SPECTRAL PROPERTIES OF THE ALIGNMENT MATRICES IN MANIFOLD LEARNING

HONGYUAN ZHA* AND ZHENYUE ZHANG†

Abstract. Local methods for manifold learning generate a collection of local parameterizations which is then aligned to produce a global parameterization of the underlying manifold. The alignment procedure is carried out through the computation of a partial eigendecomposition of a so-called alignment matrix. In this paper, we present an analysis of the eigen-structure of the alignment matrix giving both necessary and sufficient conditions under which the null space of the alignment matrix recovers the global parameterization. We show that the gap in the spectrum of the alignment matrix is proportional to the square of the size[1] of the overlap of the local parameterizations thus deriving a quantitative measure of how stably the null space can be computed numerically. We also give a perturbation analysis of the null space of the alignment matrix when the computation of the local parameterizations is subject to error. Our analysis provides insights into the behaviors and performance of local manifold learning algorithms.

1. Introduction. Consider the following unsupervised learning problem: we are given a *parameterized manifold* $\mathcal{M}$ of dimension d embedded into the m-dimensional Euclidean space $\mathcal{R}^m$, $d < m$, and $\mathcal{M} = f(\Omega)$ with a mapping $f : \Omega \to \mathcal{R}^m$, where Ω is open in $\mathcal{R}^d$ [9, section 5.22]; suppose we have a set of points $x_1, \cdots, x_N$, sampled possibly with noise from the manifold $\mathcal{M}$,

$$x_i = f(\tau_i) + \epsilon_i, \quad i = 1, \ldots, N, \tag{1.1}$$

where the $\{\epsilon_i\}$ represent noise; we are interested in recovering the $\{\tau_i\}$ and/or the mapping $f(\cdot)$ from the noisy data $\{x_i\}$. This problem is generally known as manifold learning or nonlinear dimension reduction, and has generated much research interest in the machine learning and statistics communities [10, 12]. A class of *local* methods for manifold learning starts with estimating a collection of local structures around each sample point x_i and then aligns (either implicitly or explicitly) those local structures to obtain estimates for $\{\tau_i\}$ by computing a partial eigendecomposition of an alignment matrix. Examples of local methods include LLE (Locally Linear Embedding) [10], manifold charting [2], geodesic null space analysis [3], Hessian LLE [5], LTSA (Local Tangent Space Alignment) [17], and the modified LLE (MLLE) [13]. Those methods have been applied to analyzing high-dimensional data arising from application areas such as computer vision, speech analysis as well as molecular dynamics simulations.

In contrast to the ever-increasing use of manifold learning methods and the frequent appearance of new algorithms, little has been done to assess the performance of those methods, even though manifold learning methods in general tend to be rather sensitive to the selection of several tuning parameters [3, 4, 17]. Usually one applies a manifold learning algorithm to a high-dimensional data set, sometimes one recovers satisfactory parameter vectors, sometimes one obtains catastrophic folds in the computed parameterization and one needs to tune the parameters and try again. This

*Division of Computational Science and Engineering, College of Computing, Georgia Institute of Technology, Atlanta, GA, 30332-0280, zha@cc.gatech.edu. The work of this author was supported in part by NSF grants CCF-0430349 and DMS-0736328. A preliminary version of a subset of the results reported in this paper were published without proof in [16].

† Department of Mathematics, Zhejiang University, Yuquan Campus, Hangzhou, 310027, P. R. China. zyzhang@zju.edu.cn. Corresponding author. The work of this author was supported in part by NSFC (project 10771194) and NSF grant CCF-0305879.

[1]The precise definition of the size of the overlap is given in section 6.

is a rather unsatisfactory situation which calls for more research into the robustness issues and the performance issues of manifold learning algorithms.

One source of the catastrophic folds in the computed parameterization is the variety of errors involved in manifold learning including the noise in the data, the approximation errors in learning the local structures, and the numerical errors in computing the eigenspace of the alignment matrix. It is not surprising that these errors will degrade the accuracy of the computed parameter vectors. However, in addition to those issues, there is another important question that has been largely ignored in the past: assuming, in the ideal noise-free case, that the local structures are exactly known and the eigenvector space is exactly computed, will the local manifold learning algorithms produce the true parameter vectors? The answer may actually be negative: it very much depends on how the local structures overlap with one another. If we cannot obtain the true parameter vectors in the noise-free case with all the computations done without error, then we cannot expect to do something reasonable when the data as well as the computations are subject to error.

The objective of this paper is to gain a better understanding of the key alignment procedure used in local manifold learning methods by analyzing the eigen-structure of the alignment matrix. We focus on the representative alignment matrix used in LTSA, and we address two questions in particular: 1) under what conditions wa can recover the parameter vectors $\{\tau_i\}$ from the null space of the alignment matrix if the alignment matrix is computed exactly, and 2) how stable this null space is if the computation of the alignment matrix is subject to error. We motivate the importance of addressing the two problems using a simple example in section 3 after a brief review of LTSA in section 2. We then approach the two problems as follows: in section 4, we address the issue of how errors in computing the local parameterizations will affect the null space of the alignment matrix. This allows us to focus on the spectral properties of the ideal alignment matrices and separate the local error issues from the rest of the discussions; section 5 is the main part of the paper, where we propose the concept of *affine rigidity* to precisely address the first question above. We then establish a variety of conditions to characterize when an alignment matrix is affinely rigid. Along the way we also prove some properties of the alignment matrix that will have computational significance; in section 6, we address the second question by proving a lower bound for the smallest nonzero eigenvalue of the alignment matrix.

REMARK. Though only the alignment matrices of LTSA are discussed in detail, we believe that similar approaches can be applied to the analysis of other local methods such as LLE, Hessian LLE, or even Laplacian eigenmap [1]. (See Appendix A for a brief discussion of the alignment matrices used in LLE and Laplacian eigenmap.)

NOTATION. We use e to denote a column vector of all 1's the dimension of which should be clear from the context. $\mathcal{N}(\cdot)$ and $\text{span}(\cdot)$ denote the null space and the range space of a matrix, respectively. For an index set $I_i = \{i_1, \ldots, i_k\}$, $A(:, I_i)$ denotes the submatrix of A consisting of columns of A with indices in I_i. A similar definition $A(I_i, :)$ is for the rows. We also represent the submatrix consisting of vectors x_j, $j \in I_i$ by $X_i = [\ldots, x_j, \ldots]$ with $j \in I_i$ (in the increasing order of the index j). For a set of submatrices $T_i = T(:, I_i)$, $i \in J_j$, we denote by T_{J_j} the submatrix $T(:, \cup_{i \in J_j} I_i)$. $\|\cdot\|_2$ is the spectrum norm and $\|\cdot\|_F$ is the Frobenius norm of a matrix. The superscript T denotes matrix transpose. $A^\dagger$ denotes the Moore-Penrose generalized inverse of A. For an Hermitian matrix A of order n, $\lambda_1(A) \le \cdots \le \lambda_n(A)$ denote the eigenvalues of A in nondecreasing order. The identity matrix is denoted by I or $I^{(d)}$ if its order d is indicated. Finally, $\lambda^+_{\min}(A)$ denotes the smallest nonzero eigenvalue of a positive

semi-definite matrix.

2. Alignment Matrices of LTSA. We first review how the LTSA alignment matrices are constructed [17]. For a given set of sample points $\{x_i\}$, we begin with building a connectivity graph on top of those sample points which specifies, for each sample point, which subset of the sample points constitutes its neighborhood [10]. Let the set of neighbors for the sample point x_i be $X_i = [x_{i_1}, \ldots, x_{i_{k_i}}]$, including x_i itself. We approximate those neighbors using a d-dimensional (affine) linear subspace,

$$x_{i_j} \approx \bar{x}_i + Q_i \theta_j^{(i)}, \quad Q_i = [q_1^{(i)}, \ldots, q_d^{(i)}], \qquad j = 1, \ldots, k_i.$$

Here d is the dimension of the manifold,[2] $\bar{x}_i \in \mathcal{R}^m$, $Q_i \in \mathcal{R}^{m \times d}$ is orthonormal, and $\theta_j^{(i)} \in \mathcal{R}^d$ are the local coordinates of x_{i_j}'s associated with the basis matrix Q_j. The optimal least-square-fitting is determined by solving the following problem

$$\min_{c, Q, \{\theta_j\}: Q^T Q = I^{(d)}} \sum_{j=1}^{k_i} \|x_{i_j} - (c + Q\theta_j)\|_2^2.$$

That is, $\bar{x}_i$ is the mean of the x_{i_j}'s and $\theta_j^{(i)} = Q_i^T (x_j - \bar{x}_i)$. Using the singular value decomposition of the centered matrix $X_i - \bar{x}_i e^T$, Q_i can be computed as the matrix of the right singular vectors corresponding to the d largest singular values of $X_i - \bar{x}_i e^T$ [6]. We postulate that in each neighborhood, the corresponding global parameter vectors $T_i = [\tau_{i_1}, \ldots, \tau_{i_{k_i}}]$ of $T \in \mathcal{R}^{d \times N}$ differ from the local ones $\Theta_i = [\theta_1^{(i)}, \ldots, \theta_{k_i}^{(i)}]$ by a local affine transformation. The errors of the optimal affine transformation are then given by

$$\min_{c_i, L_i} \sum_{j=1}^{k_i} \|\tau_{i_j} - (c_i + L_i \theta_j^{(i)})\|_2^2 = \min_{c_i, L_i} \|T_i - (c_i e^T + L_i \Theta_i)\|_F^2 = \|T_i \widetilde{\Phi}_i\|_F^2, \tag{2.1}$$

where $\widetilde{\Phi}_i$ is the orthogonal projection whose null space is spanned by the columns of $[e, \Theta_i^T]$.[3] Note that if Θ_i is affinely equal to T_i, i.e., $\mathrm{span}([e, \Theta_i^T]) = \mathrm{span}([e, T_i^T])$, then $T_i \widetilde{\Phi}_i = 0$. In general, $T_i \widetilde{\Phi}_i \neq 0$, and we seek to compute the parameter vectors $\{\tau_i\}$ by minimizing the following objective function,

$$\sum_i^N \Big(\min_{c_i, L_i} \sum_{j=1}^{k_i} \|\tau_{i_j} - (c_i + L_i \theta_j^{(i)})\|_2^2 \Big) = \sum_i^N \|T_i \widetilde{\Phi}_i\|_F^2 = \mathrm{tr}(T \widetilde{\Phi} T^T) \tag{2.2}$$

over $T = [\tau_1, \ldots, \tau_N]$. Here

$$\widetilde{\Phi} = \sum_{i=1}^N S_i \widetilde{\Phi}_i S_i^T \tag{2.3}$$

is the *alignment matrix* with $S_i \in R^{N \times k_i}$, the 0-1 *selection matrix*, such that $T_i = T S_i$. Imposing certain normalization conditions on T such as $TT^T = I^{(d)}$ and $Te = 0$, the corresponding optimization problem,

$$\min_{TT^T = I^{(d)},\ Te = 0} \mathrm{tr}(T \widetilde{\Phi} T^T) \tag{2.4}$$

[2] We assume d is known which can be estimated using a number of of existing methods [8, 15].

[3] $\widetilde{\Phi}_i$ can be represented as $\widetilde{\Phi}_i = I - [e, \Theta_i^T][e, \Theta_i^T]^\dagger \in R^{k_i \times k_i}$.

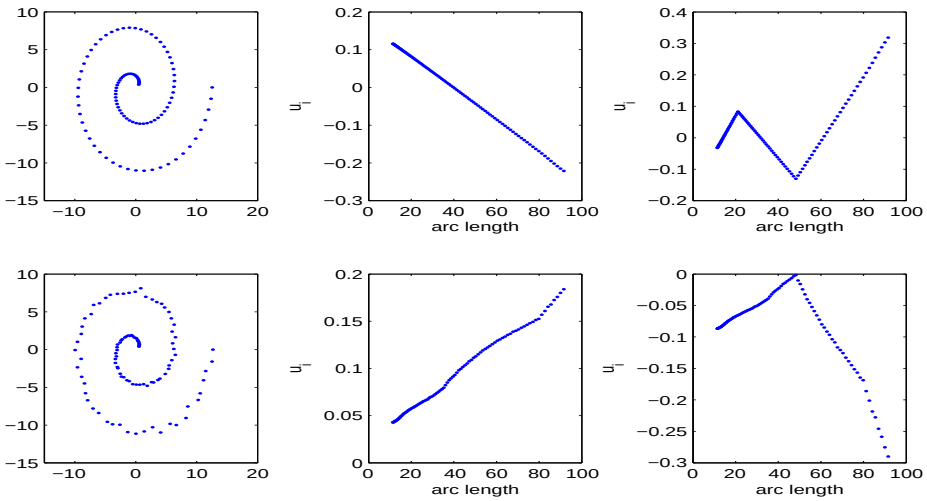

FIG. 3.1. *The spiral data points (left column) and the second eigenvector u of $\widetilde{\Phi}$ when $\widetilde{\Phi}$ has two small eigenvalues (middle column) or more than two small eigenvalues (right column) close to zero.*

is solved by computing the eigenvectors corresponding to $\lambda_2, \cdots, \lambda_{d+1}$ of $\widetilde{\Phi}$, here the eigenvalues are arranged in nondecreasing order, i.e., $\lambda_1 = 0 \leq \lambda_2 \leq \cdots \leq \lambda_{d+1} \leq \cdots \leq \lambda_N$. We remark that if T is a solution to the problem (2.4), then QT is also a solution for any orthogonal matrix Q of order d.

3. An Illustrative Example. In the ideal case when we have $\mathrm{span}([e, \Theta_i^T]) = \mathrm{span}([e, T_i^T])$, it is not difficult to see that T^T belongs to the null space of the alignment matrix (cf. section 4), and it looks like we can just use the (approximate) null space of the alignment matrix to compute the parameter vectors as suggested before. Unfortunately, the null space may contain unwanted information in addition to $\mathrm{span}([e, T^T])$, depending on how the neighborhoods overlap with each other. In the following, we present a simple example to illustrate this phenomenon.

In what follows, we call each X_i (or the corresponding T_i) a *section*. Our analysis is general enough that the X_i can correspond to an arbitrary subset of the sample points. So henceforth, the sample points $x_1, \ldots, x_N$ are grouped into s (possibly overlapping) sections $X_1, \ldots, X_s$, and the i-th section X_i is denoted by the points $\{x_j \,|\, j \in I_i\}$ with the index subset $I_i \subseteq \{1, \ldots, N\}$.

EXAMPLE 1. We generate $N = 100$ two-dimensional points

$$x_i = [t_i \cos(t_i),\, t_i \sin(t_i)]^T, \quad i = 1, \ldots, 100,$$

sampled from the one-dimensional spiral curve with $t_1, \ldots, t_N$ equally spaced in the interval $[\pi/5, 2\pi]$ with $t_0 = \pi/5$ and $t_N = 2\pi$. See the upper-left panel of Figure 3.1 for the set of the two-dimensional sample points. It is well known that a regular smooth curve is isometric to its arc length. The exact arc length coordinate τ_i for the sample point x_i on the spiral curve is given by $\tau_i = \displaystyle\int_{t_0}^{t_i} \sqrt{1 + t^2}\, dt$.

First we choose 19 sections $X_i = X(:, I_i)$, $i = 1, \ldots, 19$, with the index subsets,

$$I_i = (5(i-1)+1) : (5i+2), \quad i = 1, \ldots, 18, \quad I_{19} = 91 : 100.$$

Thus, each pair of two consecutive sections share exactly *two* points. We construct the alignment matrix $\widetilde{\Phi}$ as defined by (2.3): initially set $\widetilde{\Phi} = 0$ and update its principle submatrix $\widetilde{\Phi}(I_i, I_i)$ one-by-one as follows,

$$\widetilde{\Phi}(I_i, I_i) := \widetilde{\Phi}(I_i, I_i) + \widetilde{\Phi}_i, \quad i = 1, \cdots, 19.$$

The orthogonal projection $\widetilde{\Phi}_i$ is given by $\widetilde{\Phi}_i = I - P_i P_i^T$ with $P_i = [\frac{1}{\sqrt{k_i}} e, v_i]$. Here v_i is the eigenvector of $(X_i - \bar{x}_i e^T)^T (X_i - \bar{x}_i e^T)$ corresponding its largest eigenvalue. The resulting alignment matrix $\widetilde{\Phi}$ has two smallest eigenvalues 10^{-16} in magnitude and the third smallest eigenvalue is about 10^{-5}, distinguishable from the two smallest eigenvalues. The solution of problem (2.4) with $d = 1$ is given by the eigenvector $u = [u_1, \ldots, u_N]^T$ of $\widetilde{\Phi}$ corresponding to the second smallest eigenvalue, that is an affine approximation of the arc length coordinates of the sample points. Ideally, u is affinely equivalent to the arc length τ, i.e., there are $a \neq 0$ and b such that $u_i = a\tau_i + b$ for all i. In the middle panel of the top row of Figure 3.1, we plot the computed $\{u_i\}$ against the arc length coordinates $\{\tau_i\}$. The plotted points are approximately on a straight line, indicating an accurate recovery of $\{\tau_i\}$ within an affine transformation.

If the minimal number of the shared points among some of the consecutive sections is reduced to one, $\widetilde{\Phi}$ may have more than two small eigenvalues close to zero. The corresponding eigenvectors contain not only e and $\tau = [\tau_1, \ldots, \tau_N]^T$ but also other unwanted vectors. For example, if we delete the last columns in the two sections X_6 and X_{13}, respectively, then the two consecutive sections X_6 and X_7 share *one* point only. So do X_{13} and X_{14}. This weakens the overlap between X_6 and X_7, as well as that between X_{13} and X_{14}. As a result, $\widetilde{\Phi}$ has four eigenvalues close to zero (there are four computed smallest eigenvalues of magnitude 10^{-16}) and four linearly independent eigenvectors which include e, τ and two other vectors. Since the computed eigenvector of the second smallest eigenvalue is a linear combination of the four eigenvectors, it generally will not give the correct approximation to τ. In the top right panel of Figure 3.1, we plot such a computed eigenvector u against τ, showing that it is no longer proportional to τ. Similar phenomenon occurs for noisy data as well, see the bottom row of Figure 3.1 where we added noise to the spiral curve data.

This example clearly shows the importance of the null space structure of the alignment matrix in recovering the parameter vectors. In particular, lack of overlap among the sections will result in a null space producing incorrect parameter vectors.

4. Perturbation Analysis of the Alignment Matrix. Affine errors introduced in the local coordinates will produce an inaccurate alignment matrix which determines if the resulting parameter vectors acceptable or not. In this section, we consider the effects of local approximation errors on the alignment matrix and its null space. We will make use of matrix perturbation analysis on the alignment matrix. Our approach consists of the following two parts: 1) error estimation of the approximation of alignment matrix in terms of the local errors and 2) perturbation analysis of null space of the alignment matrix resulted by the approximation error. In particular, we will show that the local errors are magnified by the condition numbers of the centered sections $\hat{T}_i = T_i - \bar{t}_i e^T$, where $\bar{t}_i$ is the mean of columns in T_i. In addition to the error in the alignment matrix due to the local approximations, the nonzero smallest eigenvalue of the exact alignment matrix Φ is also crucial to the determination of the accuracy of the computed parameter vectors.

To this end, let $X_1, \ldots, X_s$ be s sections of the sample points $x_1, \ldots, x_N$ given in (1.1) and $T_1, \ldots, T_s$ the corresponding sections of the parameter vectors $\tau_1, \ldots, \tau_N$.

Denote by I_i the index subset of the section i with size $k_i = |I_i|$, i.e.,

$$X_i = \{x_j \,|\, j \in I_i\}, \quad T_i = \{\tau_j \,|\, j \in I_i\}.$$

The local coordinates, denoted by Θ_i, of points in section X_i are generally not equal to T_i within an affine transformation. The optimal affine error is

$$\|E_i\|_2 = \min_{c,L} \|T_i - (ce^T + L\Theta_i)\|_2. \tag{4.1}$$

As shown in (2.1), $\|E_i\|_2 = \|T_i\widetilde{\Phi}_i\|_2$. We now consider how the local errors affect the alignment matrix.

Denote by Φ the alignment matrix constructed by the exact parameter vector sections $T_1, \ldots, T_s$, and $\widetilde{\Phi}$ the alignment matrix constructed by the sections of local coordinates $\Theta_1, \ldots, \Theta_s$, as in (2.3),

$$\widetilde{\Phi} = \sum_{i=1}^{s} S_i \widetilde{\Phi}_i S_i^T, \quad \Phi = \sum_{i=1}^{s} S_i \Phi_i S_i^T,$$

where, $\widetilde{\Phi}_i$ and Φ_i are the orthogonal projections with null spaces

$$\mathcal{N}(\widetilde{\Phi}_i) = \mathrm{span}([e, \Theta_i^T]), \quad \mathcal{N}(\Phi_i) = \mathrm{span}([e, T_i^T]),$$

respectively. We assume that both $[e, \Theta_i^T]$ and $[e, T_i^T]$ are of full-column rank, and $k_i \geq d+2$ for all i to insure that $\widetilde{\Phi}_i$ and Φ_i are not identically zero. It is easy to verify that $[e, T_i^T]$ is of full column rank if and only if the centered matrix $\hat{T}_i = T_i - \bar{t}_i e^T$ is of full row rank. In that case, $\hat{T}_i$ has a finite condition number defined by $\kappa(\hat{T}_i) = \|\hat{T}_i\|_2 \|\hat{T}_i^\dagger\|_2$, which will appear in our error bound below.

THEOREM 4.1. *Let $\|E_i\|_2$ denote the local error defined in (4.1) and $\kappa(\hat{T}_i)$ the condition number of $\hat{T}_i$. Then*

$$\|\widetilde{\Phi} - \Phi\|_2 \leq \sum_{i=1}^{s} \frac{\|E_i\|_2}{\|\hat{T}_i\|_2} \kappa(\hat{T}_i). \tag{4.2}$$

Proof. The error matrix $\widetilde{\Phi} - \Phi$ is clearly given by $\widetilde{\Phi} - \Phi = \sum_{i=1}^{s} S_i(\widetilde{\Phi}_i - \Phi_i)S_i^T$, and hence, $\|\widetilde{\Phi} - \Phi\|_2 \leq \sum_i \|\widetilde{\Phi}_i - \Phi_i\|_2$. What we need to do is to bound the errors $\|\widetilde{\Phi}_i - \Phi_i\|_2$. Since both $\widetilde{\Phi}_i$ and Φ_i are orthogonal projections with the same rank, by Theorem 2.6.1 of [6] we have that

$$\|\widetilde{\Phi}_i - \Phi_i\|_2 = \|(I - \Phi_i)\widetilde{\Phi}_i\|_2.$$

We can write $I - \Phi_i = \frac{1}{k_i} ee^T + \hat{T}_i^\dagger \hat{T}_i$, because $I - \Phi_i$ is the orthogonal projection onto $\mathrm{span}([e, T_i^T])$. It follows from $e^T\widetilde{\Phi}_i = 0$ that

$$\|\widetilde{\Phi}_i - \Phi_i\|_2 = \|\hat{T}_i^\dagger \hat{T}_i \widetilde{\Phi}_i\|_2 \leq \|\hat{T}_i^\dagger\|_2 \|\hat{T}_i\widetilde{\Phi}_i\|_2 = \frac{\|E_i\|_2}{\|\hat{T}_i\|_2} \kappa(\hat{T}_i).$$

The error bound in (4.2) follows immediately by summing the above error bounds. □

It is gratifying to see that the local errors affect the alignment matrix in a linear fashion, albeit by a factor which is the condition number of $\hat{T}_i$. We remark that these condition numbers may be made smaller if we increase the size of the neighborhoods.

Now we consider the perturbation analysis of the null space of the alignment matrix. The following theorem gives an error bound for this approximation (related to Theorem 4.1 in [18]) in terms of the the smallest nonzero eigenvalue $\lambda_{\min}^{+}(\Phi)$ of Φ and the approximation error $\|\widetilde{\Phi} - \Phi\|_2$.

THEOREM 4.2. *Let $r = \dim(\mathcal{N}(\Phi))$ and let U be an eigenvector matrix of $\widetilde{\Phi}$ corresponding to the r smallest eigenvalues. Denote $\lambda_{\min}^{+} = \lambda_{\min}^{+}(\Phi)$, $\epsilon = \|\widetilde{\Phi} - \Phi\|_2$. If $\epsilon < \frac{1}{4}\lambda_{\min}^{+}$ and $4\epsilon^2(1 - \lambda_{\min}^{+} + 2\epsilon) < (\lambda_{\min}^{+} - 2\epsilon)^3$, then there exists an orthonormal basis matrix G of $\mathcal{N}(\Phi)$ such that*

$$\|U - G\|_2 \leq \frac{2\epsilon}{\lambda_{\min}^{+} - 2\epsilon}. \tag{4.3}$$

Proof. Let G_0 be an orthonormal basis matrix of $\mathcal{N}(\Phi)$ and G_1 the orthogonal complement of G_0, i.e., $[G_0, G_1]$ is an orthogonal matrix. By the standard perturbation theory [11, Theorem V.2.7] for invariant subspaces, there is a matrix P satisfying

$$\|P\|_2 \leq \frac{2\epsilon}{\lambda_{\min}^{+} - 2\epsilon} \tag{4.4}$$

such that $\widetilde{U} = (G_0 + G_1 P)(I + P^T P)^{-1/2}$ is an orthogonal basis matrix of an invariant subspace of $\widetilde{\Phi}$. By simple calculation, we have that

$$\begin{aligned}
\|\widetilde{U} - G_0\|_2 &= \left\| \begin{bmatrix} (I + P^T P)^{-1/2} - I \\ P(I + P^T P)^{-1/2} \end{bmatrix} \right\|_2 \\
&= \left\| \begin{bmatrix} (I + P^T P)^{-1/2} - I \\ P(I + P^T P)^{-1/2} \end{bmatrix}^T \begin{bmatrix} (I + P^T P)^{-1/2} - I \\ P(I + P^T P)^{-1/2} \end{bmatrix} \right\|_2^{1/2} \\
&= \|2(I - (I + P^T P)^{-1/2})\|_2^{1/2} \\
&\leq \|P\|_2.
\end{aligned}$$

The error bound (4.3) follows from the above bound and (4.4) if we can prove that $U = \widetilde{U} Q^T$ holds with an orthogonal matrix Q of order r and we also set $G = G_0 Q^T$. This is equivalent to proving that the invariant subspace $\mathrm{span}(\widetilde{U})$ of $\widetilde{\Phi}$ is associated with the r smallest eigenvalues of $\widetilde{\Phi}$. We just need to show that $\|\widetilde{\Phi}\widetilde{U}\|_2 < \lambda_{r+1}(\widetilde{\Phi})$.

We first estimate $\lambda_{r+1}(\widetilde{\Phi})$. By eigenvalue perturbation theory of symmetric matrices [11], $|\lambda_i(\widetilde{\Phi}) - \lambda_i(\Phi)| \leq \|\widetilde{\Phi} - \Phi\|_2$. It follows that

$$\lambda_{r+1}(\widetilde{\Phi}) \geq \lambda_{\min}^{+} - \epsilon,$$

since Φ is positive semidefinite and $\lambda_{r+1}(\Phi) = \lambda_{\min}^{+}$. On the other hand, by (4.4),

$$\begin{aligned}
\|\widetilde{\Phi}\widetilde{U}\|_2 = \|\widetilde{U}^T\widetilde{\Phi}\widetilde{U}\|_2 &\leq \|\widetilde{U}^T(\widetilde{\Phi} - \Phi)\widetilde{U}\|_2 + \|\widetilde{U}^T\widetilde{\Phi}\widetilde{U}\|_2 \\
&< \|\widetilde{\Phi} - \Phi\|_2 + \frac{\|P\|_2^2}{1 + \|P\|_2^2} \\
&\leq \epsilon + \frac{4\epsilon^2}{4\epsilon^2 + (\lambda_{\min}^{+} - 2\epsilon)^2} \\
&< \lambda_{\min}^{+} - \epsilon,
\end{aligned}$$

because $4\epsilon^2(1 - \lambda_{\min}^{+} + 2\epsilon) < (\lambda_{\min}^{+} - 2\epsilon)^3$. Thus $\|\widetilde{\Phi}\widetilde{U}\|_2 < \lambda_{r+1}(\widetilde{\Phi})$. ∎

We now explain why Theorem 4.2 illustrates the importance of $\mathcal{N}(\Phi)$ and $\lambda^+_{\min}(\Phi)$ in understanding the alignment procedure in manifold learning. As we will show in the next section, it is always true that $\mathrm{span}([e, T^T]) \subseteq \mathcal{N}(\Phi)$. Theorem 4.2 shows that the true parameter vectors can be obtained, up to the error bound in (4.3), from the invariant subspace of the computed alignment matrix $\widetilde{\Phi}$ corresponding to its smallest eigenvalues, provided the errors introduced to the alignment matrix are relatively small. The smallest positive eigenvalue $\lambda^+_{\min}(\Phi)$ of the true alignment matrix determines how much error is allowed in the computed alignment matrix for a reliable recovery of the parameter vectors by LTSA. Specifically, if $\mathcal{N}(\Phi) = \mathrm{span}([e, T^T])$, good approximation in the local coordinate matrices Θ_i and a not too small $\lambda^+_{\min}(\Phi)$ will guarantee that the eigenvector matrix of $\widetilde{\Phi}$ corresponding to the $d+1$ smallest eigenvalues will give a good approximation of the parameter vectors T up to an affine transformation.

5. The Null Space of the Alignment Matrix. This section focuses on the null space of the ideal alignment matrix Φ. We will establish conditions under which the equality $\mathcal{N}(\Phi) = \mathrm{span}([e, T^T])$ holds. The section is divided into the following five parts: 1) we first establish some general properties about the null space of the alignment matrix; 2) we then present a necessary *and* sufficient condition for $\mathcal{N}(\Phi) = \mathrm{span}([e, T^T])$ in the special case when we have two sections, i.e., $s = 2$; 3) we give necessary conditions for the general case $s \geq 3$; 4) we also present sufficient conditions for the general case $s \geq 3$, and 5) finally we establish an interesting contraction property of $\mathcal{N}(\Phi)$ when some sections are merged into super-sections.

5.1. General properties of $\mathcal{N}(\Phi)$. It follows from the definition of Φ that

$$\Phi[e, T^T] = \sum_i S_i \Phi_i S_i^T [e, T^T] = \sum_i S_i \Phi_i [e, T_i^T] = 0,$$

which implies that

$$\mathrm{span}([e, T^T]) \subseteq \mathcal{N}(\Phi). \tag{5.1}$$

Consider a null vector $v \in \mathcal{N}(\Phi)$. Denote by $v_i = S_i v$ the restriction of v to the section T_i, $i = 1, \ldots, s$. Since each term $S_i \Phi_i S_i^T$ in Φ is positive semidefinite, $\Phi v = 0$ implies $S_i \Phi_i S_i^T v = 0$, hence the restriction v_i must be a null vector of Φ_i. So $v_i \in \mathrm{span}([e, T_i^T])$ by the definition of Φ_i, and therefore, it can be represented as

$$v_i = [e, T_i^T] w_i, \quad w_i \in \mathcal{R}^{d+1}. \tag{5.2}$$

The vector w_i defines an affine transformation from $\mathcal{R}^d$ to $\mathcal{R}$, $w_i : \tau \to [1, \tau^T] w_i \equiv w_i(\tau)$. Notice that the common part of each pair v_i and v_j should be equal, i.e.,

$$[e, T_{ij}^T] w_i = [e, T_{ij}^T] w_j, \tag{5.3}$$

where T_{ij} is the intersection of T_i and T_j.

DEFINITION 5.1. *Let $w = \{w_1, \ldots, w_s\}$ be a set of $(d+1)$-dimensional vectors. We call w a* certificate *for the collection $\{T_1, \ldots, T_s\}$ if the conditions (5.3) hold for all $i \neq j$. In particular, w is a* trivial *certificate if all w_i's are equal to each other.*

As we mentioned above, each certificate $w = \{w_1, \ldots, w_s\}$ defines a collection of s linear affine maps from $\mathcal{R}^d$ to $\mathcal{R}$: $\tau \to [1, \tau^T] w_i$. If we restrict the i-th map w_i on the columns of T_i, then w defines a function on the N columns of T to $\mathcal{R}$:

$$w : \ \tau \in T_i \to [1, \tau^T] w_i, \quad i = 1, \cdots, s,$$

where $\tau \in T_i$ means that τ is a column of T_i. There is no ambiguity for vectors belonging to the intersection of two sections, say T_i and T_j, since the conditions (5.3) hold. Thus, w maps $T \in \mathcal{R}^{d\times N}$ to a vector $v \in \mathcal{R}^N$ whose j-th component is defined by $w(\tau_j) = [1, \tau_j^T]w_i$ if the j-th column $\tau_j \in T_i$, i.e.,

$$v = w(T) \equiv [w(\tau_1), \cdots, w(\tau_N)]^T.$$

What we are interested is the set $\mathcal{W} = \mathcal{W}_{\{T_i\}}$ of all certificates of a *fixed* collection $\{T_i\}$ of T. It is easy to verify that $\mathcal{W}$ is a linear space with the usual addition and scalar multiplication operations. For the fixed collection $\{T_i\}$ of T, let us denote by ϕ the mapping from $\mathcal{W}$ to $\mathcal{R}^N$ determined by $w(T)$:

$$\phi : w \to v = w(T), \tag{5.4}$$

and denote it as $v = \phi(w)$. It is easy to verify that ϕ is a linear map.

There is a close relation between $\mathcal{N}(\Phi)$ and the certificate space $\mathcal{W}$ through the linear map ϕ for the considered collection $\{T_1, \ldots, T_s\}$: for a given $w \in \mathcal{W}$, consider the restriction v_i of vector $v = \phi(w)$ to T_i. By definition, v_i is given by (5.2) for $i = 1, \cdots, s$. It follows that v is a null vector of Φ. On the other hand, we have shown that for each $v \in \mathcal{N}(\Phi)$, there is a certificate $w = \{w_1, \ldots, w_s\}$ satisfying (5.2). This implies $v = \phi(w)$. Therefore, ϕ is an onto-map from $\mathcal{W}$ to $\mathcal{N}(\Phi)$. Since we always assume that each $[e, T_i^T]$ is of full column rank, ϕ is also one-to-one and hence *isomorphic*. Specially, ϕ maps a trivial certificate to a vector in $\text{span}([e, T^T]) \subset \mathcal{N}(\Phi)$.

THEOREM 5.2. *1) The null space $\mathcal{N}(\Phi)$ and the certificate space $\mathcal{W}$ are isomorphic to each other and the linear transformation ϕ defined above is an isomorphism between the two linear spaces. Moreover, the subspace of all trivial certificates is isomorphic to the subspace* $\text{span}([e, T^T])$ *of* $\mathcal{N}(\Phi)$.

2) The equality $\mathcal{N}(\Phi) = \text{span}([e, T^T])$ *holds if and only if* $\{T_1, \ldots, T_s\}$ *has only trivial certificates.*

We single out those collections that have only trivial certificates.

DEFINITION 5.3. *We call a collection* $\{T_1, \ldots, T_s\}$ affinely rigid *if it has only trivial certificates.*

Geometrically, those are the collections the overlaps among their sections are strong and exhibit certain rigidity reminiscent of graph rigidity discussed in [7]. In particular, part 2) of Theorem 5.2 can be restated as

$$\mathcal{N}(\Phi) = \text{span}([e, T^T]) \text{ if and only if } \{T_1, \ldots, T_s\} \text{ is affinely rigid.}$$

5.2. Necessary and sufficient conditions of affine rigidity for $s = 2$. Consider the case when $s = 2$, i.e., $\Phi = S_1\Phi_1S_1^T + S_2\Phi_2S_2^T$ for two sections T_1 and T_2. In this case, we can characterize affine rigidity using an intuitive geometric notion defined below.

DEFINITION 5.4. *We say two sections T_i and T_j are* fully overlapped *if the intersection $T_{ij} = T_i \cap T_j$ is not empty and $[e, T_{ij}^T]$ is of full column-rank.*

Clearly, T_i and T_j are fully overlapped if they share at least *two* distinct points in the one-dimensional case $d = 1$, or if they share at least three points that are not co-linear in the two-dimensional case. Using this concept, we can establish the following result.

THEOREM 5.5. $\{T_1, T_2\}$ *is affinely rigid if and only if T_1 and T_2 are fully overlapped.*

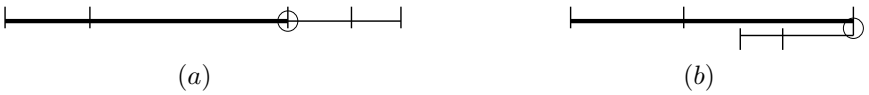

FIG. 5.1. *Two possible layouts for the global coordinates.*

Proof. We only show the necessity. Let us assume that T_1 and T_2 are not fully overlapped. Since $[e, T_{12}^T]$ is not of full column rank, we can find distinct w_1 and w_2 such that $T_{12}w_1 = T_{12}w_2$. Thus $w = \{w_1, w_2\}$ is a non-trivial certificate for $\{T_1, T_2\}$. Hence, $\{T_1, T_2\}$ is not affinely rigid by Theorem 5.2. □

We now illustrate the case when a pair of sections are not fully overlapped by a simple example with $d = 1$.

EXAMPLE 2. Consider the situation depicted in Figure 5.1. The data set has five points marked by short vertical bars. Two sections are considered as shown in panel (a) of Figure 5.1 with a thick line segment and a thin line segment connecting the points in each section. The first section consists of the left three points, and the second one consists of the right three points. The two sections share a single point denoted by a circle. We can fold the second section around the point marked by circle, while keeping the first section unchanged, see the resulting layout shown in the panel (b). This example clearly shows that the collection of the two sections is not affinely rigid.

The algebraic picture of the above is also clear. Let $\tau_1, \ldots, \tau_5$ be real numbers denoting the five different points. $T = [\tau_1, \ldots, \tau_5]$, $T_1 = [\tau_1, \tau_2, \tau_3]$ and $T_2 = [\tau_3, \tau_4, \tau_5]$, giving $T_{12} = \tau_3$. It is easy to verify that $[e,\, T_1^T]$ and $[e,\, T_2^T]$ are of full rank. However $[e, T_{12}^T]$ has a nonzero null vector $w_0 = [\tau_3, -1]^T$. Thus, for each certificate $w = \{w_1, w_2\}$ of $\{T_1, T_2\}$, $w' = \{w_1, w_2 + w_0\}$ is a different certificate of $\{T_1, T_2\}$. One of w and w' must be non-trivial, and hence the collection of sections is not affinely rigid.

5.3. Necessary conditions of affine rigidity for $s \geq 3$. For the case when a collection has three or more sections, we can partition the sections into two subsets, say $\{T_{i_1}, \ldots, T_{i_k}\}$ and $\{T_{i_{k+1}}, \ldots, T_{i_s}\}$, and consider the union of the sections in each subset,

$$T_{1:k} = T_{i_1} \cup \ldots \cup T_{i_k}, \quad T_{k+1:s} = T_{i_{k+1}} \cup \ldots \cup T_{i_s}.$$

The following theorem shows that affine rigidity of T implies that $T_{1:k}$ and $T_{k+1:s}$ are fully overlapped.

THEOREM 5.6. *If the collection $\{T_1, \ldots, T_s\}$ is affinely rigid, then for any partitions $\{T_{i_1}, \ldots, T_{i_k}\}$ and $\{T_{i_{k+1}}, \ldots, T_{i_s}\}$ with $1 \leq k < s$, $\cup_{j=1}^k T_{i_j}$ and $\cup_{j=k+1}^s T_{i_j}$ are fully overlapped.*

Proof. We prove this theorem by reduction to absurdity. If there is a partition, without loss of generality we denote the partition as, $\{T_1, \cdots, T_k\}$ and $\{T_{k+1}, \ldots, T_s\}$ $(k < s)$ such that the two super-sections $T_{1:k} = \cup_{j=1}^k T_j$ and $T_{k+1:s} = \cup_{j=k+1}^s T_j$ are not fully overlapped, then there are $(d+1)$-dimensional vectors $w' \neq w''$ such that

$$[e, T_{1:k,k+1:s}^T]w' = [e, T_{1:k,k+1:s}^T]w'',$$

where $T_{1:k,k+1:s}$ is the intersection of $T_{1:k}$ and $T_{k+1:s}$. Define $w = \{w_1, \ldots, w_s\}$ with

$$w_1 = \cdots = w_k = w', \quad w_{k+1} = \cdots = w_s = w'',$$

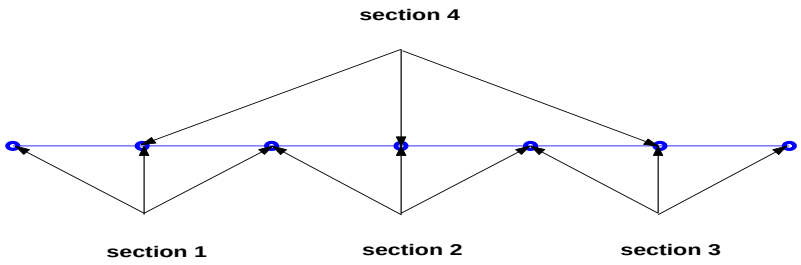

FIG. 5.2. *Overlapping patterns of four sections.*

It is obvious that $w = \{w_1, \ldots, w_s\}$ is a non-trivial certificate for $\{T_1, \ldots, T_s\}$. By Definition 5.3, $\{T_1, \ldots, T_s\}$ is not an affinely rigid, a contradiction to the assumption of the theorem. □

The necessary condition shown above is, however, not sufficient if $s > 2$. Below is a counterexample for $s = 4$ and an arbitrary k with $1 \leq k < s$.

EXAMPLE 3. Consider a data set of seven one-dimensional points

$$\{-3, -2, -1, 0, 1, 2, 3\}$$

and an associated collection of four sections ($s = 4$),

$$T_1 = [-3, -2, -1], \quad T_2 = [-1, 0, 1], \quad T_3 = [1, 2, 3], \quad T_4 = [-2, 0, 2].$$

See Figure 5.2 for each section denoted by arrows emitting from a single point. Clearly each section T_i and the union of the rest are fully overlapped. Let $k = 1, 2$ or 3. Consider any partitions $\{T_{i_1}, \ldots, T_{i_k}\}$ and $\{T_{i_{k+1}}, \ldots, T_{i_4}\}$. It is easy to verify that the union $\cup_{j=1}^{k} T_{i_j}$ and $\cup_{j=k+1}^{4} T_{i_j}$ are always fully overlapped, since they share two or more distinct points in the line. We show, however, that $\{T_1, \cdots, T_4\}$ is not affinely rigid. To this end, we represent each Φ_i explicitly as follows: $\Phi_i = \frac{1}{6} q q^T$ with $q = [1, -2, 1]^T$, due to each $[e, T_i^T]^T$ has the same null space span(q). Let $z = [0, 0, 0, 1, 2, 2, 2]^T$. The restrictions $z_i = S_i^T z$ of z corresponding to T_i are

$$S_1^T z = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \quad S_2^T z = \begin{bmatrix} 0 \\ 1 \\ 2 \end{bmatrix}, \quad S_3^T z = \begin{bmatrix} 2 \\ 2 \\ 2 \end{bmatrix}, \quad S_4^T z = \begin{bmatrix} 0 \\ 1 \\ 2 \end{bmatrix},$$

respectively. Since $z_i^T q = 0$ for $i = 1, \cdots, 4$, we conclude that z is a null vector of Φ. However, $z \notin \text{span}([e, T^T])$. So $\mathcal{N}(\Phi) \neq \text{span}([e,\ T^T])$, or equivalently, $\{T_1, \cdots, T_4\}$ is not affinely rigid.

In the above example, any pair of sections are not fully overlapped. However, it is also possible that a collection is affinely rigid even if any pair of its sections are not fully overlapped. Here is a simple example: Let T be the matrix of three vertices of a regular triangle and T_1, T_2, T_3 be three sections each consists of two vertices. The resulting collection is affinely rigid but T_i and T_j are not fully overlapped for $i \neq j$.

5.4. Sufficient conditions of affine rigidity for $s \geq 3$. We associate a collection of sections $\{T_1, \ldots, T_s\}$ with a graph G constructed as follows: its s vertices represent the s sections, where there is an edge between vertices i and j if sections T_i and T_j are fully overlapped. The following theorem gives a sufficient condition for

affine rigidity of a collection of sections based on the connectedness of its associated graph G.

THEOREM 5.7. *The collection $\{T_1, \ldots, T_s\}$ is affinely rigid if its associated graph G is connected.*

Proof. We need to show if $w = \{w_1, \ldots, w_s\}$ is a certificate of T, then G is connected implies that w is a trivial certificate.

Consider any pair of w_i and w_j. Because G is connected, there is a path, say $i_1 = i, \cdots, i_r = j$, connecting vertices i and j. The adjacency between i_k and i_{k+1} implies that T_{i_k} and $T_{i_{k+1}}$ are fully overlapped, i.e., $\mathcal{N}([e, T^T_{i_k, i_{k+1}}]) = \{0\}$. It follows from (5.3) with $i = i_k$ and $j = i_{k+1}$ that $w_{i_k} = w_{i_{k+1}}$ for $k = 1, \cdots, r-1$. Hence $w_i \equiv w_{i_1} = w_{i_2} = \ldots = w_{i_r} \equiv w_j$. □

Now we consider the case when the graph G of $\{T_1, \ldots, T_s\}$ is not connected. Let the *connected components* of G be $\{G_1, \cdots, G_r\}$, i.e., each G_j is a connected subgraph of G and there are no edges between vertices in different subgraphs. We denote by J_j the index set of the vertices in subgraph G_j, and merge the sections T_k, $k \in J_j$ into a super-section

$$T_{J_j} = \cup_{k \in J_j} T_k,$$

i.e., the matrix consisting of column vectors in $\{T_k, k \in J_j\}$. This collection of super-sections $\{T_{J_1}, \ldots, T_{J_r}\}$ produces an alignment matrix $\widehat{\Phi}$. We show that both Φ and $\widehat{\Phi}$ share a common null space.

THEOREM 5.8. *Let $\{T_{J_1}, \cdots, T_{J_r}\}$ be the super-sections obtained by merging connected components of $\{T_1, \cdots, T_s\}$. Then $\mathcal{N}(\widehat{\Phi}) = \mathcal{N}(\Phi)$.*

Proof. Consider a null vector v of Φ, $v = \phi(w)$ with a certificate $w = \{w_1, \ldots, w_s\}$ of $\{T_1, \ldots, T_s\}$. Due to the connectedness of subgraphs G_j, the sub-collection $\{T_k,\ k \in J_j\}$ is affinely rigid by Theorem 5.7, and hence, each subset $\{w_k,\ k \in J_j\}$ is a trivial certificate for the sub-collection $\{T_k,\ k \in J_j\}$, i.e., all $w_k, k \in J_j$ are equal to each other. We simply denote them by w_{J_j}, i.e., $w_k = w_{J_j}$ for $k \in J_j$, $j = 1, \cdots, r$. The set $\hat{w} = \{w_{J_1}, \ldots, w_{J_r}\}$ is clearly a certificate of $\{T_{J_1}, \ldots, T_{J_r}\}$. It is easy to verify that $\phi(w) = \hat{\phi}(\hat{w})$, where $\hat{\phi}$ is the isomorphic mapping from the certificate space of $\{T_{J_1}, \ldots, T_{J_r}\}$ to the null space of the alignment matrix $\widehat{\Phi}$. Thus $v = \phi(w) = \hat{\phi}(\hat{w}) \in \mathcal{N}(\widehat{\Phi})$. On the other hand, any null vector of $\widehat{\Phi}$ also belongs to $\mathcal{N}(\Phi)$. □

The above theorem says that merging sections with connected associated graphs does not change the null space of the alignment matrix. Equivalently, the affine rigidity of the original sections can be detected from the affine rigidity of the resulting super-sections. This fact motivates us to consider the connectedness of the associated graph $\hat{G}$ for the collection of the super-sections $\{T_{J_1}, \cdots, T_{J_r}\}$, where there is an edge between two vertices if the associated super sections are fully overlapped. We call $\hat{G}$ a coarsening of G. By Theorem 5.7 and Theorem 5.8, $\{T_1, \ldots, T_s\}$ is affinely rigid if $\hat{G}$ is connected. This coarsening procedure can be repeated, i.e., by finding the connected components of $\hat{G}$ and so on. This coarsening procedure terminates, if

1) the current graph has only one vertex, or
2) the current graph has two or more vertices and all vertices are isolated.

We call the graph obtained in the last step of the above coarsening procedure the coarsest graph and denote it by G^*. We also use $|G|$ to denote the number of vertices in a graph G. One can easily prove the following result by Theorems 5.5 and 5.7.

THEOREM 5.9. *Let G^* be the coarsest graph of the collection $\{T_1, \ldots, T_s\}$. Then*
(1) $\{T_1, \ldots, T_s\}$ is affinely rigid if $|G^*| = 1$, and
(2) $\{T_1, \ldots, T_s\}$ is not affinely rigid if $|G^*| = 2$, or if $|G^*| = 3$ and $d = 1$.

Proof. We just prove that if $d = 1$ and $|G^*| = 3$, then $\{T_1, T_2, T_3\}$ is not affinely rigid. We show this by constructing a non-trivial certificate for T.

Without loss of generality, we assume that the intersection T_{ij} between T_i and T_j is not empty for $i \neq j$. Since T_i and T_j are not fully overlapped for $i \neq j$, $\text{rank}([e, T_{ij}^T]) < d + 1 = 2$ and hence $\text{rank}([e, T_{ij}^T]) = 1$.

Now for the construction of a non-trivial certificate $w = \{w_1, w_2, w_3\}$, we can assume $w_3 = 0$ without loss of generality. Thus, $w = \{w_1, w_2, w_3\}$ is non-trivial if and only if either w_1 or w_2 is not zero. The conditions given in (5.3) now state

$$[e, T_{12}^T]w_1 = [e, T_{12}^T]w_2, \quad [e, T_{23}^T]w_2 = 0, \quad [e, T_{31}^T]w_1 = 0.$$

We rewrite the equations in the following matrix form:

$$\begin{bmatrix} [e, T_{12}^T] & [e, T_{12}^T] \\ 0 & [e, T_{23}^T] \\ [e, T_{13}^T] & 0 \end{bmatrix} \begin{bmatrix} w_1 \\ -w_2 \end{bmatrix} = 0. \tag{5.5}$$

Because the rank of the coefficient matrix is less than or equal to three and its column number is no less than four, the above linear equations have a nonzero solution $\begin{bmatrix} w_1 \\ -w_2 \end{bmatrix}$. Therefore, a non-trivial certificate exists for the collection $\{T_1, T_2, T_3\}$. By Theorem 5.2, $\{T_1, T_2, T_3\}$ is not affinely rigid. □

Unfortunately, we still cannot conclude that the original collection is not affinely rigid for the more general case $|G^*| > 3$. Here is a counterexample with $d = 1$ from Example 3.

EXAMPLE 4. We change the first section in Example 3 by adding the last point to it, and keep other sections unchanged,

$$T_1 = [-3, -2, -1, 3], \quad T_2 = [-1, 0, 1], \quad T_3 = [1, 2, 3], \quad T_4 = [-2, 0, 2].$$

Any two sections in the collection are not fully overlapped since the size of each intersection T_{ij} is one and $[e, T_{ij}^T]$ is a 1×2 matrix that is not of full collum rank. So there are no edges in the associated graph, i.e., $G^* = G$ and $|G^*| = 4$. However, the collection is still affinely rigid.

Appendix B shows the existence of an affinely rigid collection with $|G^*| = s$ for any s and d satisfying $3 \leq s \leq d + 1$. Of course, one can also construct a collection with $|G^*| = s$ that is not affinely rigid. Appendix C gives geometric conditions for affinely rigid collections with $|G^*| = 3$ and $d = 2$.

5.5. Merging sections. In the last subsection, we discuss a coarsening procedure that involves merging connected components, i.e., merging the sections in a connected component into a super-section. This kind of coarsening procedure preserves the null space (cf. Theorem 5.8). In this subsection, we further discuss the merging process with regard to: 1) merging components that are not necessarily connected; and 2) merging sections that do not form a connected component but the corresponding sub-collection is affinely rigid. We will show that for 1) the size of null space does not increase while for 2) the null space remains unchanged.

THEOREM 5.10. *Let Φ and $\widehat{\Phi}$ be the two alignment matrices of $\{T_1, \ldots, T_s\}$ and $\{T_{J_1}, \ldots, T_{J_t}\}$, respectively, where T_{J_j} is the super-section merging sections $\{T_i,\ i \in J_j\}$, $j = 1, \ldots, t$. Then $\mathcal{N}(\widehat{\Phi}) \subseteq \mathcal{N}(\Phi)$.*

Proof. Given a certificate $\hat{w} = \{w_{J_1}, \ldots, w_{J_t}\}$ for the collection $\{T_{J_1}, \ldots, T_{J_t}\}$. It can be split to a certificate $w = \{w_1, \ldots, w_s\}$ of $\{T_1, \ldots, T_s\}$ with $w_k = w_{J_j}$, for

$k \in J_j$, $j = 1, \ldots, t$. By definition, $\phi(w) = \hat{\phi}(\hat{w})$ with the isomorphic mappings ϕ and $\hat{\phi}$ as defined in (5.4). So each null vector $\phi(\hat{w})$ of $\widehat{\Phi}$ is also a null vector of Φ, i.e., $\mathcal{N}(\widehat{\Phi}) \subseteq \mathcal{N}(\Phi)$. □

Theorem 5.10 suggests that one can modify the alignment matrix Φ by merging the sections in order to push the null space to the desired subspace $\text{span}([e, T^T])$. For example, if merge any two sections given in Example 3, the graph of the resulting sections can be recursively coarsened to a connected graph and hence for the modified $\widehat{\Phi}$, $\mathcal{N}(\widehat{\Phi}) = \text{span}([e, T^T])$ holds by Theorem 5.9.

The following theorem shows that merging affinely rigid sub-collections cannot change the null space of the alignment matrix. It generalizes Theorem 5.8 slightly and has a similar proof which will not be repeated here.

THEOREM 5.11. *Let Φ and $\widehat{\Phi}$ be the two alignment matrices of $\{T_1, \ldots, T_s\}$ and $\{T_{J_1}, \ldots, T_{J_t}\}$, respectively. If each sub-collection $\{T_i,\ i \in J_j\}$ is affinely rigid for $j = 1, \ldots, t$, then $\mathcal{N}(\widehat{\Phi}) = \mathcal{N}(\Phi)$.*

6. The Smallest Nonzero Eigenvalue(s) of the Alignment Matrix. How well $\mathcal{N}(\Phi)$ can be determined numerically depends on the magnitude of its smallest nonzero eigenvalue(s). This has significant ramifications when we need to use eigenvectors of an approximation of Φ corresponding to small eigenvalues to recover the parameter vectors $\{\tau_i\}$. Theorem 4.2 gives further elaboration in this regard. The objective of this section is to establish bounds for the smallest nonzero eigenvalue. We first give a characterization of the smallest nonzero eigenvalue $\lambda^+_{\min}(\Phi)$ of Φ.

THEOREM 6.1. *Let $\Phi_i = Q_i Q_i^T$ be the orthogonal projections such that $\mathcal{N}(\Phi_i) = \text{span}([e, T_i^T])$ and Q_i orthonormal. Let $H = (H_{ij})$ be a block matrix with blocks $H_{ij} = (S_i Q_i)^T (S_j Q_j)$, S_i is the selection matrix for T_i. Then $\lambda^+_{\min}(\Phi) = \lambda^+_{\min}(H)$.*

Furthermore, if $s = 2$, then

$$\lambda^+_{\min}(\Phi) = 1 - \max\{\, \sigma :\ \sigma \in \sigma(H_{12}),\, \sigma < 1 \,\},$$

where $\sigma(\cdot)$ denotes the set of singular values of a matrix.

Proof. Let $R = [S_1 Q_1, \cdots, S_s Q_s]$. By definition, $\Phi = RR^T$ and $H = R^T R$. It is well known that Φ and H have the same nonzero eigenvalues, since 1) the eigenvalue equation $RR^T x = \lambda x$ implies $R^T R y = \lambda y$ with $y = R^T x \neq 0$, while $R^T R y = \lambda y$ yields $RR^T z = \lambda z$ with $z = Ry$, and 2) the condition $\lambda \neq 0$ guarantees that x, y, and z are nonzero simultaneously. So $\lambda^+_{\min}(\Phi) = \lambda^+_{\min}(H)$.

For the case when $s = 2$, $H = I + \begin{bmatrix} 0 & H_{12} \\ H_{12}^T & 0 \end{bmatrix}$. Notice that the eigenvalues of $\begin{bmatrix} 0 & H_{12} \\ H_{12}^T & 0 \end{bmatrix}$ are given by $\{\sigma_1, \ldots, \sigma_{d+1}, -\sigma_1, \ldots, -\sigma_{d+1}\}$, where $\sigma_1 \geq \ldots \geq \sigma_{d+1}$ are the singular values of H_{12} [6]. Assume that $1 = \sigma_1 = \ldots = \sigma_\ell > \sigma_{\ell+1} \geq \cdots \geq \sigma_{d+1}$, then $\lambda^+_{\min}(H) = 1 - \sigma_{\ell+1}$. □

6.1. Submatrices H_{ij} of H. Now we focus on the matrix H, and proceed to derive an expression of its submatrices $H_{ij} = (S_i Q_i)^T (S_j Q_j) = Q_i^T S_i^T S_j Q_j$. Denote by $T_{\bar{i}\bar{j}}^c$ the remainder of T_i by deleting $T_{\bar{i}\bar{j}}$. Without loss of generality, we write $T_i = [T_{\bar{i}\bar{j}}^c, T_{\bar{i}\bar{j}}]$.

We first derive an expression for Q_i which will allow us to relate the centered matrix $T_{\bar{i}\bar{j}}^c - t_{\bar{i}\bar{j}} e^T$ to H_{ij}, here $t_{\bar{i}\bar{j}}$ is the mean of the columns in $T_{\bar{i}\bar{j}}$. To this end, we partition

$$[e, (T_i - t_{\bar{i}\bar{j}} e^T)^T] = \begin{bmatrix} (e, (T_{\bar{i}\bar{j}}^c - t_{\bar{i}\bar{j}} e^T)^T) \\ (e, (T_{\bar{i}\bar{j}} - t_{\bar{i}\bar{j}} e^T)^T) \end{bmatrix} \equiv \begin{bmatrix} B_{\bar{i}\bar{j}}^c \\ B_{\bar{i}\bar{j}} \end{bmatrix}$$

and split it as $[e,(T_i - t_{ij}e^T)^T] = A_1 + A_2$, where

$$A_1 = [e,(T_i - t_{ij}e^T)^T]B_{ij}^\dagger B_{ij} = \begin{bmatrix} B_{ij}^c B_{ij}^\dagger B_{ij} \\ B_{ij} \end{bmatrix},$$
$$A_2 = [e,(T_i - t_{ij}e^T)^T]\big(I - B_{ij}^\dagger B_{ij}\big) = \begin{bmatrix} B_{ij}^c\big(I - B_{ij}^\dagger B_{ij}\big) \\ 0 \end{bmatrix}.$$

It is known that Q_i is orthogonal to $[e, T_i^T]$ if and only if Q_i is orthogonal to $[e,(T_i - t_{ij}e^T)^T]$, or equivalently, Q_i is orthogonal to both A_1 and A_2 since $\mathrm{span}([e,(T_i - t_{ij}e^T)^T]) = \mathrm{span}(A_1) \cup \mathrm{span}(A_2)$. Because of the structures of A_1 and A_2, one can construct such a Q_i as follows: Let V_{ij} be an orthogonal basis matrix of the orthogonal complement space of B_{ij}, and V_i an orthogonal basis matrix of the subspace orthogonal to $B_{ij}^c\big(I - B_{ij}^\dagger B_{ij}\big)$. Then the two matrices

$$C_1 = \begin{bmatrix} 0 \\ V_{ij} \end{bmatrix} \quad \text{and} \quad C_2 = \begin{bmatrix} -V_i \\ (B_{ij}^c B_{ij}^\dagger)^T V_i \end{bmatrix}$$

are orthogonal to both A_1 and A_2. Note that C_1 and C_2 are orthogonal to each other since the columns of $(B_{ij}^\dagger)^T$ are still in the range space of B_{ij}. C_1 is also orthonormal and C_2 can be normalized by multiplying it with $D_i = (I + V_i^T(B_{ij}^c B_{ij}^\dagger)(B_{ij}^c B_{ij}^\dagger)^T V_i)^{-1/2}$ from the right. So we can set Q_i to be the following orthonormal matrix,

$$Q_i = [C_1, C_2 D_i] = \begin{bmatrix} 0 & -V_i D_i \\ V_{ij} & (B_{ij}^c B_{ij}^\dagger)^T V_i D_i \end{bmatrix},$$

It follows that $Q_{ij} = [V_{ij},\ (B_{ij}^c B_{ij}^\dagger)^T V_i D_i]$. Similarly, $Q_{ji} = [V_{ij},\ (B_{ji}^c B_{ij}^\dagger)^T V_j D_j]$, where $B_{ji}^c = [e,(T_{ji}^c - t_{ij}e^T)^T]$, V_j is the basis matrix of the subspace orthogonal to $B_{ji}^c(I - B_{ij}^\dagger B_{ij})$, and $D_j = (I + V_j^T(B_{ji}^c B_{ij}^\dagger)(B_{ji}^c B_{ij}^\dagger)^T V_j)^{-1/2}$. Now we can represent $H_{ij} = Q_{ij}^T Q_{ji}$ as

$$H_{ij} = \begin{bmatrix} (V_i D_i)^T B_{ij}^c B_{ij}^\dagger (B_{ji}^c B_{ij}^\dagger)^T V_j D_j & \\ & I_{ij} \end{bmatrix} \equiv \begin{bmatrix} P_{ij} & \\ & I_{ij} \end{bmatrix}.$$

Note that

$$\|P_{ij}\|_2^2 \le \frac{\|B_{ij}^c B_{ij}^\dagger\|_2^2 \|B_{ji}^c B_{ij}^\dagger\|_2^2}{(1 + \|B_{ij}^c B_{ij}^\dagger\|_2^2)(1 + \|B_{ji}^c B_{ij}^\dagger\|_2^2)} < 1. \tag{6.1}$$

Therefore, the singular values of H_{ij} less than one consist of the singular values of P_{ij}.

6.2. Estimation of the singular values of P_{ij}. The matrix $B_{ij}^\dagger$ is given by

$$B_{ij}^\dagger = \begin{bmatrix} e^\dagger \\ (T_{ij} - t_{ij}e^T)^{T\dagger} \end{bmatrix},$$

since the first column of $B_{ij} = [e,(T_{ij} - t_{ij}e^T)^T]$ is orthogonal to the other columns. It follows that $B_{ij}^c B_{ij}^\dagger = ee^\dagger + \big((T_{ij} - t_{ij}e^T)^\dagger (T_{ij}^c - t_{ij}e^T)\big)^T$. Define $\eta_{ij} = \sigma_{\min}(T_{ij} - t_{ij}e^T)$, the smallest nonzero singular value of $T_{ij} - t_{ij}e^T$. We obtain

$$\begin{aligned} \|B_{ij}^c B_{ij}^\dagger\|_2^2 &\le 1 + \|(T_{ij} - t_{ij}e^T)^\dagger (T_{ij}^c - t_{ij}e^T)\|_2^2 \\ &\le 1 + \|(T_{ij} - t_{ij}e^T)^\dagger\|_2^2 \|T_{ij}^c - t_{ij}e^T\|_2^2 \\ &= 1 + \|T_{ij}^c - t_{ij}e^T\|_2^2/\eta_{ij}^2. \end{aligned}$$

Similarly, $\|B_{ji}^c B_{ij}^\dagger\|_2^2 \le 1 + \|T_{ji}^c - t_{ij}e^T\|_2^2/\eta_{ij}^2$. Substituting this bound into (6.1), we obtain that

$$\|P_{ij}\|_2^2 \le \frac{(\eta_{ij}^2 + \|T_{ij}^c - t_{ij}e^T\|_2^2)(\eta_{ij}^2 + \|T_{ji}^c - t_{ij}e^T\|_2^2)}{(2\eta_{ij}^2 + \|T_{ij}^c - t_{ij}e^T\|_2^2)(2\eta_{ij}^2 + \|T_{ji}^c - t_{ij}e^T\|_2^2)} \le \Big(1 - \frac{\eta_{ij}^2}{2\eta_{ij}^2 + \delta_{ij}^2}\Big)^2,$$

where $\delta_{ij} = \max\{\|T_{ij}^c - t_{ij}e^T\|_2, \|T_{ji}^c - t_{ij}e^T\|_2\}$. We have

$$\|P_{ij}\|_2 \le 1 - \frac{\eta_{ij}^2}{2\eta_{ij}^2 + \delta_{ij}^2}.$$

For the case $s = 2$, we have

$$\lambda_{\min}^+ \ge 1 - \|P_{12}\|_2 \ge \frac{\eta_{ij}^2}{2\eta_{ij}^2 + \delta_{ij}^2}$$

Therefore we have proved the following quantitative result giving a lower bound on the smallest nonzero eigenvalue of the alignment matrix for $s = 2$.

THEOREM 6.2. *Assume $s = 2$ and let $\delta_{12} = \max\{\|T_{12}^c - t_{12}e^T\|_2, \|T_{21}^c - t_{12}e^T\|_2\}$. Then*

$$\lambda_{\min}^+(\Phi) \ge \frac{\sigma_{\min}^2(T_{12} - t_{12}e^T)}{2\sigma_{\min}^2(T_{12} - t_{12}e^T) + \delta_{12}^2} \approx \Big(\frac{\sigma_{\min}(T_{12} - t_{12}e^T)}{\delta_{12}}\Big)^2.$$

Recall from Definition 5.4 that T_1 and T_2 are fully overlapped if $[e, T_{12}^T]$ is of full column rank. This condition is equivalent to $\sigma_{\min}(T_{12} - t_{12}e^T) > 0$. Therefore, $\sigma_{\min}(T_{12} - t_{12}e^T)$ can be considered as a quantitative measure of the *size* of the overlap between T_1 and T_2. The above theorem states that the null space can be well determined if T_1 and T_2 have a reasonably large overlap.

REMARK. For $s > 2$ case, it is still not clear how to formulate the concept of the size of overlaps and derive a similar bound.

6.3. Merging sections improves spectral gap. In section 5.5, we showed that merging sections to super-sections may reduce the null space of the alignment matrix. A natural question to ask is what effects merging sections will have on the smallest nonzero eigenvalue of the alignment matrix. To address this issue we look at a slightly more general notion of alignment matrix, we consider weighted alignment matrix defined as

$$\Phi(\alpha) = \sum_{i=1}^{s} \alpha_i S_i \Phi_i S_i^T, \quad \alpha_i > 0, \;\; i = 1, \ldots, s.$$

Obviously, $\mathcal{N}(\Phi(\alpha)) = \mathcal{N}(\Phi)$, but $\lambda_{\min}^+(\Phi(\alpha))$ and $\lambda_{\min}^+(\Phi)$ may be different. We remark that the results in the previous sections also hold for the weighted alignment matrices with trivial modifications.

THEOREM 6.3. *Let $\Phi(\alpha)$ be a weighted alignment matrix of $\{T_1, \ldots, T_s\}$ defined above, and T_{J_j} the super-section merging $\{T_k,\ k \in J_j\}$, $j = 1, \ldots, r$. Let*

$$\widehat{\Phi}(\hat{\alpha}) = \sum_{j=1}^{r} \hat{\alpha}_j \hat{S}_j \widehat{\Phi}_j \hat{S}_j^T, \quad \hat{\alpha}_j = \sum_{k \in J_j} \alpha_k, \;\; j = 1, \ldots, r$$

be the weighted alignment matrix of $\{T_{J_1}, \ldots, T_{J_r}\}$. *If each sub-collection* $\{T_k,\ k \in J_j\}$ *is affinely rigid,* $j = 1, \ldots, r$, *then*

$$\lambda^+_{\min}\big(\widehat{\Phi}(\hat{\alpha})\big) \geq \lambda^+_{\min}\big(\Phi(\alpha)\big).$$

Proof. Let $k \in J_j$, $S_{jk} = \hat{S}_j^T S_k$ and $\Phi_{jk} = S_{jk}\Phi_k S_{jk}^T$. We show $\widehat{\Phi}_j \geq \Phi_{jk}$, i.e., $\widehat{\Phi}_j - \Phi_{jk}$ is positive semidefinite. It is known that $\mathcal{N}(\widehat{\Phi}_j) \subseteq \mathcal{N}(\Phi_{jk})$. So $\mathcal{R}(\Phi_{jk}) \subseteq \mathcal{R}(\widehat{\Phi}_j)$.[4] We see that $\widehat{\Phi}_j - \Phi_{jk}$ is the orthogonal projection onto the orthogonal complement $\mathcal{R}(\Phi_{jk})^\perp$ of $\mathcal{R}(\Phi_{jk})$ restricted to $\mathcal{R}(\widehat{\Phi}_j)$. So $\widehat{\Phi}_j \geq \Phi_{jk}$.

Applying the above result and using $\hat{S}_j S_{jk} = S_k$, we see that

$$\widehat{\Phi}(\hat{\alpha}) = \sum_{j=1}^{r}\sum_{k\in J_j} \alpha_k \hat{S}_j \widehat{\Phi}_j \hat{S}_j^T \geq \sum_{j=1}^{r}\sum_{k\in J_j} \alpha_k \hat{S}_j \Phi_{jk} \hat{S}_j^T = \sum_{j=1}^{r}\sum_{k\in J_j} \alpha_k S_k \Phi_k S_k^T = \Phi(\alpha).$$

It yields that $\lambda_i\big(\widehat{\Phi}(\hat{\alpha})\big) \geq \lambda_i^+\big(\Phi(\alpha)\big)$ for all i. The result of the theorem follows immediately since by Theorem 5.11, $\mathcal{N}(\widehat{\Phi}(\hat{\alpha})) = \mathcal{N}(\widehat{\Phi}) = \mathcal{N}(\Phi) = \mathcal{N}(\Phi(\alpha))$. □

7. Concluding Remarks. The spectral properties of the alignment matrix play an essential role in using local methods for manifold learning. The results proved in this paper represent the first step towards a better understanding of those spectral properties and their interplay with the geometric properties of the set of the local neighborhoods. There are still several issues that deserve further investigation. One of the issues is how to derive a set of conditions which are both necessary and sufficient for the null space of the alignment matrix to recover the parameter vectors. In a sense, the problem is akin to the graph rigidity problem [7]. The other issue is how to improve the quantitative results proved in section 6 under more general conditions, i.e., for cases $s > 2$.

Several algorithmic implications of our analysis can be further explored. Larger overlaps among the local neighborhoods tend to give better conditioned null spaces of the alignment matrices and thus larger sections are favored. However, the accuracy of the local linear fitting methods used in LLE or LTSA generally will suffer on large sections, especially in high curvature regions. One possibility is to use Isomap [12] or a high-order local fitting for larger sections. The alignment matrix framework used in this paper is quite versatile and can serve as the basis for incorporating several kinds of prior information in manifold learning, for example, we may know a priori, the low dimensional parameters $\{\tau_i\}$ for a subset of the sample points [14]. Those possibilities will be further investigated in future research.

Acknowledgement. We thank the anonymous referees for their detailed and insightful comments, suggestions and corrections that greatly improved the presentation of the paper. We also thank one of the referees for the ideas resulting in the simplified proof in Theorem 6.3.

REFERENCES

[1] M. Belkin and P. Niyogi. Laplacian eigenmaps and spectral techniques for embedding and clustering. *Advances in Neural Information Processing Systems* 14, MIT Press, 2002.

[4] Here $\mathcal{R}(A)$ is the range space of A. It is also the orthogonal complement of $\mathcal{N}(A^T)$.

[2] M. Brand. Charting a manifold. *Advances in Neural Information Processing Systems* 15, MIT Press, 2003.
[3] M. Brand. From subspace to submanifold methods. *British Machine Vision Conference* (BMVC), 2004.
[4] M. Brand. Nonrigid Embeddings for Dimensionality Reduction. *European Conference on Machine Learning* (ECML), 2005.
[5] D. Donoho and C. Grimes. Hessian eigenmaps: new tools for nonlinear dimensionality reduction. *Proceedings of National Academy of Science*, 100:5591–5596, 2003.
[6] G. H. Golub and C. F Van Loan. *Matrix computations*. Johns Hopkins University Press, Baltimore, Maryland, 3nd edition, 1996.
[7] B. Hendrickson. Conditions for unique graph realizations. *SIAM J. Computing*, 21:65–84, 1992.
[8] E. Levina and P. Bickel. Maximum Likelihood Estimation of Intrinsic Dimension. *Advances in Neural Information Processing Systems 17*, MIT Press, Cambridge, MA, 2005.
[9] J. Munkres. Analysis on Manifolds. Westview Press, 1991.
[10] S. Roweis and L. Saul. Nonlinear dimensionality reduction by locally linear embedding. *Science*, 290: 2323–2326, 2000.
[11] G. W. Stewart and Ji-Guang Sun. *Matrix perturbation theory*. Academic Press, 1990.
[12] J. Tenenbaum, V. De Silva and J. Langford. A global geometric framework for nonlinear dimension reduction. *Science*, 290:2319–2323, 2000.
[13] J. Wang and Z. Zhang. MLLE: modified locally linear embedding using multiple weights. *Advances in Neural Information Processing Systems 19*, MIT Press, Cambridge, MA, 2007.
[14] X. Yang, H. Fu, H. Zha and J. Barlow. Semi-Supervised nonlinear dimensionality reduction. *Proceedings of International Conference on Machine Learning* (ICML), 2006.
[15] X. Yang, S. Michea and H. Zha. Conical dimension as an intrisic dimension estimator. *Proceedings of SIAM International Conference on Data Mining*, 2007.
[16] H. Zha and Z. Zhang. Spectral Analysis of Alignment in Manifold Learning. *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing* (ICASSP), 2005.
[17] Z. Zhang and H. Zha. Principal manifolds and nonlinear dimensionality reduction via tangent space alignment. *SIAM J. Scientific Computing*, 26:313–338, 2004.
[18] Z. Zhang and H. Zha. A domain decomposition method for fast manifold learning. *Advances in Neural Information Processing Systems 18*, MIT Press, Cambridge, MA, 2006.

Appendix A. Alignment matrices in LLE and Laplacian eigenmap. The algorithm LLE [10] solves the following optimization problem

$$\min_{YY^T=I_d} \sum_i \|y_i - \sum_j y_j w_{ji}\|_2^2, \tag{1.1}$$

where $Y = [y_1, \ldots, y_N] \in \mathcal{R}^{d\times N}$ and $\{w_{ji}\}$ are the local weights determined by the optimal linear combination of x_i using its neighbors (not including x_i itself) with the constraint $\sum_j w_{ji} = 1$. Using the notation for the index set $J_i = \{i_1, \ldots, i_k\}$ of the neighbors including x_i itself, we can write

$$y_i - \sum_j y_j w_{ji} = Y S_i w_i, \quad w_i \in \mathcal{R}^k,$$

where the t-th component of w_i, $t = 1, \ldots, k$, is 1 if $i = i_t$, or $-w_{ji}$ if $j = i_t \neq i$. So

$$\|y_i - \sum_j y_j w_{ji}\|_2^2 = \mathrm{tr}(Y S_i w_i w_i^T S_i^T Y^T) = \mathrm{tr}(Y S_i \Phi_i^{LLE} S_i^T Y^T)$$

with $\Phi_i^{LLE} = w_i w_i^T$, a rank-one semidefinite matrix. So LLE minimizes the trace of $Y\Phi^{LLE}Y^T$ under the normalization constraint $YY^T = I_d$ with the alignment matrix

$$\Phi^{LLE} = \sum_i S_i \Phi_i^{LLE} S_i^T.$$

Obviously, Φ_i^{LLE} has null vector e.

In Laplacian eigenmap [1], the optimal problem solved is

$$\min_{Y\Delta Y^T=I_d} \sum_i \sum_{j\in J_i} \|y_i - y_j\|_2^2 w_{ji}, \tag{1.2}$$

where w_{ji} are also local positive weights and $\Delta = \mathrm{diag}(\delta_1, \ldots, \delta_N)$ with $\delta_i = \sum_{j\in J_i} w_{j,i}$. We have with $Y_i = [y_{i_1}, \ldots, y_{i_k}]$

$$\begin{aligned}\sum_{j\in J_i} \|y_i - y_j\|_2^2 w_{ji} &= \|[(y_i - y_{i_1})\sqrt{w_{i_1,i}}, \ldots, (y_i - y_{i_k})\sqrt{w_{i_k,i}}]\|_F^2 \\ &= \|Y_i(e_t e^T - I)D_i^{1/2}\|_F^2 \\ &= \mathrm{tr}(YS_i(e_t e^T - I)D_i(e_t e^T - I)^T S_i^T Y^T),\end{aligned}$$

where, as before, t the index such that $i = i_t$ and $D_i = \mathrm{diag}(w_{i_1,i}, \ldots, w_{i_k,i})$, Thus we can represent the problem (1.2) as

$$\min_{YDY^T=I_d} \mathrm{tr}\Big(Y\big(\sum_i S_i(e_t e^T - I)D_i(e_t e^T - I)^T S_i^T\big)Y^T\Big).$$

Denoting $Z = Y\Delta^{1/2}$, the above problem is equivalent to

$$\min_{ZZ^T=I_d} \mathrm{tr}\Big(Z\big(\sum_i \Delta^{-1/2} S_i(e_t e^T - I)D_i(e_t e^T - I)^T S_i^T \Delta^{-1/2}\big)Z^T\Big).$$

We can write $\Delta^{-1/2}S_i = S_i\Delta_i^{-1/2}$ with $\Delta_i = \mathrm{diag}(\delta_{i_1}, \ldots, \delta_{i_k})$. The optimization problem now reads

$$\min_{ZZ^T=I_d} \mathrm{tr}(Z\Phi^{Lap}Z^T), \quad \Phi^{Lap} = \sum_i S_i\Phi_i^{Lap}S_i^T$$

is the alignment matrix with the local ones

$$\Phi_i^{Lap} = \Delta_i^{-1/2}(e_t e^T - I)D_i(e_t e^T - I)^T\Delta_i^{-1/2}.$$

Note that is the normalization constraint $Y\Delta Y^T = I_d$ is replaced by the orthogonal normalization $YY^T = I_d$, then $\Phi_i^{Lap} = (e_t e^T - I)D_i(e_t e^T - I)^T$ and Φ_i^{Lap} has a null vector e as Φ_i^{LLE}.

Appendix B. Existence of Affinely Rigid Collections. The following proposition shows the existence of an affinely rigid collection with $|G^*| = s$ for any s and d satisfying $3 \le s \le d+1$.

PROPOSITION B.1. *Let s and d satisfy $3 \le s \le d+1$. Then there is an affinely rigid collection in d-dimensional space such that $|G^*| = s$.*

Proof. We construct the required collection with sd points in the d-dimensional space $\mathcal{R}^d$, explicitly. Let $a_1, \cdots, a_{d-1} \in \mathcal{R}^d$ be $d-1$ linearly independent vectors orthogonal to $e \in \mathcal{R}^d$, and let $b_1, \cdots, b_s \in \mathcal{R}^d$ be s vectors such that $b_2 - b_1, \cdots, b_s - b_1$ are linearly independent and each $b_i - b_j$ has no zero components for $i \ne j$. Now we have sd different points

$$[a_{i1}, a_{i2}, \cdots, a_{i,d-1}, b_{ij}]^T, \quad i = 1, \cdots, d, \quad j = 1, \cdots, s,$$

that form T, where $a_k = [a_{1k}, \cdots, a_{d,k}]^T$ and $b_k = [b_{1k}, \cdots, b_{d,k}]^T$. Denoting $A = [a_1, \cdots, a_{d-1}]$, we see that $T = \left[\begin{bmatrix} A^T \\ b_1^T \end{bmatrix}, \begin{bmatrix} A^T \\ b_2^T \end{bmatrix}, \cdots, \begin{bmatrix} A^T \\ b_s^T \end{bmatrix}\right]$. We consider the collection of the s sections

$$T_i = \left[\begin{bmatrix} A^T \\ b_{i-1}^T \end{bmatrix}, \begin{bmatrix} A^T \\ b_i^T \end{bmatrix}\right], \quad i = 1, \ldots, s,$$

with $b_0 = b_s$, each has $2d$ points. This collection has the overlapping $T_{i,i+1} = \begin{bmatrix} A^T \\ b_i^T \end{bmatrix}$, $i = 1, \ldots, s-1$, $T_{s,1} = \begin{bmatrix} A^T \\ b_s^T \end{bmatrix}$, and other T_{ij}'s, $i \neq j$, are empty. Denoting $B = [e, A]$, a nonsingular matrix of order d, we see that $[e, T_i^T] = \begin{bmatrix} B & b_{i-1} \\ B & b_i \end{bmatrix}$ are of full column rank. However, for nonempty intersections $T_{i,j}$ of the sections, none of the matrices

$$[e, T_{i,i+1}^T] = [B, b_i], \quad i = 1, \ldots, s-1, \quad [e, T_{s,1}^T] = [B, b_s] \tag{2.1}$$

has full column rank since the column number is larger than the row number. Hence the associated graph $G = G^*$ and $|G^*| = s$.

Now we consider the collection rigidness. Let $w = \{w_1, \ldots, w_s\}$ be a certificate of the collection. The overlap conditions (5.3) now become

$$[e, T_{i,i+1}^T](w_i - w_{i+1}) = 0, \quad i = 1, \ldots, s-1, \quad [e, T_{s,1}^T](w_s - w_1) = 0. \tag{2.2}$$

We partition

$$w_i - w_{i+1} = \begin{bmatrix} \delta_i \\ \eta_i \end{bmatrix}, \ i = 1, \ldots, s-1, \quad w_s - w_1 = \begin{bmatrix} \delta_s \\ \eta_s \end{bmatrix}$$

with $\delta_i \in \mathcal{R}^d$, $\eta_i \in \mathcal{R}$, conforming to the column partition in (2.1). Substituting them and (2.1) into (2.2), we obtain

$$B\delta_i = -\eta_i b_i, \quad i = 1, \ldots, s. \tag{2.3}$$

It follows from $\sum_i \delta_i = 0$ and $\sum \eta_i = 0$ that

$$\sum_i \eta_i (b_i - b_1) = -B \sum_i \delta_i = 0,$$

which implies that $\eta_2 = \ldots = \eta_s = 0$ since $b_2 - b_1, \ldots, b_s - b_1$ are linearly independent, giving $\eta_1 = 0$ since $\sum \eta_i = 0$. By (2.3), $\delta_i = 0$, $i = 1, \ldots, s$, because B is nonsingular. It follows that w must be trivial. Therefore, the collection is affinely rigid. □

Appendix C. Geometric conditions for affinely rigid collections. We now give geometric conditions characterizing affinely rigid collections with $|D^*| = 3$ and $d = 2$. Consider a collection of three sections $\{T_1, T_2, T_3\}$ in $\mathcal{R}^2$ such that the associated graph G is already the coarsest, i.e., each $[e, T_i^T]$ is of full column rank while $[e, T_{ij}^T]$ is not of full column rank for $i \neq j$.

If there is a T_{ij} such that it is empty or contains only one point, then the coefficient matrix in (5.5) will be of column rank less than 6; additionally, there is nonzero $[w_1^T, -w_2^T]^T$ satisfying the equation in (5.5). Thus $w = \{w_1, w_2, w_3\}$ with $w_3 = 0$ is a nontrivial certificate of the collection and hence the collection is not affinely rigid. So in the following discussion, we can assume that each intersection T_{ij} contains at least two different points. In fact, all the points in T_{ij} must be on a line segment

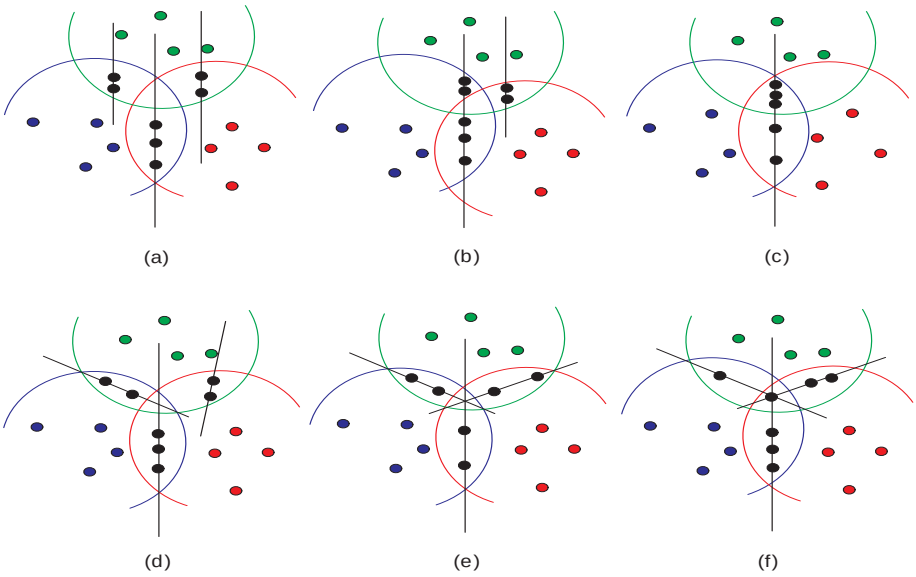

FIG. C.1. *Illustrations of the geometry analysis of the rigidity of collections with* $|G^*| = 3$ *and* $d = 2$.

ℓ_{ij}, because $[e, T_{ij}^T]$ is not of full column rank. Whether the collection is affinely rigid depends on the geometric positions of these line segments as illustrated in Figure C.1 with the following three situations: 1) the line segments are parallel to each others, 2) the line segments are not parallel and they do not meet at a point, and 3) the line segments are not parallel but they meet at a point. We discuss these cases below.

Let $w = \{w_1, w_2, w_3\}$ be a certificate of the collection $\{T_1, T_2, T_3\}$. Because all the points in T_{ij} must be on the line segment ℓ_{ij}, the certificate conditions (5.3) $[1, \tau^T]w_i = [1, \tau^T]w_j$ for $\tau \in T_{ij}$ are equivalent to

$$[1, u_{ij}^T](w_i - w_j) = 0, \quad [0, v_{ij}^T](w_i - w_j) = 0$$

with a point $u_{ij} \in T_{ij}$ and v_{ij}, a vector gives the direction of ℓ_{ij}. Writing the equations above in terms of $z_1 = w_1 - w_3$ and $z_2 = w_2 - w_3$ ($w_1 - w_2 = z_1 - z_2$), we have six equations with respect to z_1 and z_2 for determining the certificate $w = \{w_1, w_2, w_3\}$.

$$[1, u_{12}^T](z_1 - z_2) = 0, \quad [1, u_{13}^T]z_1 = 0, \quad [1, u_{23}^T]z_2 = 0, \tag{3.1}$$

$$[0, v_{12}^T](z_1 - z_2) = 0, \quad [0, v_{13}^T]z_1 = 0, \quad [0, v_{23}^T]z_2 = 0. \tag{3.2}$$

Obviously, w is trivial if and only if $[z_1, z_2]$ is zero.

For case 1) $\ell_{12}, \ell_{13}, \ell_{23}$ are parallel to each others, then $v_{12} = v_{13} = v_{23}$ and the three equations in (3.2) reduce to the two equations $[0, v^T]z_1 = 0$ and $[0, v^T]z_2 = 0$ (v is the common direction for the three line segments). So (3.1-3.2) give at most five equations for the six variables in the two vectors z_1 and z_2. For case 2) $\ell_{12}, \ell_{13}, \ell_{23}$ are not parallel but meet at a point u, we can set $u_{12} = u_{13} = u_{23} = u$. The three equations with respect to u_{ij} in (3.1) reduce to the two equations $[1, u^T]z_1 = 0$ and $[1, u^T]z_2 = 0$. Thus, (3.1-3.2) also give at most five equations for six variables. Therefore, in both the cases, (3.1-3.2) must be underdetermined and have a nonzero solution on (z_1, z_2). Thus, we have a nontrivial certificate and the collection $\{T_1, T_2, T_3\}$ is not affinely rigid.

For case 3) $\ell_{12}, \ell_{13}, \ell_{23}$ are not parallel and they do not meet at a common point. Without loss of generality, assume that ℓ_{13} and ℓ_{23} are not parallel and they meet at τ_0. Of course, $\tau_0 \notin \ell_{12}$. Consider a certificate $w = \{w_1, w_2, w_3\}$. We denote $w_i(\tau) = [1, \tau^T]w_i$ and use the equality $w_i(X) = w_j(X)$ which means that $w_i(x) = w_j(x)$ for all $x \in X$. By (5.3),

$$w_1(\ell_{12}) = w_2(\ell_{12}), \quad w_1(\ell_{13}) = w_3(\ell_{13}), \quad w_2(\ell_{23}) = w_3(\ell_{23}). \tag{3.3}$$

The last two equalities imply that

$$w_1(\tau_0) \overset{\tau_0 \in \ell_{13}}{=} w_3(\tau_0) \overset{\tau_0 \in \ell_{23}}{=} w_2(\tau_0).$$

Together with $w_1(\ell_{12}) = w_2(\ell_{12})$, we see that $w_1(\tau) = w_2(\tau)$ holds for at least three points which are not co-linear. We conclude that $w_1 = w_2$. It follows that $w_1(\ell_{13} \cup \ell_{23}) = w_3(\ell_{13} \cup \ell_{23})$. Since $\ell_{13} \neq \ell_{23}$, we also have that $w_1 = w_3$. Therefore, w must be trivial and the collection is affinely rigid.

We summarize the above discussions in the following result.

THEOREM C.1. *The collection with $|G^*| = 3$ and $d = 2$ is affinely rigid if and only if the three line segments $\ell_{12}, \ell_{13}, \ell_{23}$ corresponding to the super sections of the associated coarsest graph are not parallel to each other and they do not meet at one point.*