\mathcal{S}_i$  and there is an edge between node  $\mathcal{S}_i$  and node  $\mathcal{S}_j$  if they are fully overlapped, according to Definition 2.1.

Finally, noting  $\Phi_i [e, T^T] = E_i P_i E_i^T [e, T^T] = E_i P_i [e, T_i^T] = 0$ , we arrive at (2.8).  $\square$

We next show the reverse inclusion first for a collection of two subsets and then generally.

LEMMA 2.5. *Let  $\Phi_i$  and  $\Phi$  be defined as in (2.3) and (2.4), respectively. If  $\{\mathcal{S}_1, \mathcal{S}_2\}$  is a fully overlapped covering of  $\mathcal{S}$ , then  $\text{null}\{\Phi\} = \text{span}\{[e, T^T]\}$ .*

*Proof.* Let  $Q$  be a basis of  $\text{null}(\Phi)$ . By (2.7),

$$\text{span}(Q) \subset \{x \mid E_i^T x \in \text{span}([e, T_i^T])\}$$

which implies that  $E_i^T Q \in \text{span}([e, T_i^T])$ , or equivalently, there is a  $W_i \in \mathbb{R}^{(d+1) \times m}$ , where  $m = \dim \text{null}(\Phi)$ , such that  $E_i^T Q = [e, T_i^T] W_i$ , i.e.,

$$Q(\mathbf{I}_i, :) = [e, T(:, \mathbf{I}_i)^T] W_i.$$

Let  $\mathbf{I}_{1,2} = \mathbf{I}_1 \cap \mathbf{I}_2$ . Then

$$(2.9) \quad [e, T(:, \mathbf{I}_{1,2})^T] (W_1 - W_2) = 0.$$

But  $\mathcal{S}_1$  and  $\mathcal{S}_2$  are *fully overlapped*; so  $[e, T(:, \mathbf{I}_{1,2})^T]$  has full column rank by Definition 2.1 and its equivalent condition (2.6). Thus,  $W_1 = W_2$  by (2.9). Let  $W \equiv W_1 = W_2$ . We then have  $E_i^T Q = [e, T_i^T] W = E_i^T [e, T^T] W$  for  $i = 1, 2$ . Therefore  $Q = [e, T^T] W$  because  $\{\mathcal{S}_i, 1 \leq i \leq 2\}$  is a covering of  $\mathcal{S}$ .  $\square$

THEOREM 2.6. *Let  $\Phi_i$  and  $\Phi$  be defined as in (2.3) and (2.4), respectively, and let  $\{\mathcal{S}_i, i = 1, \dots, s\}$  be a covering of  $\mathcal{S}$ . If it is fully overlapped, then  $\text{null}(\Phi) = \text{span}([e, T^T])$ .*

*Proof.* We prove the theorem by induction in  $s$ . The case  $s = 2$  has been already dealt with in Lemma 2.5. So, suppose the theorem is true for any collection with at most  $s - 1$  subsets. We now prove it for a collection with  $s$  subsets.

Since  $\{\mathcal{S}_i, i = 1, \dots, s\}$  is fully overlapped, let  $\{\mathcal{S}_i, i = 1, \dots, p\}$  and  $\{\mathcal{S}_i, i = p+1, \dots, s\}$  be the partition such that both  $\{\mathcal{S}_i, i = 1, \dots, p\}$  and  $\{\mathcal{S}_i, i = p+1, \dots, s\}$  are fully overlapped and the two sets  $\hat{\mathcal{S}}_1 \equiv \bigcup_{i=1}^p \mathcal{S}_i$  and  $\hat{\mathcal{S}}_2 \equiv \bigcup_{i=p+1}^s \mathcal{S}_i$  are fully overlapped. Denote accordingly the matrices consisting of the columns of  $\hat{\mathcal{S}}_1$  and  $\hat{\mathcal{S}}_2$  by  $\hat{T}_1$  and  $\hat{T}_2$ , respectively. For the collection  $\{\hat{\mathcal{S}}_1, \hat{\mathcal{S}}_2\}$ , let  $\hat{P}_i$ ,  $\hat{\Phi}_i$  and  $\hat{\Phi}$  be the matrices defined according to (2.3) and (2.4). By Theorem 2.5,

$$\text{null}(\hat{\Phi}) = \text{null}(\hat{\Phi}_1) \cap \text{null}(\hat{\Phi}_2) = \text{span}([e, T^T]).$$

Now,  $\{\mathcal{S}_i, i = 1, \dots, p\}$  is a covering collection of  $\hat{\mathcal{S}}_1$ . From this, we can construct  $\tilde{P}_i$ ,  $\tilde{\Phi}_i$  and  $\tilde{\Phi}$  according to (2.3) and (2.4). Since  $\{\mathcal{S}_i, i = 1, \dots, p\}$  is fully overlapped, by the induction assumption, we have  $\text{null}(\tilde{\Phi}) = \text{span}([e, \hat{T}_1^T])$ . Clearly,  $\tilde{P}_i = P_i$ . Since  $\tilde{\Phi}_i$  is the embedding of  $\tilde{P}_i$  into  $\mathbb{R}^k$  ( $k$  is the number of columns in  $\hat{T}_1$  while  $\Phi_i$  is the embedding of  $P_i$  into  $\mathbb{R}^N$ ), the embedding of  $\tilde{\Phi}_i$  into  $\mathbb{R}^N$  is  $\Phi_i$ . Indeed, if  $E \in \mathbb{R}^{N \times k}$  is the selection matrix corresponding to the index set of  $\hat{T}_1$  such that  $TE = \hat{T}_1$ , then

$$(2.10) \quad \Phi_i = E \tilde{\Phi}_i E^T.$$

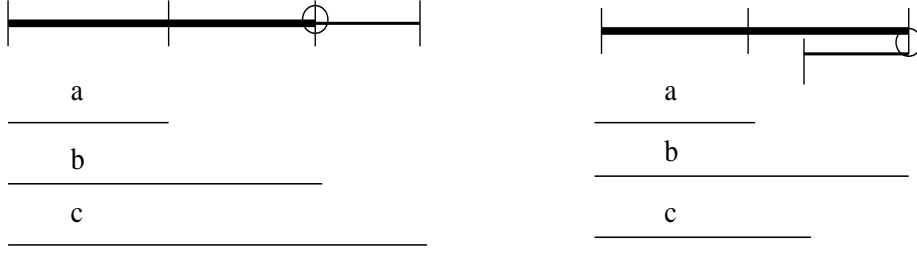


FIG. 2.1. Two possible layouts for the global coordinates.

Noting that  $\text{null}(\tilde{\Phi}) = \text{span}([e, \hat{T}_1^T])$  and  $\text{null}(\hat{P}_1) = \text{span}([e, \hat{T}_1^T])$ , we see

$$\text{null}(E\tilde{\Phi}E^T) = \text{null}(E\hat{P}_1E^T).$$

Thus,

$$\begin{aligned} \text{null}(\Phi_1 + \cdots + \Phi_p) &= \text{null}(E(\tilde{\Phi}_1 + \cdots + \tilde{\Phi}_p)E^T) \\ &= \text{null}(E\tilde{\Phi}E^T) \\ &= \text{null}(E\hat{P}_1E^T) \\ &= \text{null}(\hat{\Phi}_1). \end{aligned}$$

Similarly,

$$\text{null}(\Phi_{p+1} + \cdots + \Phi_s) = \text{null}(\hat{\Phi}_2).$$

Finally,

$$\begin{aligned} \text{null}(\Phi) &= \text{null}(\Phi_1 + \cdots + \Phi_p) \cap \text{null}(\Phi_{p+1} + \cdots + \Phi_s) \\ &= \text{null}(\hat{\Phi}_1) \cap \text{null}(\hat{\Phi}_2) \\ &= \text{span}\{[e, T^T]\}, \end{aligned}$$

as was to be shown.  $\square$

REMARK 2.1. We first give an example with  $d = 1$  to illustrate that the *fully overlapped* assumption is really needed. Let  $u, v, w, x$  be four different real numbers, and  $\mathcal{S} = \{u, v, w, x\}$ . Assume that we have two subsets  $\mathcal{S}_1 = \{u, v, w\}$  and  $\mathcal{S}_2 = \{w, x\}$  that share a single point  $w$ . It is easy to verify that  $[e, T_1^T]$  and  $[e, T_2^T]$  have full (column) rank. We have  $P_1 = pp^T$  with  $p$  orthogonal to  $e$  and  $T_1^T$ , and  $P_2 = 0$ . So  $\Phi = E_1P_1E_1^T$ . It can be verified that

$$\text{a basis for } \text{null}(\Phi) \text{ is: } e, [T_1, 0]^T, [0, 0, 0, 1]^T,$$

while  $\dim \text{span}([e, T^T]) = 2$ , and thus  $\text{null}(\Phi) \supset \text{span}([e, T^T])$  is strict.

This has also an interesting geometric interpretation. The basis vectors for  $\text{null}(\Phi)$ , as illustrated above, can be chosen as  $[1, 1, 1, 1]^T$ ,  $[\eta, \eta + a, \eta + b, \eta + c]^T$  with  $a = \theta_1, b = \theta_1 + \theta_2, c = \theta_1 + \theta_2 + \theta_3$  and  $\eta$  an arbitrary real number and  $[\eta, \eta + a, \eta + b, \eta + c]^T$  with  $a = \theta_1, b = \theta_1 + \theta_2, c = \theta_1 + \theta_2 - \theta_3$  and  $\eta$  an arbitrary real number. Geometrically, the second vector corresponds to the coordinates of the

four points in the left panel of Figure 2.1 and the third vector corresponds to the coordinates of the four points in the right panel of Figure 2.1. While  $\mathcal{S}_1$  fixes the first three points  $\eta, \eta + a, \eta + b$ , the fourth point as fixed by  $\mathcal{S}_2$  can either be extended or folded back from  $\eta + b$ , because  $\mathcal{S}_1$  and  $\mathcal{S}_2$  are not fully overlapped. Thus,  $\mathcal{S}$  can not be uniquely determined from  $\mathcal{S}_1$  and  $\mathcal{S}_2$  if  $\mathcal{S}_1$  and  $\mathcal{S}_2$  are not fully overlapped.

The theorem shows that  $T$  can be reconstructed from the alignment matrix  $\Phi$ . However, the alignment matrix as defined above assumes that we have  $\tau_i$  for each subsets. In practice, we do not have  $\tau_i$  available but we may construct for each subset (local patch) a local coordinate system that is an isometric transformation (or an affine transformation) of  $\tau_i$ 's. Fortunately, with such local coordinate systems, the same alignment matrix is constructed and hence  $T$  recovered. We show this fact in the following theorem but will discuss more details on the use of these results in the next section.

**THEOREM 2.7.** *Let  $\{\mathcal{S}_i, i = 1, \dots, s\}$ , given by (2.1), be a covering of  $\mathcal{S}$ , and let  $\{\theta_1^{(i)}, \dots, \theta_{k_i}^{(i)}\} \subset \mathbb{R}^d$  which is isometric in the Euclidean distance to  $\{\tau_{i_1}, \dots, \tau_{i_{k_i}}\}$ . Set*

$$(2.11) \quad \Theta_i = \left[ \theta_1^{(i)}, \dots, \theta_{k_i}^{(i)} \right].$$

and define  $Q_i$  to be the orthogonal projection onto the orthogonal complement of  $\text{span}\{[e, \Theta_i^T]\}$ , and

$$(2.12) \quad \Psi_i = E_i Q_i E_i^T, \quad \Psi = \sum_{i=1}^s \Psi_i.$$

Then

$$\Psi = \Phi.$$

In particular,  $\text{span}\{[e, T^T]\} \subset \text{null}(\Psi)$  and, if  $\{\mathcal{S}_i, i = 1, \dots, s\}$  is fully overlapped, then  $\text{null}\{\Psi\} = \text{span}\{[e, T^T]\}$ .

*Proof.* Since  $\{\theta_1^{(i)}, \dots, \theta_{k_i}^{(i)}\}$  is isometric to  $\{\tau_{i_1}, \dots, \tau_{i_{k_i}}\}$ ,  $\{\theta_1^{(i)} - \bar{\theta}_i, \dots, \theta_{k_i}^{(i)} - \bar{\theta}_i\}$  is also isometric to  $\{\tau_{i_1} - \bar{\tau}_i, \dots, \tau_{i_{k_i}} - \bar{\tau}_i\}$ , where

$$\bar{\theta}_i = \frac{1}{k_i} \sum_{j=1}^{k_i} \theta_j^{(i)}, \quad \bar{\tau}_i = \frac{1}{k_i} \sum_{j=1}^{k_i} \tau_{i_j}.$$

Then the distance matrix of  $\{\theta_1^{(i)} - \bar{\theta}_i, \dots, \theta_{k_i}^{(i)} - \bar{\theta}_i\}$  and that of  $\{\tau_{i_1} - \bar{\tau}_i, \dots, \tau_{i_{k_i}} - \bar{\tau}_i\}$  are the same, implying  $\hat{\Theta}_i^T \hat{\Theta}_i = \hat{T}_i^T \hat{T}_i$ , where

$$\hat{\Theta}_i = \left[ \theta_1^{(i)} - \bar{\theta}_i, \dots, \theta_{k_i}^{(i)} - \bar{\theta}_i \right], \quad \hat{T}_i = \left[ \tau_{i_1} - \bar{\tau}_i, \dots, \tau_{i_{k_i}} - \bar{\tau}_i \right].$$

Therefore,  $\hat{\Theta}_i$  and  $\hat{T}_i$  have the same singular values and the same right singular vectors. Form the singular value decompositions of  $\hat{\Theta}_i$  and  $\hat{T}_i$ , we deduce that there is an orthogonal matrix  $V_i$  such that

$$(2.13) \quad \hat{\Theta}_i = V_i \hat{T}_i.$$

Thus,

$$(2.14) \quad \text{span}([e, \Theta_i^T]) = \text{span}([e, T_i^T])$$

and hence  $Q_i = P_i$  defined earlier. Therefore,  $\Psi = \Phi$ . Now the theorem follows from Theorem 2.5.  $\square$

**3. Reconstruction of Global Coordinates.** Theorem 2.7 implies that if the underlying manifold is *locally* isometric to a parametric space and for each local set  $\mathcal{S}_i$  (or patch), an isometric coordinate system can be computed, then a global isometric coordinate is recovered by calculating  $\text{null}(\Psi)$ . Suppose

$$g : \Omega \subset \mathbb{R}^d \rightarrow \mathcal{M} \subset \mathbb{R}^n$$

be an isometry mapping, and let  $\mathbf{X} = \{x_1, \dots, x_N\} \subset \mathcal{M}$  consisting of sample points on  $\mathcal{M}$ , where  $x_i = g(\tau_i)$  ( $1 \leq i \leq N$ ). Consider a covering  $\{\mathbf{X}_i, i = 1, \dots, s\}$  of  $\mathbf{X}$ , where  $\mathbf{X}_i = \{x_{i_1}, \dots, x_{i_{k_i}}\}$ . Set accordingly  $\mathcal{S}_i = \{\tau_{i_1}, \dots, \tau_{i_{k_i}}\}$  and  $\mathcal{S} = \{\tau_1, \dots, \tau_N\}$ . We say  $\{\mathbf{X}_i, i = 1, \dots, s\}$  is a fully overlapped covering of  $\mathbf{X}$  if  $\{\mathcal{S}_i, i = 1, \dots, s\}$  is a fully overlapped covering of  $\mathcal{S}$ .

Suppose that all points in each subset  $\mathbf{X}_i$  are ‘‘sufficiently local’’ such that

$$d(\tau_{i_p}, \tau_{i_q}) = d_{\mathcal{M}}(x_{i_p}, x_{i_q}), \quad \text{for } 1 \leq p, q \leq k_i,$$

where  $d(\cdot, \cdot)$  is the Euclidean distance and  $d_{\mathcal{M}}(\cdot, \cdot)$  is the geodesic distance along  $\mathcal{M}$ . Let  $\theta_1^{(i)}, \dots, \theta_{k_i}^{(i)} \in \mathbb{R}^d$  be some locally isometric coordinates constructed for the corresponding points in  $\mathbf{X}_i$ , i.e.,

$$d(\theta_p^{(i)}, \theta_q^{(i)}) = d_{\mathcal{M}}(x_{i_p}, x_{i_q}), \quad \text{for } 1 \leq p, q \leq k_i.$$

Now, construct  $\Psi$  by (2.12) from the  $\theta_j^{(i)}$ ’s and Theorem 2.7 applies. From  $\text{null}\{\Psi\} = \text{span}\{[e, T^T]\}$ , the global coordinates can be reconstructed with a post normalization step. We propose the following general procedure for a given input of data.

**ALGORITHM 3.1. Reconstruction of Global Coordinates:**

Given  $\mathbf{X} = \{x_1, \dots, x_N\} \subset \mathbb{R}^n$ .

1. Construct a fully overlapped covering  $\{\mathbf{X}_i, i = 1, \dots, s\}$  with  $\mathbf{X}_i = \{x_{i_1}, \dots, x_{i_{k_i}}\}$ .
2. For each  $\mathbf{X}_i$ , construct centered local coordinates  $\theta_1^{(i)}, \dots, \theta_{k_i}^{(i)} \subset \mathbb{R}^d$ . This can be done by the projection into the tangent space as in LTSA [14] if the points in  $\mathbf{X}_i$  are confined to a small neighborhood or by the ISOMAP algorithm [12] for a more general  $\mathbf{X}_i$ . Also, determine a patch  $\mathbf{X}_p$  such that its local coordinates are constructed with least errors.
3. Construct  $\Psi$  from  $\Theta_i = [\theta_1^{(i)}, \dots, \theta_{k_i}^{(i)}]$  as in (2.12) and compute an orthonormal basis  $[e/\sqrt{N}, Z^T]$  for the eigenspace of  $\Psi$  corresponding to the  $d + 1$  smallest eigenvalue (or the  $d + 1$  smallest eigenvectors) where  $Z^T \in \mathbb{R}^{N \times d}$ .
4.  $T = WZ$  is the global coordinate, where  $W = \Theta_p \hat{Z}_p^+$  and  $\hat{Z}_p = ZE_i(I - \frac{1}{k_p} ee^T)$ .

We give some remarks to explain each of the steps in the algorithm.

A covering  $\{\mathbf{X}_i, i = 1, \dots, s\}$  is constructed through partitioning of  $\mathbf{X}$ , typically into local patches. That  $\{\mathbf{X}_i, i = 1, \dots, s\}$  is fully overlapped at Step 1 is not verifiable in general because the dimension  $d$  in Definition 2.1 may not be known, but in practice, the condition (2.5) can generally be assumed to be true as long as there is a sufficient number (i.e. greater than  $d$ ) of intersection points. We also note that  $d$  is often a parameter that we choose as the dimension of the coordinate systems for representing the given data.

Step 2 has been discussed earlier [11, 12, 14]. Briefly, if local patches  $\mathbf{X}_i$  consists of points in a small neighborhood as in LTSA [14], a local coordinate system can be constructed from an orthonormal basis of a  $d$ -dimensional subspace that best



approximates  $\text{span}\{x_{i_1} - \bar{x}_i, \dots, x_{i_{k_i}} - \bar{x}_i\}$  where  $\bar{x}_i = \frac{1}{k_i} \sum_{j=1}^{k_i} x_{i_j}$ . This is computed from the singular value decomposition of  $[x_{i_1} - \bar{x}_i, \dots, x_{i_{k_i}} - \bar{x}_i]$ . Then up to an error bounded by the  $(d+1)$ -st singular value  $\sigma_{d+1}$ , the constructed coordinates  $\{\theta_1^{(i)}, \dots, \theta_{k_i}^{(i)}\}$  are approximately isometric to  $\{x_{i_1}, \dots, x_{i_{k_i}}\}$  in Euclidean distance. Using  $\sigma_{d+1}/\sigma_1$  as a measure of this approximation error, we can identify the patch  $\mathbf{X}_p$  that is best represented by the local coordinates computed.  $\mathbf{X}_p$  is used in step 4 for post normalization.

We further note that, using the first-order Taylor expansion, we have

$$g(\tilde{\tau}) = g(\tau) + J_g(\tau) \cdot (\tilde{\tau} - \tau) + O(\|\tilde{\tau} - \tau\|^2)$$

where  $J_g(\tau) \in \mathbb{R}^{m \times d}$  is the Jacobi matrix. Since we assume that  $g(\cdot)$  is a local isometry, then  $J_g(\tau)$  has orthonormal columns for each  $\tau$ . Applying this to the points in  $\mathbf{X}_i$  and ignoring the second order terms, we have that  $\{\tau_{i_1}, \dots, \tau_{i_{k_i}}\}$  is approximately isometric to  $\{x_{i_1}, \dots, x_{i_{k_i}}\}$  in Euclidean distance and hence approximately isometric to the computed local coordinates  $\{\theta_1^{(i)}, \dots, \theta_{k_i}^{(i)}\}$ .

For step 3, an approximate null space of  $\Psi$  is obtained from the  $d+1$ -smallest eigenvectors, which can be computed by the Lanczos or block Lanczos algorithm combined with shift-and-invert [1], its implicitly restarted version [9], or if the inverse is difficult, by the inverse-free preconditioned Krylov subspace method [10]. The known eigenvector  $e$  should be used as an initial vector. The sparse matrix  $\Psi$  can be assembled from its local elements, see [14].

We now discuss the recovery of  $T$  at step 4 through a post normalization. Assume that  $T$  is centered, i.e.,  $Te = 0$ . By Theorem 2.7,  $\text{null}(\Psi) = \text{span}[e, T^T]$  and therefore  $\text{span}[e, Z^T] = \text{span}[e, T^T]$ . However, we do not necessarily have  $Z = T$ . Nevertheless, there is some nonsingular  $U \in \mathbb{R}^{d \times d}$  such that  $T^T = Z^T U^T$ . To determine  $U$ , we consider one section  $T_i = TE_i = [\tau_{i_1}, \dots, \tau_{i_{k_i}}]$  of  $T$  and realign the corresponding section of  $Z$  with  $T_i$  through the computed local coordinates  $\Theta_i$ . Let  $\hat{T}_i = T_i \left(I - \frac{1}{k_i} ee^T\right)$  be the centered  $T_i$ . If  $\{\theta_j^{(i)}\}$  are exactly isometric to  $\{\tau_{i_j}\}$ , then their centered coordinates are related by an orthogonal transformation, i.e. there exists an orthogonal matrix  $V_i$  such that  $V_i \hat{T}_i = \Theta_i$ , see (2.13). Let  $W = V_i U$ . We have

$$(3.1) \quad \Theta_i = V_i \hat{T}_i = V_i U Z E_i \left(I - \frac{1}{k_i} ee^T\right) = V_i U Z_i \left(I - \frac{1}{k_i} ee^T\right) = W \hat{Z}_i.$$

where  $Z_i = Z E_i$  is the section of  $Z$  corresponding to  $T_i$  as a section of  $T$  and  $\hat{Z}_i = Z_i \left(I - \frac{1}{k_i} ee^T\right)$  is the column-centered  $Z_i$ . Hence, using (3.1), we can determine  $W$  from the computed local coordinates  $\Theta_i$  and a corresponding section of  $Z$ . From  $W$ , we have  $WZ = V_i UZ = V_i T$ . With  $V_i$  being an orthogonal matrix,  $WZ$  determines  $T$  modulo a rigid motion. Therefore, we recover  $T$  by  $WZ$ .

However, (3.1) is based on the assumption that the local coordinates  $\{\theta_j^{(i)}\}$  constructed are exactly isometric to global coordinates  $\{\tau_{i_j}\}$ . In practice, the local coordinates constructed are only approximately isometric to  $\{\tau_{i_j}\}$  (see discussions about Step 2 above). To obtain best realignment, we therefore need to use a patch  $\mathbf{X}_p$ , for which the local approximation error is the least and then find some  $W$  by solving

$$(3.2) \quad \min_W \|\Theta_p - W \hat{Z}_p\|_F$$

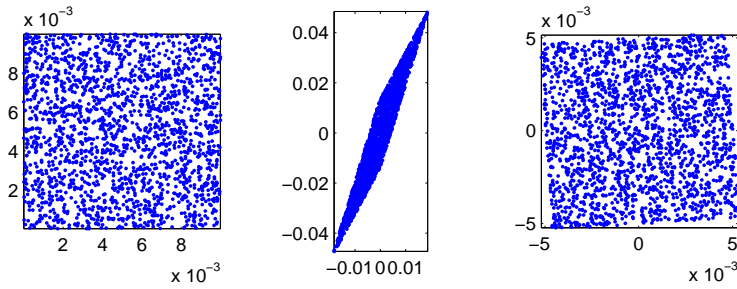


FIG. 3.1. (Left) Plot of the original 2D coordinates; (middle) plot of the coordinates recovered by LTSA; (right) plot of the coordinates recovered by LTSA with normalization.

which gives  $W = \Theta_p \widehat{Z}_p^+$ , where  $\widehat{Z}_p^+$  is the Moore-Penrose pseudo inverse of  $\widehat{Z}_p$ .

In [14],  $Z$  with orthonormal rows are directly used to approximately recover  $T$  but it is only approximately a linear transformation of  $T$ . In light of the discussion above, however,  $Z$  need to be normalized by the matrix  $W$  to recover a coordinate system that is an orthogonal transformation of the original one. The effectiveness of this will be illustrated in the following numerical example. As a by-product, we have shown that LTSA with the post normalization step 4 can recover locally isometric embedding up to a rigid motion.

We present two numerical examples of local isometric maps to demonstrate benefits of the post-normalization. In both examples, local patches are constructed from the  $k$ -nearest neighborhood of each point. In constructing local coordinates by SVD, we use  $\sigma_{d+1}/\sigma_1$  to determine the patch that has the least local errors, which is used for post normalization.

EXAMPLE 1. We consider 2000 random sample points on a 2-d parametric surface in  $\mathbb{R}^3$  defined by

$$x_i = [\cos(s_i), t_i, \sin(s_i)]^T$$

where  $s_i$  and  $t_i$  are uniformly distributed random numbers between 0 and 0.01. We use LTSA and LTSA with the normalization (step 4) to determine a parametrization of these points. We have tested with several values of  $k$  in the range  $7 \leq k \leq 30$  and the results are similar. We present the one with  $k = 15$  in Figure 3.1. On the left panel is the original coordinates  $(s_i, t_i)$ , in the middle is the coordinates found by LTSA, and on the right is the coordinates found by LTSA with the normalization. Clearly, without the normalization, LTSA obtains only an affine transformation of the original coordinates but with normalization, it recovers the square (with a rotation) as well as the original scale.

EXAMPLE 2. We consider 2000 random sample points on a 2-d parametric surface in  $\mathbb{R}^4$  defined by

$$x_i = [\cos(s_i) \sin(t_i), \sin(s_i) \sin(t_i), \sin(s_i) \cos(t_i), \cos(s_i) \cos(t_i)]^T$$

where  $(s_i, t_i)$  are random points on the unit half disk generated by

$$(s_i, t_i) = r_i(\cos(\theta_i\pi), \sin(\theta_i\pi))$$

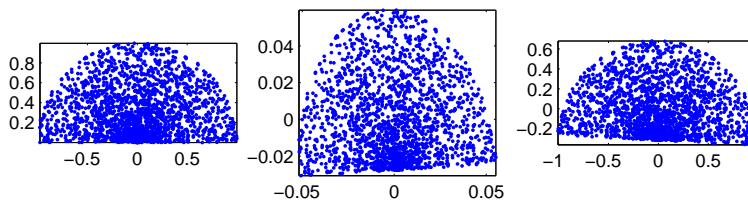


FIG. 3.2. (Left) Plot of the original 2D coordinates; (middle) plot of the coordinates recovered by LTSA; (right) plot of the coordinates recovered by LTSA with normalization.

with  $r_i$  and  $\theta_i$  uniformly distributed random numbers between 0 and 1. We use LTSA and LTSA with the post normalization (step 4) to determine a parametrization of these points. We have tested with several values of  $k$  in the range  $7 \leq k \leq 30$  and the results are similar. We present the one with  $k = 15$  in Figure 3.2. On the left panel is the original coordinates  $(s_i, t_i)$ , in the middle is the coordinates found by LTSA, and on the right is the coordinates found by LTSA with the normalization. Again, without the normalization, LTSA obtains only an elongated half disk (an affine transformation) and with normalization, we recover the half disk (with a rotation) as well as the original scale.

**4. Concluding Remarks.** To better understand the performance of manifold learning algorithms such as LTSA, it is important to gain deeper insights into the alignment process. In this paper we have analyzed the alignment process through characterizing the null space structure of the alignment matrix and its relationship with the geometric properties of the set of sample points. Several topics will be investigated in future research including (1) a quantitative characterization of the spectral properties of the alignment matrix; and (2) an analysis of the reconstruction error with respect to the error in the construction of local coordinates as well as in the data itself (i.e. noisy data). Matrix analysis techniques such as matrix perturbation theory will certainly play important role in the understanding of the performance of manifold learning algorithms which are at the current forefront of machine learning research.

#### REFERENCES

- [1] Z. Bai, J. Demmel, J. Dongarra, A. Ruhe and H. van der Vorst, editors, *Templates for the solution of Algebraic Eigenvalue Problems: A Practical Guide*. SIAM, Philadelphia, 2000.
- [2] M. BRAND, Charting a manifold, *Advances in Neural Information Processing Systems* 15, MIT Press, 2003.
- [3] J. DEMMEL, *Applied Numerical Linear Algebra*, SIAM, Philadelphia, 1997.
- [4] D. DONOHO AND C. GRIMES, Hessian Eigenmaps: new tools for nonlinear dimensionality reduction, *Proceedings of National Academy of Science*, 5591-5596, 2003.
- [5] Q. DU AND X. WANG, Tessellation and Clustering by Mixture Models and Their Parallel Implementations, *Proceedings of the fourth SIAM international conference on Data Mining*, 257-268, 2004.
- [6] G. H. GOLUB AND C. F. VAN LOAN, *Matrix Computations*, Johns Hopkins University Press, Baltimore, Maryland, 3rd ed., 1996.
- [7] T. HASTIE, R. TIBSHIRANI AND J. H. FRIEDMAN, *The Elements of Statistical Learning*, Springer, Berlin, 2001.

- [8] A. K. JAIN AND M. N. MURTY AND P. J. FLYNN, *Data Clustering: A Review*, ACM Computing Surveys, 31, (1999), pp. 264–323.
- [9] LEHOUCQ, R., SORENSON, D., AND YANG, C. *ARPACK Users' Guides, Solution of Large Scale Eigenvalue Problems with Implicitly Restarted Arnoldi Method*, SIAM, Philadelphia, 1998.
- [10] J. MONEY AND Q. YE, *Algorithm 845: EIGFIP: A MATLAB Program for Solving Large Symmetric Generalized Eigenvalue Problems*, ACM Trans. Math. Softwares, 31(2005):270-279.
- [11] S. T. ROWEIS AND L. K. SAUL, *Nonlinear dimensionality reduction by locally linear embedding*, Science, 290 (2000), pp. 2323–2326.
- [12] J. B. TENENBAUM, V. DE SILVA, AND J. C. LANGFORD, *A global geometric framework for nonlinear dimensionality reduction*, Science, 290 (2000), pp. 2319–2323.
- [13] Y. TEH AND S. ROWEIS, *Automatic Alignment of Hidden Representations*, *Advances in Neural Information Processing Systems*, 15, MIT Press, 2003.
- [14] Z. ZHANG AND H. ZHA, *Principal manifolds and nonlinear dimension reduction via local tangent space alignment*, SIAM Journal of Scientific Computing, 26 (2004), pp. 313–338.