EIGENVALUES OF AN ALIGNMENT MATRIX IN NONLINEAR MANIFOLD LEARNING

CHI-KWONG LI *, REN-CANG LI †, AND QIANG YE ‡

Abstract. Alignment algorithm is an effective method recently proposed for nonlinear manifold learning (or dimensionality reduction). By first computing local coordinates of a data set, it constructs an *alignment matrix* from which a global coordinate is obtained from its null space. In practice, the local coordinates can only be constructed approximately and so is the alignment matrix. This together with roundoff errors requires that we compute the the eigenspace associated with a few smallest eigenvalues of an approximate alignment matrix. For this purpose, it is important to know the first nonzero eigenvalue of the alignment matrix or a lower bound in order to computationally separate the null space. This paper bounds the smallest nonzero eigenvalue, which serves as an indicator of how difficult it is to correctly compute the desired null space of the approximate alignment matrix.

Key words. Smallest nonzero eigenvalues, alignment matrix, overlapped partition, nonlinear manifold learning, dimensionality reduction.

subject classifications. 65F15, 15A18.

1. Introduction

Given an $N\times\ell$ matrix Z and s submatrices $Z_j\in\mathbb{C}^{k_j\times\ell}$ (for $1\le j\le s$) consisting of certain rows of Z, let P_{Z_j} be the orthogonal projector in $\mathbb{C}^{k_j}$ onto the column space of Z_j, and $P_{Z_j}^{\perp}=I-P_{Z_j}$. Embed $P_{Z_j}^{\perp}$ into $\mathbb{C}^{N\times N}$ according to the position of the rows of Z_j in Z and denote the resulting $N\times N$ matrix by Φ_j (see (2.3) in Section 2 for details). The matrix

$$P\equiv\sum_{j=1}^{s}\Phi_j. \tag{1.1}$$

is called an *alignment matrix*. This definition is abstracted from and slightly more general than the one in [9], where Z's first column is all ones. Nonetheless most analysis and the results there regarding the null space of P can be carried over in a straightforward way. For example, it is proved under a condition called *fully overlap* among $\{Z_j\}$ that the null space of P is the span of Z [9]. With this property of the alignment matrix, we can reconstruct the rows of Z, up to a linear transformation, from the local projectors P_{Z_j}. This forms a theoretical basis for the LTSA (Local Tangent Space Alignment) algorithm of [11] recently developed for the problem of nonlinear manifold learning.

In nonlinear manifold learning [5, 8], one is concerned with determining a suitable parametrization for a set of given high-dimensional data points lying in a nonlinear

*Department of Mathematics, College of William and Mary, Williamsburg, VA 23187-8795 (`ckli@math.wm.edu`). C.-K. Li is an honorary professor of the University of Hong Kong, and an honorary professor the Heilongjiang University. His research was partially supported by a USA NSF grant, and a HK RGC grant.

†Department of Mathematics, University of Texas at Arlington, P.O. Box 19408, Arlington, TX 76019-0408 (`rcli@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.

‡Department of Mathematics, University of Kentucky, Lexington, KY 40506 (`qye@ms.uky.edu`). Supported in part by the National Science Foundation under Grants CCR-0098133 and DMS-0411502.

manifold, which is also known as (nonlinear) dimensionality reduction. Several methods have been proposed recently for this problem [1, 3, 5, 7, 11]. The alignment matrix was first introduced in the LTSA method [11] in which a local coordinate system is first constructed for a small neighborhood (i.e., a patch) around each sample points and all local coordinates are then aligned together to arrive at a global coordinate. The process of aligning the local coordinates together is achieved through the alignment matrix. We note, however, that the alignment matrix can be used in a more general setting to align coordinates for subsets of data points that are not necessarily local [9]. In this context, the rows of Z are the unknown global coordinates of the high-dimensional data points and the rows of Z_j correspond to the coordinates of the data points in a subset (e.g., a local patch). Then the rows of Z, up to a linear transformation, is constructed from the projectors P_{Z_j} by computing the null space of the alignment matrix.

In practice, only an approximation of the alignment matrix is available. This together with roundoff and/or data errors require that we compute the eigenspace associated with a few smallest eigenvalues that are considered the perturbations of the zero eigenvalues. However, if the perturbations cause the zero eigenvalues to become as large in magnitude as the smallest nonzero eigenvalue of the alignment matrix, it is not possible to determine how many smallest eigenvalues and which of them should be considered zeros. For this purpose, it is important to investigate the first nonzero eigenvalue of the alignment matrix or a lower bound in order to computationally separate the null space.

This paper presents a lower bound on the smallest nonzero eigenvalue, which serves as an indicator of how difficult it is to correctly compute the desired null space of the alignment matrix. An implication of our bound is that the smallest nonzero eigenvalue depends on the "amount" of overlap among Z_j and hence it is necessary to maintain sufficient overlap among Z_j in practice. Our study is based on an ideal situation, namely P is uncontaminated, while contaminated P in practice likely has no nonzero eigenvalues. Thus such simplification becomes somewhat necessary. Nevertheless our effort here represents a step forward to acquire better understanding towards instructively how much overlaps among Z_j for robust recovery of Z, which, translated into the language of nonlinear manifold learning [9, 11], how much overlaps among local patches for robust recovery of global coordinates.

Our investigation into the null space and eigenvalues of this so-called alignment matrix P, abstracted from and slightly more general than its counterpart in nonlinear manifold learning, may be of interest in its own right from matrix theoretical point of view.

The rest of this paper is organized as follows. In Section 2, we set up our framework to study the alignment matrix. We then derive the lower bound at stages, first for the case $s=2$ in Section 3 and then for the general case in Section 4. We shall also discuss when the fully overlap condition is a necessary condition in Section 4.

Notation. As we have done already, denote by $\mathbb{C}^{m\times n}$ the set of all $m\times n$ complex matrices, $\mathbb{C}^n=\mathbb{C}^{n\times 1}$, and $\mathbb{C}=\mathbb{C}^1$. Denote by I_n the $n\times n$ identity matrix, and sometimes simply I when its size is clear from the context. Let $\|X\|_2$ be the spectral norm of a matrix X, i.e., its largest singular value, and $\mathsf{eig}(X)$ be the set of the eigenvalues of a square X. X^* and X^{T} denote the conjugate transpose, the transpose of a matrix or vector X, respectively. $X\preceq Y$ for two Hermitian matrices X and Y means that $Y-X$ is positive semi-definite, and accordingly $X\succeq Y$ means $Y\preceq X$.

For $1\le i\le j\le n$, $i:j$ is the set of integers from i to j inclusive and $i:i=\{i\}$. For vector u and matrix X, $u_{(j)}$ is u's jth entry, $X_{(i,j)}$ is the (i,j)th entry of X. Moreover, subvector $u_{(\boldsymbol{I})}$ consists of all entries $i\in\boldsymbol{I}$; submatrices $X_{(\boldsymbol{I},\boldsymbol{J})}$, $X_{(\boldsymbol{I},:)}$, and $X_{(:,\boldsymbol{J})}$ consist of intersections of all rows $i\in\boldsymbol{I}$ and all columns $j\in\boldsymbol{J}$, all rows $i\in\boldsymbol{I}$ and all columns, and all rows and all columns $j\in\boldsymbol{J}$, respectively.

2. Alignment Matrix

Material in this section is essentially taken from [9], but stated in a slightly more general term, namely in [9] Z's first column is all ones, which is not required here. Also to allow a bit more generality, we assume all involved numbers are complex, unless otherwise explicitly stated. In the context of nonlinear manifold learning [9], most likely they are real.

Let $Z\in\mathbb{C}^{N\times\ell}$, and $N>\ell$. Suppose $Z_j\in\mathbb{C}^{k_j\times\ell}$ for $1\le j\le s$ are submatrices of Z and each consists of certain rows and all columns. Let $\boldsymbol{I}_j=\{j_1,j_2,\dots,j_{k_j}\}$ be the index set for the rows of Z_j as the rows of Z, i.e.,

$$Z_j=Z_{(\boldsymbol{I}_j,:)}=(I_N)_{(\boldsymbol{I}_j,:)}\times Z\in\mathbb{C}^{k_j\times\ell}. \tag{2.1}$$

Assume throughout this paper

$$\bigcup_{j=1}^{s}\boldsymbol{I}_j=\{1,2,\dots,N\}, \tag{2.2}$$

i.e., each row of Z appears in at least one of the Z_j.

Let P_{Z_j} be the orthogonal projector in $\mathbb{C}^{k_j}$ onto the *column space* $\mathsf{span}(Z_j)$ of Z_j, and $P_{Z_j}^{\perp}=I-P_{Z_j}$ is the orthogonal projector also in $\mathbb{C}^{k_j}$ but onto the orthogonal complement of $\mathsf{span}(Z_j)$. It is known that $P_{Z_j}=Z_jZ_j^{\dagger}$, where $Z_j^{\dagger}$ is the Moore-Penrose inverse [6] of Z_j. In particular

$$P_{Z_j}=Z_j(Z_j^*Z_j)^{-1}Z_j^* \quad \text{if } Z_j \text{ has full column rank.}$$

Let Φ_j be the embedding of $P_{Z_j}^{\perp}$ into $\mathbb{C}^N$, i.e.

$$\Phi_j=\left[(I_N)_{(\boldsymbol{I}_j,:)}\right]^{\mathrm{T}}\times P_{Z_j}^{\perp}\times(I_N)_{(\boldsymbol{I}_j,:)}\in\mathbb{C}^{N\times N}. \tag{2.3}$$

Finally an $N\times N$ matrix P is constructed as

$$P=\sum_{j=1}^{s}\Phi_j. \tag{2.4}$$

It can be verified that $PZ=0$, i.e, $\mathsf{span}(Z)\subset\mathsf{null}(P)$, the *null space* of P.

DEFINITION 2.1. *This definition is recursive.*

1. Z_i *always* fully overlaps *itself regardless of its rank;*
2. Z_i *and* Z_j *for* $i\ne j$ *are* fully overlapped, *if* $Z_{(\boldsymbol{I}_i\bigcap\boldsymbol{I}_j,:)}$ *has full column rank;*
3. *The collection* $\boldsymbol{Z}=\{Z_j,1\le j\le s\}$ *for* $s\ge 3$ *is* fully overlapped, *if it can be partitioned into two nonempty disjoint subsets* $\boldsymbol{Z}_1$ *and* $\boldsymbol{Z}_2$ *each of which is a fully overlapped collection and that* $Z_{(\widetilde{\boldsymbol{I}}_1,:)}$ *and* $Z_{(\widetilde{\boldsymbol{I}}_2,:)}$ *are fully overlapped, where*

$$\widetilde{\boldsymbol{I}}_i=\bigcup_{Z_j\in\boldsymbol{Z}_i}\boldsymbol{I}_j. \tag{2.5}$$

This definition is rather general. For example it encompasses the following case: the partitioning graph of Z is connected. By the partitioning graph of Z we mean a graph whose vortices are submatrices Z_j and there is an edge connecting two vortices Z_i and Z_j if and only if Z_i and Z_j for $i \neq j$ are fully overlapped.

THEOREM 2.2. *Assume* (2.2) *holds. If* $\{Z_j, 1 \le j \le s\}$ *is fully overlapped, then* $\mathsf{null}(P) = \mathsf{span}(Z)$.

In nonlinear manifold learning, Z_j is not known but an approximations to $P_{Z_j}^{\perp}$ can be computed; so is an approximation to P whose eigenspace associated with a few smallest eigenvalues will then give an approximation to the column space of Z. This theorem says if $P_{Z_j}^{\perp}$ is exactly known, the column space of Z can be recovered exactly as $\mathsf{null}(P)$. Theorem 2.2 is an extension of the main result in [9] and can be proved by a minor modification to the argument in [9]. Later our method for deriving the eigenvalue bound will lead to another proof of the result.

COROLLARY 2.3. *Under the conditions of* Theorem 2.2,

$$\lambda_{\min}^{+}(P) P_Z^{\perp} \preceq P \preceq \lambda_{\max}(P) P_Z^{\perp},$$

where $\lambda_{\min}^{+}(P)$ *is the smallest nonzero eigenvalue of* P, *and* $\lambda_{\max}(P)$ *is the largest eigenvalue of* P.

Proof. Since P is Hermitian, it has eigen-decomposition $P = U\Lambda U^*$, where U is unitary and Λ is diagonal with the last ℓ diagonal entries being zero. Thus

$$\lambda_{\min}^{+}(P) \times \begin{array}{c} {\scriptstyle N-\ell} \\ {\scriptstyle \ell} \end{array} \!\!\overset{\begin{array}{cc} {\scriptstyle N-\ell} & {\scriptstyle \ell} \end{array}}{\begin{pmatrix} I & \\ & 0 \end{pmatrix}} \preceq \Lambda \preceq \lambda_{\max}(P) \times \begin{array}{c} {\scriptstyle N-\ell} \\ {\scriptstyle \ell} \end{array} \!\!\overset{\begin{array}{cc} {\scriptstyle N-\ell} & {\scriptstyle \ell} \end{array}}{\begin{pmatrix} I & \\ & 0 \end{pmatrix}},$$

which implies

$$\lambda_{\min}^{+}(P) U_{(:,1:N-\ell)} [U_{(:,1:N-\ell)}]^* \preceq P \preceq \lambda_{\max}(P) U_{(:,1:N-\ell)} [U_{(:,1:N-\ell)}]^*.$$

Notice $\mathsf{null}(P) = \mathsf{span}(Z) = \mathsf{null}(P_Z^{\perp})$ by Theorem 2.2 to conclude that $\mathsf{span}(U_{(:,1:N-\ell)}) = \mathsf{span}(P_Z^{\perp})$ and thus $P_Z^{\perp} = U_{(:,1:N-\ell)} [U_{(:,1:N-\ell)}]^*$. ◻

By construction, it is clear $\lambda_{\max}(P) = \|P\|_2 \le s$. However, there is not much we can say about $\lambda_{\min}^{+}(P)$ at this point. The main contribution of this paper is to present a lower bound of it.

3. The case of two submatrices

Without loss of generality, upon permuting rows of Z we may take

$$Z_1 = \begin{array}{c} {\scriptstyle m_{11}} \\ {\scriptstyle m_{12}} \end{array} \overset{\scriptstyle \ell}{\begin{pmatrix} Z_{11} \\ Z_{12} \end{pmatrix}}, \quad Z_2 = \begin{array}{c} {\scriptstyle m_{21}} \\ {\scriptstyle m_{22}} \end{array} \overset{\scriptstyle \ell}{\begin{pmatrix} Z_{21} \\ Z_{22} \end{pmatrix}}, \tag{3.1}$$

where $Z_{12} = Z_{21}$ is the common part in Z_1 and Z_2, $m_{12} = m_{21}$. Then

$$P = \begin{array}{c} {\scriptstyle m_{11}+m_{12}} \\ {\scriptstyle m_{22}} \end{array} \overset{\begin{array}{cc} {\scriptstyle m_{11}+m_{12}} & {\scriptstyle m_{22}} \end{array}}{\begin{pmatrix} P_{Z_1}^{\perp} & 0 \\ 0 & 0 \end{pmatrix}} + \begin{array}{c} {\scriptstyle m_{11}} \\ {\scriptstyle m_{12}+m_{22}} \end{array} \overset{\begin{array}{cc} {\scriptstyle m_{11}} & {\scriptstyle m_{12}+m_{22}} \end{array}}{\begin{pmatrix} 0 & 0 \\ 0 & P_{Z_2}^{\perp} \end{pmatrix}}. \tag{3.2}$$

Theorem 2.2 says if Z_{12} has full column rank (i.e., Z_1 and Z_2 are fully overlapped), then $\dim \mathsf{null}(P) = \ell$ and in fact $\mathsf{null}(P) = \mathsf{span}(Z)$ which implies P has exactly ℓ zero eigenvalues. We would like to know more about its nonzero eigenvalues, too. We shall start by looking into the eigen-structure of P without assuming Z_1 and Z_2 are fully overlapped and then specialize the results to the fully overlapped case.

3.1. Z_1 **and** Z_2 **not necessarily fully overlapped**

The case when $m_{12}=0$, i.e., there is no overlap at all between Z_1 and Z_2 is not interesting, because then

$$P=\begin{pmatrix} P_{Z_1}^{\perp} & \\ & P_{Z_2}^{\perp} \end{pmatrix}$$

a direct sum of two orthogonal projectors whose eigenvalues are either 1 or 0; the case when either $m_{11}=0$ or $m_{22}=0$, i.e., one of Z_i is part of the other, is not particularly interesting, either, because, say, if $m_{11}=0$, then $P_{Z_2}^{\perp}\le P\le 2P_{Z_2}^{\perp}$. So we shall assume $m_{12}\ge 1$, $m_{11}\ge 1$, and $m_{22}\ge 1$ in the rest of this section. The key idea of our analysis below is to find an $N\times N$ unitary matrix Q so that Q^*PQ has simple structure to allow us to determine the null space and the eigenvalues of P.

THEOREM 3.1. *Assume* $m_{12}\ge 1$*,* $m_{11}\ge 1$*, and* $m_{22}\ge 1$*.* Z_{11}*,* $Z_{12}=Z_{21}$ *and* Z_{22} *admit the following decompositions*

$$Z_{11}=U_2\times \begin{matrix} {\scriptstyle r_2} \\ {\scriptstyle m_{11}-r_2} \end{matrix} \overset{\begin{matrix} {\scriptstyle r_1} & {\scriptstyle r_2} & {\scriptstyle \ell-r_1-r_2} \end{matrix}}{\begin{pmatrix} \widetilde{M}_1 & \Sigma_2 & 0 \\ M_1 & 0 & 0 \end{pmatrix}} \times \overset{\begin{matrix} {\scriptstyle r_1} & {\scriptstyle \ell-r_1} \end{matrix}}{\begin{pmatrix} I & 0 \\ 0 & V_2^* \end{pmatrix}} V_1^*, \tag{3.3}$$

$$Z_{12}=Z_{21}=U_1\times \begin{matrix} {\scriptstyle r_1} \\ {\scriptstyle m_{12}-r_1} \end{matrix} \overset{\begin{matrix} {\scriptstyle r_1} & {\scriptstyle \ell-r_1} \end{matrix}}{\begin{pmatrix} \Sigma_1 & 0 \\ 0 & 0 \end{pmatrix}} V_1^*, \tag{3.4}$$

$$Z_{22}=U_3\times \begin{matrix} {\scriptstyle r_3} \\ {\scriptstyle m_{22}-r_3} \end{matrix} \overset{\begin{matrix} {\scriptstyle r_1} & {\scriptstyle r_3} & {\scriptstyle \ell-r_1-r_3} \end{matrix}}{\begin{pmatrix} \widetilde{M}_2 & \Sigma_3 & 0 \\ M_2 & 0 & 0 \end{pmatrix}} \times \overset{\begin{matrix} {\scriptstyle r_1} & {\scriptstyle \ell-r_1} \end{matrix}}{\begin{pmatrix} I & 0 \\ 0 & V_3^* \end{pmatrix}} V_1^*, \tag{3.5}$$

where $U_1(m_{12}\times m_{12})$*,* $U_2(m_{11}\times m_{11})$*,* $U_3(m_{22}\times m_{22})$*,* V_1 $(\ell\times\ell)$*, and* V_2 *and* V_3 *(both* $(\ell-r_1)\times(\ell-r_1)$*) are unitary,* Σ_1 *and* Σ_2 *are diagonal with positive diagonal entries. In particular*

$$r_1=\mathsf{rank}(Z_{12}), r_2=\mathsf{rank}((Z_{11}V_1)_{(:,r_1+1:\ell)}), r_3=\mathsf{rank}((Z_{22}V_1)_{(:,r_1+1:\ell)}). \tag{3.6}$$

Proof. Equation (3.4) is the singular value decomposition (SVD) of Z_{12}. Consider the submatrix of the last $\ell-r_1$ columns of $Z_{11}V_1$ and let its SVD be

$$(Z_{11}V_1)_{(:,r_1+1:\ell)}=U_2\times \begin{matrix} {\scriptstyle r_2} \\ {\scriptstyle m_{11}-r_2} \end{matrix} \overset{\begin{matrix} {\scriptstyle r_2} & {\scriptstyle \ell-r_1-r_2} \end{matrix}}{\begin{pmatrix} \Sigma_2 & 0 \\ 0 & 0 \end{pmatrix}} V_2^*. \tag{3.7}$$

Now notice $U_2^*Z_{11}V_1=(U_2^*(Z_{11}V_1)_{(:,1:r_1)}\ \ U_2^*(Z_{11}V_1)_{(:,r_1+1:\ell)})$, together with (3.7), to arrive at (3.3) with $\widetilde{M}_1$ and M_1 being the top r_2 rows and the bottom $m_{11}-r_2$ rows of $U_2^*(Z_{11}V_1)_{(:,1:r_1)}$, respectively. Similarly let the SVD of the submatrix consisting of the last $\ell-r_1$ columns of $Z_{22}V_1$ be

$$(Z_{22}V_1)_{(:,r_1+1:\ell)}=U_3\times \begin{matrix} {\scriptstyle r_3} \\ {\scriptstyle m_{22}-r_2} \end{matrix} \overset{\begin{matrix} {\scriptstyle r_3} & {\scriptstyle \ell-r_1-r_3} \end{matrix}}{\begin{pmatrix} \Sigma_3 & 0 \\ 0 & 0 \end{pmatrix}} V_3^* \tag{3.8}$$

to lead to (3.5) with $\widetilde{M}_2$ and M_2 being the top r_3 rows and the bottom $m_{22}-r_3$ rows of $U_3^*(Z_{22}V_1)_{(:,1:r_1)}$, respectively. ∎

In what follows, we shall use $X \overset{\text{cols}}{\Leftrightarrow} Y$ to mean $\mathsf{span}(X)=\mathsf{span}(Y)$ for convenience. Note from (3.4) that

$$Z_{12}=Z_{21}=U_1\times \begin{array}{r}\\ {\scriptstyle r_1}\\ {\scriptstyle m_{12}-r_1}\end{array}\!\!\begin{array}{c}\begin{array}{cc}{\scriptstyle r_1} & {\scriptstyle \ell-r_1}\end{array}\\ \begin{pmatrix}\Sigma_1 & 0\\ 0 & 0\end{pmatrix}\end{array}\times \begin{array}{c}\begin{array}{cc}{\scriptstyle r_1} & {\scriptstyle \ell-r_1}\end{array}\\ \begin{pmatrix} I & 0\\ 0 & V^*\end{pmatrix}\end{array} V_1^*$$

for any $(\ell-r_1)\times(\ell-r_1)$ matrix V. In particular set $V=V_2$ and V_3 to get

$$\begin{aligned}\begin{pmatrix}U_2^* & \\ & U_1^*\end{pmatrix}\begin{pmatrix}Z_{11}\\ Z_{12}\end{pmatrix} &\overset{\text{cols}}{\Leftrightarrow} \begin{array}{r}{\scriptstyle r_2}\\ {\scriptstyle m_{11}-r_2}\\ {\scriptstyle r_1}\\ {\scriptstyle m_{12}-r_1}\end{array}\!\!\overset{\begin{array}{cc}{\scriptstyle r_1} & {\scriptstyle r_2}\end{array}}{\begin{pmatrix}\widetilde{M}_1 & \Sigma_2\\ M_1 & 0\\ \Sigma_1 & 0\\ 0 & 0\end{pmatrix}} \overset{\text{cols}}{\Leftrightarrow} \begin{array}{r}{\scriptstyle r_2}\\ {\scriptstyle m_{11}-r_2}\\ {\scriptstyle r_1}\\ {\scriptstyle m_{12}-r_1}\end{array}\!\!\overset{\begin{array}{cc}{\scriptstyle r_1} & {\scriptstyle r_2}\end{array}}{\begin{pmatrix}\widetilde{W}_1 & I\\ W_1 & 0\\ I & 0\\ 0 & 0\end{pmatrix}}\\ &\overset{\text{cols}}{\Leftrightarrow} \widetilde{Z}_1 \overset{\text{def}}{=} \begin{array}{r}{\scriptstyle r_2}\\ {\scriptstyle m_{11}-r_2}\\ {\scriptstyle r_1}\\ {\scriptstyle m_{12}-r_1}\end{array}\!\!\overset{\begin{array}{cc}{\scriptstyle r_1} & {\scriptstyle r_2}\end{array}}{\begin{pmatrix}0 & I\\ W_1 & 0\\ I & 0\\ 0 & 0\end{pmatrix}}\begin{pmatrix}R_1 & \\ & I\end{pmatrix},\end{aligned} \tag{3.9}$$

where

$$W_1=M_1\Sigma_1^{-1},\quad R_1=(I+W_1^*W_1)^{-1/2}. \tag{3.10}$$

Set

$$\widetilde{Z}_1^{\perp}= \begin{array}{r}{\scriptstyle r_2}\\ {\scriptstyle m_{11}-r_2}\\ {\scriptstyle r_1}\\ {\scriptstyle m_{12}-r_1}\end{array}\!\!\overset{\begin{array}{cc}{\scriptstyle m_{11}-r_2} & {\scriptstyle m_{12}-r_1}\end{array}}{\begin{pmatrix}0 & 0\\ I & 0\\ -W_1^* & 0\\ 0 & I\end{pmatrix}}\begin{pmatrix}D_1 & \\ & I\end{pmatrix}\qquad \text{with } D_1=(I+W_1W_1^*)^{-1/2}.$$

Then $(\widetilde{Z}_1 \quad \widetilde{Z}_1^{\perp})$ is unitary. Thus, the column space of $\widetilde{Z}_1^{\perp}$ is a basis for the orthogonal complement of $\mathsf{span}(\widetilde{Z}_1)$ in $\mathbb{C}^{k_1}$. Similarly,

$$\begin{pmatrix}U_1^* & \\ & U_3^*\end{pmatrix}\begin{pmatrix}Z_{21}\\ Z_{22}\end{pmatrix} \overset{\text{cols}}{\Leftrightarrow} \begin{array}{r}{\scriptstyle r_1}\\ {\scriptstyle m_{12}-r_1}\\ {\scriptstyle r_3}\\ {\scriptstyle m_{22}-r_3}\end{array}\!\!\overset{\begin{array}{cc}{\scriptstyle r_1} & {\scriptstyle r_3}\end{array}}{\begin{pmatrix}\Sigma_1 & 0\\ 0 & 0\\ \widetilde{M}_2 & \Sigma_3\\ M_2 & 0\end{pmatrix}} \overset{\text{cols}}{\Leftrightarrow} \widetilde{Z}_2 = \begin{array}{r}{\scriptstyle r_1}\\ {\scriptstyle m_{12}-r_1}\\ {\scriptstyle r_3}\\ {\scriptstyle m_{22}-r_3}\end{array}\!\!\overset{\begin{array}{cc}{\scriptstyle r_1} & {\scriptstyle r_3}\end{array}}{\begin{pmatrix}I & 0\\ 0 & 0\\ 0 & I\\ W_2 & 0\end{pmatrix}}\begin{pmatrix}R_2 & \\ & I\end{pmatrix}, \tag{3.11}$$

where

$$W_2=M_2\Sigma_1^{-1},\quad R_2=(I+W_2^*W_2)^{-1/2}, \tag{3.12}$$

and

$$\widetilde{Z}_2^{\perp}= \begin{array}{r}{\scriptstyle r_1}\\ {\scriptstyle m_{12}-r_1}\\ {\scriptstyle r_3}\\ {\scriptstyle m_{22}-r_3}\end{array}\!\!\overset{\begin{array}{cc}{\scriptstyle m_{22}-r_3} & {\scriptstyle m_{12}-r_1}\end{array}}{\begin{pmatrix}-W_2^* & 0\\ 0 & I\\ 0 & 0\\ I & 0\end{pmatrix}}\begin{pmatrix}D_2 & \\ & I\end{pmatrix}\qquad \text{with } D_2=(1+W_2W_2^*)^{-1/2}$$

has orthonormal columns spanning the orthogonal complement of $\mathsf{span}(\widetilde{Z}_2)$ in $\mathbb{C}^{k_2}$. Let

$$G_1 = \begin{matrix} {\scriptstyle m_{11}+m_{12}} \\ {\scriptstyle m_{22}} \end{matrix} \begin{pmatrix} \widetilde{Z}_1^{\perp} \\ 0 \end{pmatrix}, \quad G_2 = \begin{matrix} {\scriptstyle m_{11}} \\ {\scriptstyle m_{12}+m_{22}} \end{matrix} \begin{pmatrix} 0 \\ \widetilde{Z}_2^{\perp} \end{pmatrix},$$

and

$$G = (G_1 \quad G_2) = \begin{matrix} \\ {\scriptstyle r_2} \\ {\scriptstyle m_{11}-r_2} \\ {\scriptstyle r_1} \\ {\scriptstyle m_{12}-r_1} \\ {\scriptstyle r_3} \\ {\scriptstyle m_{22}-r_3} \end{matrix} \begin{matrix} {\scriptstyle m_{11}-r_2} & {\scriptstyle m_{12}-r_1} & {\scriptstyle m_{22}-r_3} & {\scriptstyle m_{12}-r_1} \\ \left(\begin{matrix} 0 \\ D_1 \\ -W_1^* D_1 \\ 0 \\ 0 \\ 0 \end{matrix}\right. & \begin{matrix} 0 \\ 0 \\ 0 \\ I \\ 0 \\ 0 \end{matrix} & \begin{matrix} 0 \\ 0 \\ -W_2^* D_2 \\ 0 \\ 0 \\ D_2 \end{matrix} & \left.\begin{matrix} 0 \\ 0 \\ 0 \\ I \\ 0 \\ 0 \end{matrix}\right) \end{matrix}.$$

Set

$$Q \stackrel{\text{def}}{=} \begin{matrix} \\ {\scriptstyle m_{11}} \\ {\scriptstyle m_{12}} \\ {\scriptstyle m_{22}} \end{matrix} \begin{matrix} {\scriptstyle m_{11}} & {\scriptstyle m_{12}} & {\scriptstyle m_{22}} \\ \left(\begin{matrix} U_2 \\ \\ \\ \end{matrix}\right. & \begin{matrix} \\ U_1 \\ \\ \end{matrix} & \left.\begin{matrix} \\ \\ U_3 \end{matrix}\right) \end{matrix}. \tag{3.13}$$

Then Q is a unitary matrix, and

$$\widetilde{P} \stackrel{\text{def}}{=} Q^* P Q = Q^* \Phi_1 Q + Q^* \Phi_2 Q = G_1 G_1^* + G_2 G_2^* = G G^*.$$

Note also that the null space of $\widetilde{P}$ is the same as the null space of G^*, which is the same as the orthogonal complement of the column space of G. Let

$$G_3 = \begin{matrix} \\ {\scriptstyle r_2} \\ {\scriptstyle m_{11}-r_2} \\ {\scriptstyle r_1} \\ {\scriptstyle m_{12}-r_1} \\ {\scriptstyle r_3} \\ {\scriptstyle m_{22}-r_3} \end{matrix} \begin{matrix} {\scriptstyle r_2} & {\scriptstyle r_1} & {\scriptstyle r_3} \\ \left(\begin{matrix} I \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{matrix}\right. & \begin{matrix} 0 \\ W_1 \\ I \\ 0 \\ 0 \\ W_2 \end{matrix} & \left.\begin{matrix} 0 \\ 0 \\ 0 \\ 0 \\ I \\ 0 \end{matrix}\right) \end{matrix}.$$

Note that in G, the 4th block column is the same as the 2nd one, and the first 3 block columns are linear independent. Therefore $\mathsf{rank}(G) = m_{11} + m_{12} + m_{22} - (r_1 + r_2 + r_3)$ which implies $\dim \mathsf{null}(G^*) = r_1 + r_2 + r_3$. Evidently, $\mathsf{rank}(G_3) = r_1 + r_2 + r_3$. Therefore $\mathsf{null}(\widetilde{P}) = \mathsf{null}(G^*) = \mathsf{span}(G_3)$ because $G^* G_3 = 0$.

THEOREM 3.2. *Let all symbols keep their assignments so far in this section. Then*

1. $\dim \mathsf{null}(P) = \dim \mathsf{null}(\widetilde{P}) = r_1 + r_2 + r_3$;
2. $\mathsf{null}(\widetilde{P})$ *is the column space of* G_3 *and* $\mathsf{null}(P) = Q\,\mathsf{null}(\widetilde{P})$.
3. *Suppose* Z_1 *and* Z_2 *have full column rank. Then* $\mathsf{null}(P) = \mathsf{span}(Z)$ *if and only if* Z_1 *and* Z_2 *are fully overlapped.*

Proof. Only Item 3 needs a proof. If Z_1 and Z_2 are fully overlapped, then $r_1 = \ell$ and $r_2 = r_3 = 0$ which imply $\dim \mathsf{null}(P) = \ell$ by Item 1. Now $\dim \mathsf{span}(Z) = \ell$ and $\mathsf{span}(Z) \subset \mathsf{null}(P)$ as noted before imply $\mathsf{null}(P) = \mathsf{span}(Z)$. Suppose Z_1 and Z_2 are not fully overlapped. Then $r_1 < \ell$. Noticing that $r_1+r_2=\ell=r_1+r_3$ because Z_1 and Z_2 have full column rank, we have $\dim \mathsf{null}(P)=r_1+r_2+r_3>\ell+(\ell-r_1)>\ell$, and thus $\mathsf{null}(P)\neq\mathsf{span}(Z)$. □

REMARK 3.1. The third assertion in Theorem 3.2 was also obtained by Zha and Zhang [10] for Z_j whose first column is all ones.

Now let us look at the eigenvalues of P, which are the same as those of $\widetilde{P}=GG^*$. Apart from additional zeros, they are the same as those of

$$G^*G=\begin{array}{c} \scriptstyle m_{11}-r_2 \\ \scriptstyle m_{12}-r_1 \\ \scriptstyle m_{22}-r_3 \\ \scriptstyle m_{12}-r_1 \end{array}\begin{pmatrix} \overset{m_{11}-r_2}{I} & \overset{m_{12}-r_1}{0} & \overset{m_{22}-r_3}{D_1W_1W_2^*D_2} & \overset{m_{12}-r_1}{0} \\ 0 & I & 0 & I \\ D_2W_2W_1^*D_1 & 0 & I & 0 \\ 0 & I & 0 & I \end{pmatrix}$$

which is permutationally similar to a direct sum of

$$\begin{pmatrix} I & I \\ I & I \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} I & D_1W_1W_2^*D_2 \\ D_2W_2W_1^*D_1 & I \end{pmatrix}.$$

The former matrix has nonzero eigenvalue 2 with multiplicity $m_{12}-r_1$; the latter matrix has eigenvalues $1\pm\sigma_j$ for $j=1,\dots,k$, where $\sigma_1,\dots,\sigma_k$ are the nonzero singular values of $D_1W_1W_2^*D_2$, and the remaining eigenvalues equal to 1. Thus, we have the following.

THEOREM 3.3. *Let the nonzero singular values of $D_2W_2W_1^*D_1$ be $\sigma_1,\sigma_2,\dots,\sigma_k$. Then* $\mathsf{eig}(P)$ *consists of*

$$\begin{array}{ll} 1\pm\sigma_j, & \textit{for } 1\le j\le k, \\ 2, & \textit{with multiplicity } m_{12}-r_1, \\ 1, & \textit{with multiplicity } m_{11}+m_{22}-r_2-r_3-2k, \\ 0, & \textit{with multiplicity } r_1+r_2+r_3. \end{array}$$

We shall now bound the singular values σ_j of $D_2W_2W_1^*D_1$. First we have

$$\begin{aligned} \sigma_j &\le \|D_2W_2W_1^*D_1\|_2 \\ &= \|D_2W_2\|_2\,\|D_1W_1\|_2, \end{aligned} \tag{3.14}$$

$$\|D_iW_i\|_2=\frac{\|W_i\|_2}{\sqrt{1+\|W_i\|_2^2}}. \tag{3.15}$$

It follows from (3.3), (3.4), and (3.10) that

$$Z_{11}Z_{12}^{\dagger}=Z_{11}V_1\begin{pmatrix}\Sigma_1^{-1} & \\ & 0\end{pmatrix}U_1^*=((Z_{11}V_1)_{(:,1:r_1)}\Sigma_1^{-1}\ \ 0)U_1^*,$$

$$U_2^*(Z_{11}V_1)_{(:,1:r_1)}\Sigma_1^{-1}=\begin{pmatrix}\widetilde{M}_1 \\ M_1\end{pmatrix}\Sigma_1^{-1}=\begin{pmatrix}\widetilde{M}_1\Sigma_1^{-1} \\ W_1\end{pmatrix}.$$

They yield

$$\|W_1\|_2\le\|U_2^*(Z_{11}V_1)_{(:,1:r_1)}\Sigma_1^{-1}\|_2=\|Z_{11}Z_{12}^{\dagger}\|_2 \quad \text{with equality if } \widetilde{M}_1=0 \text{ or } r_2=0. \tag{3.16}$$

Since the construction of W_2 is similar to the construction of W_1, one can give a similar bound to W_2, namely,

$$\|W_2\|_2 \le \|Z_{22}Z_{12}^\dagger\|_2 \quad \text{with equality if } \widetilde{M}_2=0 \text{ or } r_3=0. \tag{3.17}$$

Combine (3.14) – (3.17) to get

$$\sigma_j \le \frac{\|Z_{11}Z_{12}^\dagger\|_2}{\sqrt{1+\|Z_{11}Z_{12}^\dagger\|_2^2}} \frac{\|Z_{22}Z_{12}^\dagger\|_2}{\sqrt{1+\|Z_{22}Z_{12}^\dagger\|_2^2}}.$$

We have proved

THEOREM 3.4. *The nonzero eigenvalues of P is no smaller than $1-\tau$ where*

$$\tau \overset{\text{def}}{=} \frac{\|Z_{11}Z_{12}^\dagger\|_2}{\sqrt{1+\|Z_{11}Z_{12}^\dagger\|_2^2}} \frac{\|Z_{22}Z_{12}^\dagger\|_2}{\sqrt{1+\|Z_{22}Z_{12}^\dagger\|_2^2}}. \tag{3.18}$$

Its largest eigenvalue is no greater than $1+\tau$ if $m_{12}=r_1$ and it is 2 if $m_{12}>r_1$.

3.2. Z_1 and Z_2 fully overlapped

When Z_1 and Z_2 are fully overlapped, results in the previous subsection are still valid, only simpler. Here is the list of a few notably changes to what in the previous subsection:

- $r_1=\ell$ and $r_2=r_3=0$;
- Decompositions (3.7) and (3.8) are not needed, and consequently the decompositions in Theorem 3.1 are simplified to

$$Z_{11}=M_1V_1^*,\ Z_{12}=U_1\Sigma_1V_1^*,\ Z_{22}=M_2V_1^*.$$

- (3.9) and (3.11) remain valid with $U_2=I$ and $U_3=I$;
- (3.16) and (3.17) are equalities, and in fact $W_i=Z_{ii}Z_{12}^\dagger U_1^*$ for $i=1,2$;
- Theorems 3.3 and Theorems 3.4 are valid as they are, and furthermore Theorem 3.4 has a stronger version – Theorem 3.5 below.

THEOREM 3.5. *Let τ be defined by* (3.18). *If Z_1 and Z_2 are fully overlapped, then*

$$(1-\tau)P_Z^\perp \preceq P \preceq \left\{\begin{matrix} (1+\tau) & \text{if } m_{12}=\ell, \\ 2 & \text{if } m_{12}>\ell \end{matrix}\right\} P_Z^\perp. \tag{3.19}$$

Furthermore,

$$\lambda_{\min}^+(P) \ge \frac{1}{2}\left(\frac{\sigma_{\min}^2(Z_{12})}{\sigma_{\max}^2(Z_{11})}+\frac{\sigma_{\min}^2(Z_{12})}{\sigma_{\max}^2(Z_{22})}\right)\Bigg/\left(1+\frac{\sigma_{\min}^2(Z_{12})}{\sigma_{\max}^2(Z_{11})}+\frac{\sigma_{\min}^2(Z_{12})}{\sigma_{\max}^2(Z_{22})}\right), \tag{3.20}$$

where $\sigma_{\min}$ and $\sigma_{\max}$ denote the smallest and the largest singular value respectively.

Proof. (3.19) is a consequence of Item 3 of Theorem 3.2 and the proof of Corollary 2.3. From (3.18), we have

$$\begin{aligned}
\lambda_{\min}^{+}(P) &\geq 1-\frac{1}{\sqrt{1+\|Z_{11}\|_2^{-2}\|Z_{12}^{\dagger}\|_2^{-2}}\sqrt{1+\|Z_{22}\|_2^{-2}\|Z_{12}^{\dagger}\|_2^{-2}}} \\
&\geq \frac{(1+\|Z_{11}\|_2^{-2}\|Z_{12}^{\dagger}\|_2^{-2})(1+\|Z_{22}\|_2^{-2}\|Z_{12}^{\dagger}\|_2^{-2})-1}{2(1+\|Z_{11}\|_2^{-2}\|Z_{12}^{\dagger}\|_2^{-2})(1+\|Z_{22}\|_2^{-2}\|Z_{12}^{\dagger}\|_2^{-2})} \\
&\geq \frac{\|Z_{11}\|_2^{-2}\|Z_{12}^{\dagger}\|_2^{-2}+\|Z_{22}\|_2^{-2}\|Z_{12}^{\dagger}\|_2^{-2}}{2(1+\|Z_{11}\|_2^{-2}\|Z_{12}^{\dagger}\|_2^{-2}+\|Z_{22}\|_2^{-2}\|Z_{12}^{\dagger}\|_2^{-2})} \\
&= \frac{\sigma_{\min}^2(Z_{12})/\sigma_{\max}^2(Z_{11})+\sigma_{\min}^2(Z_{12})/\sigma_{\max}^2(Z_{22})}{2\,[1+\sigma_{\min}^2(Z_{12})/\sigma_{\max}^2(Z_{11})+\sigma_{\min}^2(Z_{12})/\sigma_{\max}^2(Z_{22})]},
\end{aligned}$$

as expected. ◻

The theorem implies that if Z_{12} has full column rank but with nearly linearly dependent columns, $\sigma_{\min}(Z_{12})$ is small and the smallest nonzero eigenvalue $\lambda_{\min}^{+}(P)$ may also and can be nearly zero. In particular, $\lambda_{\min}^{+}(P)$ may be of order $\sigma_{\min}^2(Z_{12})$.

Remark 3.2. Independently, Zha and Zhang [10, Theorem 5.1] obtained, in our notation, the following result (original version was for Z_j whose first column is all ones): *Let* $P_{Z_j}^{\perp}=Q_jQ_j^*$ $(j=1,2)$ *and partition*

$$Q_1=\begin{matrix} {\scriptstyle m_{11}} \\ {\scriptstyle m_{12}} \end{matrix}\begin{pmatrix} Q_{11} \\ Q_{12} \end{pmatrix}, \quad Q_2=\begin{matrix} {\scriptstyle m_{21}} \\ {\scriptstyle m_{22}} \end{matrix}\begin{pmatrix} Q_{21} \\ Q_{22} \end{pmatrix},$$

conformally to those in (3.1). *If* Z_1 *and* Z_2 *are fully overlapped, then the smallest nonzero eigenvalue of* P *is given by* $1-\sigma_{\max}(Q_{12}^*Q_{21})$. This will obviously give the same smallest nonzero eigenvalue of P as one can deduce from Theorem 3.3, but since Q_j depends on Z_j in a nontrivial way, i.e., there is no explicit expression to write down Q_j in terms of Z_j, it is not clear if one could establish any lower bound based on $1-\sigma_{\max}(Q_{12}^*Q_{21})$ in terms of Z_j, as we did in Theorem 3.5 based on Theorem 3.3. Zha and Zhang also extended their result for more than two submatrices, the case we will be dealing with in the next section.

Next we give an example to show that the bound in Theorem 3.5 can be asymptotically attained and hence sharp.

Example 3.1. Consider $m_{11}=1=m_{22}$, $m_{12}=2$, and $\ell=2$:

$$Z_1=\begin{pmatrix} Z_{11} \\ Z_{12} \end{pmatrix}\equiv\begin{pmatrix} 1 & a \\ \hline 1 & c_1 \\ 1 & c_2 \end{pmatrix}, \quad Z_2=\begin{pmatrix} Z_{21} \\ Z_{22} \end{pmatrix}\equiv\begin{pmatrix} 1 & c_1 \\ 1 & c_2 \\ \hline 1 & b \end{pmatrix},$$

assuming $c_1\neq c_2$, i.e., $Z_{12}\equiv Z_{21}$ is nonsingular. All numbers are real. Z as such comes from nonlinear manifold learning [9]. Calculation by Maple[1] shows that the characteristic polynomial of P is

$$\Delta_1\Delta_2\lambda^2\left(\lambda^2-2\lambda+\frac{(c_1-c_2)^2\Delta}{\Delta_1\Delta_2}\right),$$

[1] http://www.maplesoft.com/.

where

$$\begin{aligned}\Delta_1 &= (a-c_1)^2+(a-c_2)^2+(c_1-c_2)^2,\\ \Delta_2 &= (b-c_1)^2+(b-c_2)^2+(c_1-c_2)^2,\\ \Delta &= (a-c_1)^2+(a-c_2)^2+(c_1-c_2)^2+(b-c_1)^2+(b-c_2)^2+(a-b)^2.\end{aligned}$$

So there are two zero eigenvalues and two nonzero ones, as expected. The two nonzero eigenvalues are

$$\begin{aligned}&1-\frac{\sqrt{\Delta_1\Delta_2-(c_1-c_2)^2\Delta}}{\sqrt{\Delta_1\Delta_2}}=\frac{(c_1-c_2)^2\Delta}{\sqrt{\Delta_1\Delta_2}+\sqrt{\Delta_1\Delta_2-(c_1-c_2)^2\Delta}},\\ &1+\frac{\sqrt{\Delta_1\Delta_2-(c_1-c_2)^2\Delta}}{\sqrt{\Delta_1\Delta_2}}.\end{aligned} \tag{3.21}$$

Calculations also show

$$\Delta_1\Delta_2-(c_1-c_2)^2\Delta=[(c_2-a)(c_2-b)+(c_1-a)(c_1-b)]^2.$$

Now let us look at what our bounds by Theorem 3.5 say. We have

$$Z_{12}^{\dagger}\equiv Z_{12}^{-1}=\frac{1}{c_2-c_1}\begin{pmatrix}c_2 & -c_1\\ -1 & 1\end{pmatrix},$$

$$Z_{11}Z_{12}^{-1}=\frac{1}{c_2-c_1}(c_2-a \quad a-c_1),\quad Z_{22}Z_{12}^{-1}=\frac{1}{c_2-c_1}(c_2-b \quad b-c_1).$$

The lower and upper bounds by Theorem 3.4 are

$$1\pm\frac{\sqrt{(a-c_1)^2+(a-c_2)^2}}{\sqrt{\Delta_1}}\,\frac{\sqrt{(b-c_1)^2+(b-c_2)^2}}{\sqrt{\Delta_2}} \tag{3.22}$$

which can be verified to be exactly the two values in (3.21) if

$$|(c_2-a)(c_2-b)+(c_1-a)(c_1-b)|=\sqrt{(a-c_1)^2+(a-c_2)^2}\sqrt{(b-c_1)^2+(b-c_2)^2},$$

i.e., when two vectors $(c_2-a \quad a-c_1)$ and $(c_2-b \quad b-c_1)$ are parallel, which happens when $c_1=c_2$. When $c_1=c_2$, however, $\{Z_1,Z_2\}$ is not fully overlapped. But by making $c_1\neq c_2$ while as close as needed, $\{Z_1,Z_2\}$ is fully overlapped and at the same time the lower and upper bounds by Theorem 3.4 can be made as close to the two values in (3.21) as wished.

4. The case of more than two submatrices

In general for $s\geq 3$, our approach in the previous section appears to break down. In what follows, we shall describe a way to recursively bound the smallest nonzero eigenvalue $\lambda_{\min}^{+}(P)$ from below. To do so, we define a function τ which takes two submatrices of Z with all columns as arguments. Given

$$\widetilde{Z}_i=Z_{(\widetilde{\boldsymbol{I}}_i,:)},\quad i=1,2,$$

we define

$$\tau(\widetilde{Z}_1,\widetilde{Z}_2)\stackrel{\rm def}{=}\frac{t_1}{\sqrt{1+t_1^2}}\,\frac{t_2}{\sqrt{1+t_2^2}},\quad t_i=\left\|Z_{(\boldsymbol{J}_i,:)}\,Z^{\dagger}_{(\widetilde{\boldsymbol{I}}_1\bigcap\widetilde{\boldsymbol{I}}_2,:)}\right\|_2, \tag{4.1}$$

where $\boldsymbol{J}_i$ is the complement set of $\widetilde{\boldsymbol{I}}_1 \bigcap \widetilde{\boldsymbol{I}}_2$ in $\widetilde{\boldsymbol{I}}_i$.

Throughout the rest of this section, we adopt in whole the notation associated with Z and Z_j as introduced in Section 2, and we assume that $\boldsymbol{Z} = \{Z_j, 1 \le j \le s\}$ is fully overlapped, and (2.2) holds.

From Definition 2.1, $\boldsymbol{Z}$ can be partitioned into two nonempty disjoint subsets $\boldsymbol{Z}_1$ and $\boldsymbol{Z}_2$ each of which is a fully overlapped collection and that

$$\widetilde{Z}_1 = Z_{(\widetilde{\boldsymbol{I}}_1,:)}, \quad \widetilde{Z}_2 = Z_{(\widetilde{\boldsymbol{I}}_2,:)} \tag{4.2}$$

are fully overlapped, where $\widetilde{\boldsymbol{I}}_1$ and $\widetilde{\boldsymbol{I}}_2$ are defined as in (2.5). By Theorem 3.5, we have

$$\left[1 - \tau(\widetilde{Z}_1, \widetilde{Z}_2)\right] P_Z^{\perp} \preceq \sum_{j=1}^{2} (I_N)^{\mathrm{T}}_{(\widetilde{\boldsymbol{I}}_j,:)} \times P^{\perp}_{\widetilde{Z}_j} \times (I_N)_{(\widetilde{\boldsymbol{I}}_j,:)}.$$

Now recursively bound $P^{\perp}_{\widetilde{Z}_j}$ in exactly the same way because $\boldsymbol{Z}_j$ is fully overlapped until the right-hand side becomes P. The following procedure recursively computes $\alpha(\boldsymbol{Z})$ that satisfies $\alpha(\boldsymbol{Z}) P_Z^{\perp} \le P$:

$$\alpha(\{Z_i\}) = 1, \tag{4.3}$$

$$\alpha(\{Z_i, Z_j\}) = 1 - \tau(Z_i, Z_j), \tag{4.4}$$

$$\alpha(\boldsymbol{Z}) = \left[1 - \tau(\widetilde{Z}_1, \widetilde{Z}_2)\right] \min\{\alpha(\boldsymbol{Z}_1), \alpha(\boldsymbol{Z}_2)\}. \tag{4.5}$$

The smallest nonzero eigenvalue $\lambda^{+}_{\min}(P)$ is then no smaller than $\alpha(\boldsymbol{Z})$.

THEOREM 4.1. *Suppose $\boldsymbol{Z} = \{Z_1, Z_2, \dots, Z_s\}$ is a fully overlapped collection, where Z_j are submatrices of $Z \in \mathbb{C}^{N \times \ell}$ as defined by* (2.1) *and* (2.2). *Let $\alpha(\boldsymbol{Z})$ be computed recursively by* (4.3) – (4.5). *Then $\alpha(\boldsymbol{Z}) P_Z^{\perp} \le P$, where alignment matrix P is defined by* (2.4).

EXAMPLE 4.1. Consider $s = 3$. Suppose Z_1 and $\widetilde{Z}_2 \stackrel{\mathrm{def}}{=} Z_{(\boldsymbol{I}_2 \bigcup \boldsymbol{I}_3,:)}$ are fully overlapped. Let $\widetilde{\boldsymbol{I}}_2 \stackrel{\mathrm{def}}{=} \boldsymbol{I}_2 \bigcup \boldsymbol{I}_3$. Then we have by Theorem 3.5

$$\left[1 - \tau(Z_1, \widetilde{Z}_2)\right] P_Z^{\perp} \preceq (I_N)^{\mathrm{T}}_{(\boldsymbol{I}_1,:)} \times P^{\perp}_{Z_1} \times (I_N)_{(\boldsymbol{I}_1,:)} + (I_N)^{\mathrm{T}}_{(\widetilde{\boldsymbol{I}}_2,:)} \times P^{\perp}_{\widetilde{Z}_2} \times (I_N)_{(\widetilde{\boldsymbol{I}}_2,:)},$$

and

$$[1 - \tau(Z_2, Z_3)]\,(I_N)^{\mathrm{T}}_{(\widetilde{\boldsymbol{I}}_2,:)} \times P^{\perp}_{\widetilde{Z}_2} \times (I_N)_{(\widetilde{\boldsymbol{I}}_2,:)} \preceq \sum_{j=2}^{3} (I_N)^{\mathrm{T}}_{(\boldsymbol{I}_j,:)} \times P^{\perp}_{Z_j} \times (I_N)_{(\boldsymbol{I}_j,:)}.$$

Put the two inequalities together to get $\alpha(\boldsymbol{Z}) P_Z^{\perp} \preceq P$ with

$$\alpha(\boldsymbol{Z}) = \left[1 - \tau(Z_1, \widetilde{Z}_2)\right] [1 - \tau(Z_2, Z_3)].$$

EXAMPLE 4.2. Consider $s = 4$. Suppose Z_1 and Z_2, Z_3 and Z_4, and $\widetilde{Z}_1 \stackrel{\mathrm{def}}{=} Z_{(\boldsymbol{I}_1 \bigcup \boldsymbol{I}_2,:)}$ and $\widetilde{Z}_2 \stackrel{\mathrm{def}}{=} Z_{(\boldsymbol{I}_3 \bigcup \boldsymbol{I}_4,:)}$ are fully overlapped pairs. Let $\widetilde{\boldsymbol{I}}_1 \stackrel{\mathrm{def}}{=} \boldsymbol{I}_1 \bigcup \boldsymbol{I}_2$ and $\widetilde{\boldsymbol{I}}_2 \stackrel{\mathrm{def}}{=} \boldsymbol{I}_3 \bigcup \boldsymbol{I}_4$. Then we have by Theorem 3.5

$$\left[1 - \tau(\widetilde{Z}_1, \widetilde{Z}_2)\right] P_Z^{\perp} \preceq (I_N)^{\mathrm{T}}_{(\widetilde{\boldsymbol{I}}_1,:)} \times P^{\perp}_{\widetilde{Z}_1} \times (I_N)_{(\widetilde{\boldsymbol{I}}_1,:)} + (I_N)^{\mathrm{T}}_{(\widetilde{\boldsymbol{I}}_2,:)} \times P^{\perp}_{\widetilde{Z}_2} \times (I_N)_{(\widetilde{\boldsymbol{I}}_2,:)},$$

and

$$[1-\tau(Z_1,Z_2)]\,(I_N)^{\mathrm{T}}_{(\widetilde{\boldsymbol{I}}_1,:)}\times P^{\perp}_{\widetilde{Z}_1}\times (I_N)_{(\widetilde{\boldsymbol{I}}_1,:)}\preceq\sum_{j=1}^{2}(I_N)^{\mathrm{T}}_{(\boldsymbol{I}_j,:)}\times P^{\perp}_{Z_j}\times (I_N)_{(\boldsymbol{I}_j,:)},$$

$$[1-\tau(Z_3,Z_4)]\,(I_N)^{\mathrm{T}}_{(\widetilde{\boldsymbol{I}}_2,:)}\times P^{\perp}_{\widetilde{Z}_2}\times (I_N)_{(\widetilde{\boldsymbol{I}}_2,:)}\preceq\sum_{j=3}^{4}(I_N)^{\mathrm{T}}_{(\boldsymbol{I}_j,:)}\times P^{\perp}_{Z_j}\times (I_N)_{(\boldsymbol{I}_j,:)}.$$

Put the three inequalities together to get $\alpha(\boldsymbol{Z})P^{\perp}_{Z}\preceq P$ with

$$\alpha(\boldsymbol{Z})=\left[1-\tau(\widetilde{Z}_1,\widetilde{Z}_2)\right]\min\{[1-\tau(Z_1,Z_2)],[1-\tau(Z_3,Z_4)]\}.$$

Example 4.3. This is for the sequentially fully overlapping case, i.e., Z_i and Z_j are fully overlapped if $i-j=\pm1$ and Z_i and Z_j have no overlap at all if $|i-j|\geq 2$. Write

$$Z_j=\begin{matrix} m_{j1}\\ m_{j2}\\ m_{j3}\end{matrix}\overset{\ell}{\begin{pmatrix} Z_{j1}\\ Z_{j2}\\ Z_{j3}\end{pmatrix}},\quad m_{11}=m_{s3}=0,\quad k_j=m_{j1}+m_{j2}+m_{j3},$$

where $Z_{j1}=Z_{j-13}$ is the overlapped part between Z_{j-1} and Z_j. Recursively, we have

$$\begin{aligned}\alpha(\{Z_1,\dots,Z_s\})=&\left[1-\tau(Z_{(1:p+m_{j-13},:)},Z_{(p+1:s,:)})\right]\\ &\times\min\{\alpha(\{Z_1,\dots,Z_{j-1}\}),\alpha(\{Z_j,\dots,Z_s\})\},\end{aligned}$$

where $p=\sum_{i=1}^{j-1}(m_{i1}+m_{i2})$, and

$$t_1=\left\|Z_{(1:p,:)}Z^{\dagger}_{j-13}\right\|_2,\ t_2=\left\|Z_{(p+m_{j1}+1:s,:)}Z^{\dagger}_{j1}\right\|_2,$$

$$\tau(Z_{(1:p+m_{j-13},:)},Z_{(p+1:s,:)})=\frac{t_1}{\sqrt{1+t_1^2}}\,\frac{t_2}{\sqrt{1+t_2^2}}.$$

The ending conditions (4.3) and (4.4) still apply. For shortest recursion, j should be picked about $s/2$, e.g., the smallest integer that is no less than $s/2$. To see how good our recursive bound is, we have tested on random Z and random Z with its first column being all ones to mimic cases from nonlinear manifold learning. Figure 4.1 plots $\lambda^{+}_{\min}(P)$ *vs.* $\alpha(\boldsymbol{Z})$ for $2\leq s\leq 10$ and $2\leq\ell\leq 5$. Our bounds for small s (about $s\leq 4$ here, especially for $s=2$) look pretty good; however, they very much underestimate $\lambda^{+}_{\min}(P)$ for big s (about $s>4$ here).

Example 4.4. Consider $Z\in\mathbb{C}^{N\times 2}$ with the first column $Z_{(:,1)}$ being all ones and the second column $Z_{(:,2)}$ being $1,2,\dots,N$, and $Z_j=Z_{(j:j+2,:)}$ $(1\leq j\leq s=N-2)$. This corresponds to the one-dimensional case in manifold learning with data points equidistantly sit on a straight line. The case is so special that it allows us to estimate more precisely our bound and $\lambda^{+}_{\min}(P)$ through analytical means. It can be seen that

$$P^{\perp}_{Z_j}=\frac{1}{6}\begin{pmatrix}1&-2&1\\-2&4&-2\\1&-2&1\end{pmatrix},\quad P=\frac{1}{6}\begin{pmatrix}1&-2&1&&&&\\-2&5&-4&1&&&\\1&-4&6&-4&1&&\\&\ddots&\ddots&\ddots&\ddots&\ddots&\\&&1&-4&6&-4&1\\&&&1&-4&5&-2\\&&&&1&-2&1\end{pmatrix}.$$

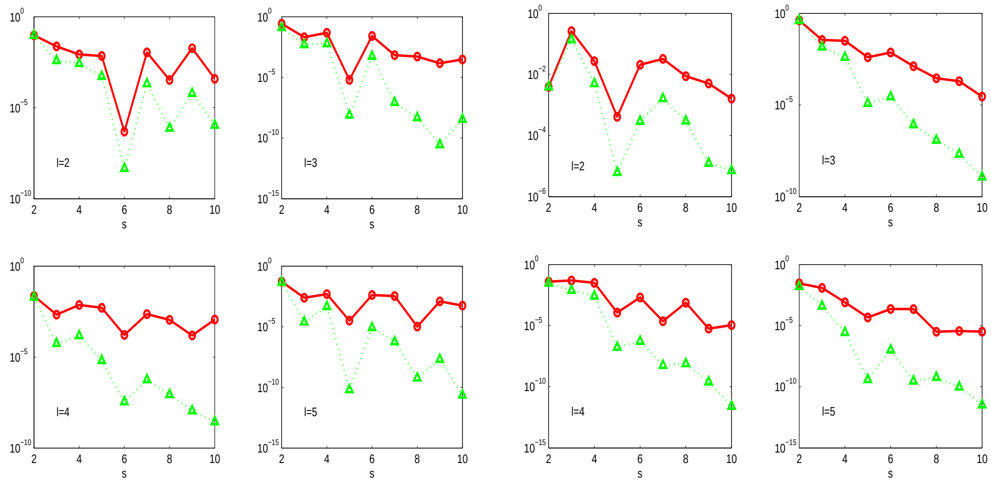

FIG. 4.1. $\lambda^{+}_{\min}(P)$ *(red* $\circ$*)* vs. $\alpha(\boldsymbol{Z})$ *(green* $\triangle$*):* $m_{j1}=m_{j3}=\ell$ *and* $m_{j2}=\ell-1$ *(except* $m_{11}=m_{s3}=0$*).* **Left** *4 plots: random* Z*;* **Right** *4 plots: random* Z *(except* $Z_{(:,1)}$ *all ones).*

It can be verified that $6P=U^{\mathrm{T}}U$, where $U=\big(e_1\ T\ e_s\big)\in\mathbb{C}^{s\times N}$, e_1 and e_s are the first and last column of the identity matrix I_s, and $T\in\mathbb{C}^{s\times s}$ is the famous tridiagonal Toeplitz matrix with diagonal entries -2 and off-diagonal entries 1. Thus $6\lambda^{+}_{\min}(P)$ is the smallest eigenvalues of

$$UU^{\mathrm{T}}=e_1e_1^{\mathrm{T}}+T^2+e_se_s^{\mathrm{T}}\succeq T^2.$$

The eigenvalue system of T is explicitly known [2, 4], and so is T^2: $T=Q\Lambda Q^{\mathrm{T}}$, where $\Lambda=\mathsf{diag}(\lambda_1,\dots,\lambda_s)$ with

$$\lambda_j=-2+2\cos\theta_j,\theta_j=\frac{j}{s+1}\pi,\quad Q_{(i,j)}=\frac{2}{\sqrt{s+1}}\sin\left(\pi\frac{ij}{s+1}\right)$$

for $1\le i,j\le s$. Therefore

$$\lambda^{+}_{\min}(P)\ge\frac{1}{6}(2-2\cos\theta_1)^2=\frac{8}{3}\sin^4\frac{\theta_1}{2}\sim\frac{\pi^4}{6(s+1)^4}=\frac{\pi^4}{6(N-1)^4} \tag{4.6}$$

for large N. We now establish an upper bound on $\lambda^{+}_{\min}(P)$. Note that

$$UU^{\mathrm{T}}=T(I+XX^{\mathrm{T}})T\preceq(1+\|X\|_2^2)T^2,\quad X=(T^{-1}e_1\ \ T^{-1}e_s).$$

This implies

$$\lambda^{+}_{\min}(P)\le\frac{1}{6}(2-2\cos\theta_1)^2(1+\|X\|_2^2). \tag{4.7}$$

We now bound $\|X\|_2^2$. We have

$$\begin{aligned}
\|X\|_2^2 &\le \|T^{-1}e_1\|_2^2+\|T^{-1}e_s\|_2^2=2\|T^{-1}e_1\|_2^2\\
&=\frac{8}{s+1}\sum_{j=1}^{s}\frac{\sin^2\theta_j}{\lambda_j^2}=\frac{2}{s+1}\sum_{j=1}^{s}\cot^2\frac{\theta_j}{2}\\
&\le \frac{2}{s+1}\cot^2\frac{\theta_1}{2}+\frac{4}{\pi}\int_{\frac{\pi}{2(s+1)}}^{\pi/2}\cot^2 t\,dt\\
&=\frac{2}{s+1}\cot^2\frac{\theta_1}{2}+\frac{4}{\pi}\left(-\cot t+\frac{\pi}{2}-t\right)\Big|_{\frac{\pi}{2(s+1)}}^{\pi/2}\\
&=\frac{2}{s+1}\cot^2\frac{\theta_1}{2}+\frac{4}{\pi}\left(\cot\frac{\pi}{2(s+1)}+\frac{\pi}{2(s+1)}-\frac{\pi}{2}\right)\\
&\sim\frac{8(N-1)}{\pi^2},
\end{aligned}$$

for large N. Combine this with (4.7) to get

$$\lambda_{\min}^{+}(P)\le\frac{1}{6}(2-2\cos\theta_1)^2(1+\|X\|_2^2)\sim\frac{4\pi^2}{3(N-1)^3}. \tag{4.8}$$

Next we estimate what we may expect from our bound $\alpha(\boldsymbol{Z})$. It can be seen that a key step in our recursive procedure for $\alpha(\boldsymbol{Z})$ is for

$$\widetilde{Z}_1=\begin{pmatrix}1 & i\\ \vdots & \vdots\\ 1 & m\\ 1 & m+1\end{pmatrix},\quad \widetilde{Z}_2=\begin{pmatrix}1 & m\\ 1 & m+1\\ \vdots & \vdots\\ 1 & j\end{pmatrix},$$

where $i<m<m+1<j$ with m about half way between i and j. Let $\widetilde{Z}_{12}$ be their common part, $\widetilde{Z}_{11}$ the part in $\widetilde{Z}_1$ excluding $\widetilde{Z}_{12}$, and $\widetilde{Z}_{22}$ the part in $\widetilde{Z}_2$ excluding $\widetilde{Z}_{12}$ also. We have

$$\widetilde{Z}_{11}\widetilde{Z}_{12}^{\dagger}=\begin{pmatrix}m-i+1 & -(m-i)\\ \vdots & \vdots\\ 3 & -2\\ 2 & -1\end{pmatrix},\quad \widetilde{Z}_{22}\widetilde{Z}_{12}^{\dagger}=\begin{pmatrix}-1 & 2\\ -2 & 3\\ \vdots & \vdots\\ -(j-m-1) & j-m\end{pmatrix}.$$

For large $j-i$ and m about half way between i and j,

$$\|\widetilde{Z}_{11}\widetilde{Z}_{12}^{\dagger}\|_2\approx\|\widetilde{Z}_{22}\widetilde{Z}_{12}^{\dagger}\|_2\sim\frac{1}{\sqrt{3}}\left(\frac{j-i}{2}\right)^{3/2}.$$

Consequently for large $j-i$

$$\alpha(\{\widetilde{Z}_1,\widetilde{Z}_2\})\sim\frac{1}{1+\frac{1}{3}\left(\frac{j-i}{2}\right)^3}\sim 3\left(\frac{2}{j-i}\right)^3.$$

This implies $\alpha(\boldsymbol{Z})$, modulo a constant factor, is approximately

$$\prod_{k=1}^{\lceil\log_2 N\rceil}3\left(\frac{2}{N/2^k}\right)^3\sim 3^{\log_2 N}\left(\frac{2^{\log_2 N}\,2^{(\log_2 N)^2/2}}{N^{\log_2 N}}\right)^3=\frac{1}{N^{(\log_2 N)/2-3-\log_2 3}}, \tag{4.9}$$

where $\lceil \log_2 N \rceil$ is the smallest integer that is no less than $\log_2 N$. Compared to (4.7) and (4.8), this very much underestimates $\lambda_{\min}^{+}(P)$ for large N, a conclusion similar to what we have made at the end of Example 4.3.

4.1. Necessary Condition In the case $s=2$, Theorem 3.2 states that the fully overlapped condition is also a necessary condition for $\mathsf{null}(P)=\mathsf{span}(Z)$, provided that all Z_i have full column rank. It turns out it is not a necessary condition in general when $s\geq 3$. We shall first give a counterexample to illustrate this and then give a result on when $\mathsf{null}(P)=\mathsf{span}(Z)$ does not hold.

EXAMPLE 4.5. Consider the 7×4 matrix

$$Z=\begin{pmatrix} 1\,0\,0\,1\\ 1\,1\,0\,0\\ 1\,1\,0\,1\\ 1\,0\,1\,0\\ 1\,0\,0\,0\\ 1\,0\,1\,1\\ 1\,1\,1\,0 \end{pmatrix}, \quad Z_1=Z_{(1:5,:)}, \quad Z_2=Z_{(3:7,:)}, \quad Z_3=Z_{(\{1,2,5:7\},:)}.$$

Then all Z_i have full column rank and they are pairwise not fully overlapped. Moreover, $\{Z_1,Z_2,Z_3\}$ is not fully overlapped. Computation by Maple's `nullspace(P)` gives a basis of 4 vectors, i.e., $\dim \mathsf{null}(P)=4$ which implies

$$\mathsf{null}(P)=\mathsf{span}(Z),$$

since $PZ=0$ and $\mathsf{rank}(Z)=4$.

While the fully overlapped condition is not a necessary condition, each Z_i in this example is fully overlapped with the union of the remaining Z_j's. The following theorem shows that this is indeed necessary for $\mathsf{null}(P)=\mathsf{span}(Z)$.

THEOREM 4.2. *Assume that Z_j has full column rank for $1\leq j\leq s$ and that $\boldsymbol{Z}=\{Z_j, 1\leq j\leq s\}$ can be partitioned into two nonempty disjoint subsets $\boldsymbol{Z}_1$ and $\boldsymbol{Z}_2$ such that the union set of $\boldsymbol{Z}_1$ and that of $\boldsymbol{Z}_2$ are not fully overlapped, i.e.,*

$$\widetilde{Z}_1=Z_{(\widetilde{\boldsymbol{I}}_1,:)}, \quad \widetilde{Z}_2=Z_{(\widetilde{\boldsymbol{I}}_2,:)}, \tag{4.10}$$

are not fully overlapped, where $\widetilde{\boldsymbol{I}}_1$ and $\widetilde{\boldsymbol{I}}_2$ are defined as in (2.5). *Then $\mathsf{span}(Z)$ is a proper subspace of $\mathsf{null}(P)$.*

Proof. Without loss of generality, let

$$\boldsymbol{Z}_1=\{Z_j, 1\leq j\leq p\} \text{ and } \boldsymbol{Z}_2=\{Z_j, p+1\leq j\leq s\}.$$

Since $\widetilde{\boldsymbol{I}}_1$ and $\widetilde{\boldsymbol{I}}_2$ are not fully overlapped, it follows from Theorem 3.2 that

$$\dim \mathsf{null}\left(\sum_{j=1}^{2}(I_N)^{\mathrm{T}}_{(\widetilde{\boldsymbol{I}}_j,:)}\times P^{\perp}_{\widetilde{Z}_j}\times (I_N)_{(\widetilde{\boldsymbol{I}}_j,:)}\right)>\ell.$$

Utilizing the fact that $\mathsf{null}(X+Y)=\mathsf{null}(X)\cap\mathsf{null}(Y)$ for two positive semi-definite

$X,Y \succeq 0$, we also have

$$\begin{aligned}
&\mathsf{null}\left(\sum_{j=1}^{2}(I_N)^{\mathrm{T}}_{(\widetilde{\boldsymbol{I}}_j,:)}\times P^{\perp}_{\widetilde{Z}_j}\times (I_N)_{(\widetilde{\boldsymbol{I}}_j,:)}\right)\\
&=\mathsf{null}\left((I_N)^{\mathrm{T}}_{(\widetilde{\boldsymbol{I}}_1,:)}\times P^{\perp}_{\widetilde{Z}_1}\times (I_N)_{(\widetilde{\boldsymbol{I}}_1,:)}\right)\bigcap\mathsf{null}\left((I_N)^{\mathrm{T}}_{(\widetilde{\boldsymbol{I}}_2,:)}\times P^{\perp}_{\widetilde{Z}_2}\times (I_N)_{(\widetilde{\boldsymbol{I}}_2,:)}\right)\\
&\subset \mathsf{null}(\Phi_1+\cdots+\Phi_p)\bigcap\mathsf{null}(\Phi_{p+1}+\cdots+\Phi_s)\\
&=\mathsf{null}(P).
\end{aligned}$$

Thus, $\dim\mathsf{null}(P) > \ell = \dim\mathsf{span}(Z)$ and the theorem is proved. ◻

The following is an interesting corollary that is not obvious from Definition 2.1.

COROLLARY 4.3. *Suppose $\boldsymbol{Z}=\{Z_1,Z_2,\dots,Z_s\}$ is a fully overlapped collection, where Z_j are submatrices of $Z\in\mathbb{C}^{N\times\ell}$ as defined by* (2.1) *and* (2.2). *Then for any two nonempty disjoint subsets $\boldsymbol{Z}_1$ and $\boldsymbol{Z}_2$ of $\boldsymbol{Z}$,*

$$\widetilde{Z}_1 = Z_{(\widetilde{\boldsymbol{I}}_1,:)},\quad \widetilde{Z}_2 = Z_{(\widetilde{\boldsymbol{I}}_2,:)},$$

must be fully overlapped, where $\widetilde{\boldsymbol{I}}_1$ and $\widetilde{\boldsymbol{I}}_2$ are defined as in (2.5).

REMARK 4.1. Zha and Zhang [10, Theorem 3.2] also established some necessary conditions for $\mathsf{null}(P)=\mathsf{span}(Z)$ in terms of full overlap as well as connected overlap for Z_j whose first column is all ones.

5. Conclusions

We have studied the eigenstructure of the alignment matrix P in a slightly more general context than in nonlinear manifold learning. It is proved that $\alpha(\boldsymbol{Z})P^{\perp}_{Z} \preceq P$ under the condition that $\boldsymbol{Z}$ is a fully overlapped collection, where $\alpha(\boldsymbol{Z})>0$ is computed recursively. For $s=2$, the bound is no worse than proportional to the square of the ratio of the smallest singular value of the matrix in the overlapped part to the largest singular value of the matrix in the non-overlapped part and this ratio can be considered as a measure of the "amount" of overlap.

From the computational point of view, the bigger the smallest nonzero eigenvalue $\lambda^{+}_{\min}(P)$, the less difficult it is to recover $\mathsf{null}(P)$ numerically. Our lower bound can be used as an indicator on the difficulty of numerically recovering $\mathsf{null}(P)$.

Another implication of our result is concerned with how to make $\lambda^{+}_{\min}(P)$ bigger – increase the overlaps as one naturally expects. But we provide a quantitative measure. Our present study contributes to the theoretical foundation of the LTSA algorithm [11]. But further studies are necessary. We mention two unanswered issues that need to be addressed in the future.

1. For $s=2$, our bound $\alpha(\boldsymbol{Z})$ is tight and asymptotically achievable, but for $s\ge 3$, the recursively computed $\alpha(\boldsymbol{Z})>0$ depends on how $\boldsymbol{Z}$ is partitioned as in Definition 2.1 and could very much underestimate $\lambda^{+}_{\min}(P)$. How do we improve the bound?
2. Any practical use of our result here remains to be investigated because in practice data errors may and will complicate the analysis and must be taken into account.

We shall leave these issues, among others, to our future studies on the subject.

Acknowledgment. Theorem 3.1 was explicitly formulated at the suggestion from an anonymous referee who also suggested us to add a simpler example – Example 4.4. Previously Theorem 3.1 was immersed in the text. The authors wish to thank his comments that improved the presentation of this paper.

REFERENCES

[1] M. Brand, *Charting a manifold*, in Advances in Neural Information Processing Systems 15, S. T. S. Becker and K. Obermayer, eds., MIT Press, Cambridge, MA, 2003, pp. 961–968.

[2] J. Demmel, *Applied Numerical Linear Algebra*, SIAM, Philadelphia, 1997.

[3] D. Donoho and C. Grimes, *Hessian eigenmaps: new tools for nonlinear dimensionality reduction*, Proceedings of National Academy of Science, 100 (2003), pp. 5591–5596.

[4] A. E. Naiman, I. M. Babuška, and H. C. Elman, *A note on conjugate gradient convergence*, Numer. Math., 76 (1997), pp. 209–230.

[5] S. T. Roweis and L. K. Saul, *Nonlinear dimensionality reduction by locally linear embedding*, Science, 290 (2000), pp. 2323–2326.

[6] G. W. Stewart and J.-G. Sun, *Matrix Perturbation Theory*, Academic Press, Boston, 1990.

[7] Y. W. Teh and S. Roweis, *Automatic alignment of local representations*, in Advances in Neural Information Processing Systems 15, S. T. S. Becker and K. Obermayer, eds., MIT Press, Cambridge, MA, 2003, pp. 841–848.

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

[9] Q. Ye, H. Zha, and R.-C. Li, *Analysis of an alignment algorithm for nonlinear manifold learning*, BIT, to appear.

[10] H. Zha and Z. Zhang, *Spectral analysis of alignment in manifold learning*, in Proceedings of Acoustics, Speech, and Signal Processing (ICASSP '05), vol. 5, March 2005, pp. 1069–1072.

[11] Z. Zhang and H. Zha, *Principal manifolds and nonlinear dimension reduction via local tangent space alignment*, SIAM J. Sci. Comput., 26 (2004), pp. 313–338.