Error Bounds For Approximate Deflating Subspaces For Linear Response Eigenvalue Problems

Wei-guo Wang
Lei-Hong Zhang
Ren-Cang Li

Technical Report 2016-01

http://www.uta.edu/math/preprint/
Error Bounds For Approximate Deflating Subspaces
For Linear Response Eigenvalue Problems

Wei-guo Wang∗ Lei-Hong Zhang† Ren-Cang Li‡

January 6, 2016

Dedicated to Professor Rajendra Bhatia on the occasion of his 65th birthday

Abstract

Consider the linear response eigenvalue problem (LREP) for \( H = \begin{bmatrix} 0 & K \\ M & 0 \end{bmatrix} \), where \( K \) and \( M \) are positive semidefinite and one of them is definite. Given a pair of approximate deflating subspaces of \( \{K, M\} \), it can be shown that LREP can be transformed into one for \( \tilde{H} \) that is nearly decoupled into two smaller LREPs upon congruence transformations on \( K \) and \( M \) that preserve the eigenvalues of \( H \). In this paper, we establish a bound on how far the pair of approximate deflating subspaces is from a pair of exact ones, using the closeness of \( \tilde{H} \) from being decoupled.

Key words. linear response eigenvalue problem, random phase approximation, perturbation, deflating subspaces

AMS subject classifications. 15A42, 65F15

1 Introduction

In computational quantum chemistry and physics, the so-called random phase approximation (RPA) describes the excitation states (energies) of physical systems in the study of collective motion of many-particle systems [1, 12, 13]. It has important applications in silicon nanoparticles and nanoscale materials and analysis of interstellar clouds [1, 2]. One important question in RPA is to compute a few eigenpairs associated with the smallest positive eigenvalues of the following eigenvalue problem:

\[
\mathcal{H} \begin{bmatrix} u \\ v \end{bmatrix} := \begin{bmatrix} A & B \\ -B & -A \end{bmatrix} \begin{bmatrix} u \\ v \end{bmatrix} = \lambda \begin{bmatrix} u \\ v \end{bmatrix}, \quad (1.1)
\]

∗School of Mathematical Sciences, Ocean University of China, Qingdao, 266100, People’s Republic of China. Email: wgwang@ouc.edu.cn. Supported in part by the National Natural Science Foundation of China Grant 11371333 and the Shandong Province Natural Science Foundation Grant ZR2013AM025.

†School of Mathematics and Shanghai Key Laboratory of Financial Information Technology, Shanghai University of Finance and Economics, 777 Guoding Road, Shanghai 200433, China. Email: longzh@163.com. Supported in part by the National Natural Science Foundation of China NSFC-11371102, and the Basic Academic Discipline Program, the 11th five year plan of 211 Project for Shanghai University of Finance and Economics.

‡Department of Mathematics, University of Texas at Arlington, P.O. Box 19408, Arlington, TX 76019 (rcli@uta.edu). Supported in part by NSF grants DMS-1317330 and CCF-1527104, and NSFC grant 11428104.
where \( A, B \in \mathbb{R}^{n \times n} \) are both symmetric matrices and \( \begin{bmatrix} A & B \\ B & A \end{bmatrix} \) is positive definite. Through a similarity transformation, this eigenvalue problem can be equivalently transformed into [1, 2, 4]

\[
Hz := \begin{bmatrix} 0 & K \\ M & 0 \end{bmatrix} \begin{bmatrix} y \\ x \end{bmatrix} = \lambda \begin{bmatrix} y \\ x \end{bmatrix},
\]

(1.2)

where \( K = A - B \) and \( M = A + B \). This eigenvalue problem was still referred to as the linear response eigenvalue problem (LREP) [1, 4, 15] and will be in this paper, too.

The condition imposed upon \( A \) and \( B \) in (1.1) implies that both \( K \) and \( M \) are symmetric and positive definite [1]. But in the rest of this paper, unless otherwise explicitly stated, we relax the positive definiteness of both \( K \) and \( M \) to that both are positive semidefinite and one of them is definite.

An important notion for LREP [1] is the so-called pair of deflating subspaces \( \{U, V\} \) by which we mean that both \( U \) and \( V \) are subspaces of \( \mathbb{R}^n \) and satisfy

\[
KU \subseteq V \quad \text{and} \quad MV \subseteq U.
\]

More discussions on this are in section 3. It is a generalization of the concept of the invariant subspace (or, eigenspace) in the standard eigenvalue problem upon considering the special structure in the LREP (1.2). This notion is not only vital in analyzing the theoretical properties such as the subspace version [1] of Thouless’s minimization principle [12] and the Cauchy-like interlacing inequalities [2], but also fundamental for several rather efficient algorithms, e.g., the Locally Optimal Block Preconditioned 4D Conjugate Gradient Method (LOBP4DCG) [2] and its space-extended variation [3], the block Chebyshev-Davidson method [11], as well as the generalized Lanczos method [9, 14, 15]. Each of these algorithms generates a sequence of approximate deflating subspace pairs \( \{U_j, V_j\} \) that hopefully converge to or contain subspaces near the pair of deflating subspaces of interest. Therefore, it is important to establish relationships between the accuracy in eigenvalue approximations and the distances from the exact deflating subspaces to their approximate ones.

Analogously to error estimate results for approximate invariant subspaces in the standard and generalized eigenvalue problems [5, 6, 7], in this paper, we will establish results on error bounds for approximate deflating subspaces of LREP (1.2).

The rest of the paper is organized as follows. In section 2, we will first collect some known results for LREP. These results are essential to our later development. Section 3 discusses the most important property of a pair of deflating subspaces: it decouples \( H \) into two smaller LREP's upon congruence transformations on \( K \) and \( M \) that preserve the eigenvalues of \( H \). In section 4, we present a way to construct the pair of exact deflating subspaces that will decouple the LREP, given a pair of approximate deflating subspaces. Our main results on the existence of the pair of exact deflating subspaces and how far it is from the pair of approximate ones are given in section 5.

Notation. \( \mathbb{K}^{n \times m} \) is the set of all \( n \times m \) matrices whose entries belong to the number field \( \mathbb{K} \), \( \mathbb{K}^n = \mathbb{K}^{n \times 1} \), and \( \mathbb{K}^1 = \mathbb{K}^1 \), where \( \mathbb{K} = \mathbb{R} \) (the set of real numbers) or \( \mathbb{C} \) (the set of complex numbers). \( I_n \) (or simply \( I \) if its dimension is clear from the context) denotes the \( n \times n \) identity matrix. All vectors are column vectors and are in boldface. For a matrix \( Z \),

1. \( Z^T \) denotes its transpose,
2. $\mathcal{R}(Z)$ is $Z$’s column space, spanned by its column vectors,

3. $\|Z\|_2$ and $\|Z\|_F$ are the spectral norm and the Frobenius norm, respectively,

4. $Z$’s submatrices $Z_{(k: \ell, i:j)}$, $Z_{(k: \ell, :)}$, and $Z_{(:, i:j)}$ consist of intersections of row $k$ to row $\ell$ and column $i$ to column $j$, row $k$ to row $\ell$, and column $i$ to column $j$, respectively,

5. when $Z$ is a square matrix and its eigenvalue set is $\text{eig}(Z)$.

2 Preliminaries

Recall that in LREP (1.2), we assume, in what follows, both $K$ and $M$ are positive semidefinite and one of them is definite. Many theoretical properties of LREP have been established in [1, 2]. In particular, LREP (1.2) has only real eigenvalues coming in pairs $\{\lambda, -\lambda\}$. Let the eigenvalues be

$$-\lambda_n \leq \cdots \leq -\lambda_1 \leq \lambda_1 \leq \cdots \leq \lambda_n. \quad (2.1)$$

Theorem 2.1 presents decompositions on $K$ and $M$, necessary for our developments later in this paper.

**Theorem 2.1.** Suppose that $K$ is semidefinite and $M$ is definite. Then the following statements are true:

(i) There exists a nonsingular $\Phi \in \mathbb{R}^{n \times n}$ such that

$$K = \Psi \Lambda^2 \Psi^T \quad \text{and} \quad M = \Phi \Phi^T, \quad (2.2)$$

where $\Lambda = \text{diag}(\lambda_1, \lambda_2, \ldots, \lambda_n)$, $\lambda_i$ are as in (2.1), and $\Psi = \Phi^{-T}$.

(ii) If $K$ is also definite, then all $\lambda_i > 0$ and $H$ is diagonalizable:

$$H \begin{bmatrix} \Psi \Lambda & \Psi \Lambda \\ \Phi & -\Phi \end{bmatrix} = \begin{bmatrix} \Psi \Lambda & \Psi \Lambda \\ \Phi & -\Phi \end{bmatrix} \begin{bmatrix} \Lambda & \Lambda \\ -\Lambda & -\Lambda \end{bmatrix}. \quad (2.3)$$

(iii) The eigen-decomposition of $KM$ and $MK$ are

$$(KM)\Psi = \Psi \Lambda^2 \quad \text{and} \quad (MK)\Phi = \Phi \Lambda^2, \quad (2.4)$$

respectively.

LREP (1.2) is equivalent to any one of the following product eigenvalue problems [1]

$$KM y = \lambda^2 y, \quad (2.5a)$$

$$MK x = \lambda^2 x. \quad (2.5b)$$

In theory, solving any one of them leads to solutions of the other two. It is tempting to solve either (2.5a) or (2.5b) because they are half the size of the original LREP (1.2). But numerically, it is preferable to solving (1.2) directly for a couple of reasons. First, $KM$ and $MK$ are no longer symmetric and thus spurious complex eigenvalues could show up due to rounding errors, and secondly if carefully designed, solving (1.2) directly costs about the same as solving either (2.5a)
or (2.5b) even though the former is twice the size of the latter. Nonetheless, for theoretic analysis there are times when it is more convenient to deal with either (2.5a) or (2.5b). For example, in seeking transformations to transform $K$ and $M$ to “simpler forms” while preserving the eigenvalues of these problems, working with either $KM$ or $MK$ for theoretic investigation seems to be more convenient. In fact, in order not to change the eigenvalues of $KM$, we may do

$$K \rightarrow Z^{-1}KX, \quad M \rightarrow X^{-1}MZ$$

to get

$$KM \rightarrow Z^{-1}(KM)Z$$

which, on $H$, becomes

$$H \rightarrow \begin{bmatrix} 0 & Z^{-1}KX \\ X^{-1}MZ & 0 \end{bmatrix} = \begin{bmatrix} Z & 0 \\ X & 0 \end{bmatrix}^{-1} H \begin{bmatrix} Z & 0 \\ X & 0 \end{bmatrix}.$$ 

But such transformations destroy the symmetry in $K$ and $M$, not to mention their definiteness property. Instead, we prefer symmetry- and definiteness-preserving transformations:

$$K \rightarrow X^T KX, \quad M \rightarrow Y^T MY,$$

with the constraint $XY^T = I_n$ to ensure that the eigenvalues of $KM$ before and after the transformation remain the same. In fact, under $XY^T = I_n$, we have

$$KM \rightarrow X^T KX \cdot Y^T MY = X^T (KM)Y = Y^{-1}(KM)Y$$

and, on $H$,

$$H \rightarrow \begin{bmatrix} 0 & X^T KX \\ Y^T MY & 0 \end{bmatrix} = \begin{bmatrix} X & Y \\ X & Y \end{bmatrix}^T H \begin{bmatrix} Y & X \\ Y & X \end{bmatrix},$$

a similarity transformation.

## 3 Deflating subspaces

Recall that $\{U, V\}$ is a pair of deflating subspaces of $\{K, M\}$ if

$$KU \subseteq V \quad \text{and} \quad MV \subseteq U. \quad (3.1)$$

Let $U \in \mathbb{R}^{n \times k}$ and $V \in \mathbb{R}^{n \times k}$ be the basis matrices for $U$ and $V$, respectively. Alternatively, (3.1) can be restated as that there exist $K_R \in \mathbb{R}^{k \times k}$ and $M_R \in \mathbb{R}^{k \times k}$ such that

$$KU = VK_R \quad \text{and} \quad MV = UM_R \quad (3.2)$$

and vice versa, or equivalently,

$$H \begin{bmatrix} V \\ U \end{bmatrix} = \begin{bmatrix} V \\ U \end{bmatrix} H_R \quad \text{with} \quad H_R := \begin{bmatrix} M_R & K_R \end{bmatrix},$$

i.e., $V \oplus U$ is an invariant subspace of $H$ [1, Theorem 2.4].

Two particular choices of $\{K_R, M_R\}$ to satisfy (3.2) are

$$K_R = (U^TV)^{-1}U^T KU, \quad M_R = (V^TU)^{-1}V^T MV; \quad (3.3a)$$

$$K_R = (V^TV)^{-1}V^T KU, \quad M_R = (U^TU)^{-1}U^T MV. \quad (3.3b)$$
In (3.3a), it is assumed that $U^TV$ is invertible, which is guaranteed if one of $K$ and $M$ is definite [1, Lemma 2.7]. By what we just proved, the associated $H_0$ with either (3.3a) or (3.3b) must have the same eigenvalues.

$\{K_r, M_r\}$ by (3.3a) relates to the structure-preserving projection $H_{sr}$ of $H$ in [2, (2.2)] that plays an important role numerically there. To highlight this particular pair, in [16] we called $\{K_r, M_r\}$ by (3.3a) a Rayleigh quotient pair of LREP (1.2) associated with $\{U, V\}$ and introduce

$$K_{rq} := (U^T)^{-1}U^TKU, \quad M_{rq} := (V^T)^{-1}V^TMV$$

for the ease of future references. Both $K_{rq}$ and $M_{rq}$ vary with different selections of $U$ and $V$ as the basis matrices of $U$ and $V$, respectively. But the eigenvalues of the induced

$$H_{rq} = \begin{bmatrix} M_{rq} & K_{rq} \end{bmatrix}.$$

do not.

In the case when $U^TV = I_k$, both $K_{rq}$ and $M_{rq}$ are symmetric and have the same definiteness property as their corresponding $K$ and $M$. But when $U^TV \neq I_k$, if one of $K$ and $M$ is definite [1, Lemma 2.7], we can always “normalize” them so that $U^TV$ is $I_k$. For example, we factorize $U^TV$ as $U^TV = W_1^TW_2$, where $W_i \in \mathbb{R}^{k \times k}$ are nonsingular, and then $UW_1^{-1}$ and $VW_2^{-1}$ are new basis matrices of $U$ and $V$, respectively, and satisfy $(UW_1^{-1})^T(VW_2^{-1}) = I_k$. For this reason, frequently we select basis matrices for a pair of deflating subspaces to have this property.

**Theorem 3.1.** Let $\{U, V\}$ be a pair of deflating subspaces of $\{K, M\}$ with $\dim(U) = \dim(V) = k$, and let the columns of $U_1$ and $V_1$ form bases for $U$ and $V$, respectively, and satisfy $U_1^TV_1 = I_k$. Then there are nonsingular matrices $[U_1, U_2]$ and $[V_1, V_2]$ such that $[U_1, U_2]^T[V_1, V_2] = I_n$, and

$$[U_1, U_2]^T K [U_1, U_2] = \begin{bmatrix} K_1 & 0 \\ 0 & K_2 \end{bmatrix}, \quad [V_1, V_2]^T M [V_1, V_2] = \begin{bmatrix} M_1 & 0 \\ 0 & M_2 \end{bmatrix}.$$

Furthermore, $\{\mathcal{R}(U_2), \mathcal{R}(V_2)\}$ is also a pair of deflating subspaces of dimension $n - k$, and

$$\text{eig}(H) = \text{eig}(H_1) \cup \text{eig}(H_2),$$

where $H_i = \begin{bmatrix} 0 & K_i \\ M_i & 0 \end{bmatrix}$ for $i = 1, 2$.

**Proof.** Let the columns of $U_2 \in \mathbb{R}^{n \times (n-k)}$ and $V_2 \in \mathbb{R}^{n \times (n-k)}$ be the basis matrices of the orthogonal complements of $V$ and $U$, respectively. In particular, $U_2^TV_1 = V_2^TU_1 = 0$. Then we have $U_2^TKU_1 = 0$ and $V_2^TMV_1 = 0$ (cf. (3.2)), which implies that $[U_1, U_2]^T K [U_1, U_2]$ and $[V_1, V_2]^T M [V_1, V_2]$ have the form (3.6).

We claim that both $[U_1, U_2]$ and $[V_1, V_2]$ are nonsingular. Consider $[U_1, U_2]$. It is sufficient to show that $U_1y = U_2z$ implies $y = 0$ and $z = 0$. We have $y = V_1^TU_1y = V_1^TU_2z = 0$ since $V_1^TU_2 = 0$. Thus $U_2z = 0$ which implies $z = 0$ because $U_2$ is the basis matrix of the orthogonal complement of $V$. Similarly $[V_1, V_2]$ is also nonsingular, and so is

$$[U_1, U_2]^T [V_1, V_2] = \begin{bmatrix} I_k & 0 \\ 0 & U_2^TV_2 \end{bmatrix}.$$
Thus $U_2^TV_2$ is nonsingular. As we just commented before stating this theorem, we can always “normalize” $U_2^TV_2$ so that it is $I_{n-k}$, which we will assume for the rest of this proof. Now use (3.6) to get

$$K[U_1, U_2] = [V_1, V_2] \begin{bmatrix} K_1 & 0 \\ 0 & K_2 \end{bmatrix} \Rightarrow KU_2 = V_2K_2,$$

and similarly, $MU_2 = V_2M_2$. Hence $\{R(U_2), R(V_2)\}$ is a pair of deflating subspaces of dimension $n - k$.

**Corollary 3.1.** Let $M$ be positive definite, and let $\{U, V\}$ be a pair of deflating subspaces with $\dim(U) = \dim(V) = k$. Then there are nonsingular matrices $U$ and $V$ such that $UV^T = I_n$, and

$$U^TKU = \begin{bmatrix} K_1 & 0 \\ 0 & K_2 \end{bmatrix}, \quad V^TMV = \begin{bmatrix} I_k & 0 \\ 0 & I_{n-k} \end{bmatrix}. \quad (3.7)$$

**Proof.** In Theorem 3.1, let $M_1 = R_1^TR_1$ and $M_2 = R_2^TR_2$ be the Cholesky decompositions. $R_i$ for $i = 1, 2$ are nonsingular. Set

$$U := [U_1, U_2] \begin{bmatrix} R_1^T \\ 0 & R_2^T \end{bmatrix}, \quad V := [V_1, V_2] \begin{bmatrix} R_1^{-1} \\ 0 & R_2^{-1} \end{bmatrix},$$

to complete the proof. \qed

### 4 Approximate Deflating Subspaces

Theorem 3.1 says that if $\{U, V\}$ is a pair of deflating subspaces, then both $K$ and $M$ can be block-diagonalizable as in (3.6). The opposite is also true. Because of $[U_1, U_2]^T[V_1, V_2] = I_n$, the congruence transformations on $K$ and $M$ there preserve eigenvalues as we argued at the end of section 2.

Now what if $\{U, V\}$ is just a pair of approximate deflating subspaces, then we won’t expect to have something like (3.6). In fact, examining the proof of Theorem 3.1, we will see now we don’t have $U_2^TKU_1 = 0$ and $V_2^TMV_1 = 0$ anymore, except that both $U_2^TKU_1$ and $V_2^TMV_1$ are expected to be small in magnitude. Thus again following through the proof, we will have $[U_1, U_2]^T[V_1, V_2] = I_n$, and

$$[U_1, U_2]^T[K[U_1, U_2] = \begin{bmatrix} K_1 & 0 \\ 0 & K_2 \end{bmatrix} =: \mathcal{K}, \quad (4.1a)$$

$$[V_1, V_2]^T[M[V_1, V_2] = \begin{bmatrix} M_1 & 0 \\ 0 & M_2 \end{bmatrix} =: \mathcal{M}, \quad (4.1b)$$

with tiny $\|K_{21}\|_2$ and $\|M_{21}\|_2$, where $U_1$ and $V_1$ are basis matrices of $\mathcal{U}$ and $\mathcal{V}$, respectively. Naturally, we expect that there will be a pair of exact deflating subspaces that is near $\{U, V\}$. Recently in [10], the authors consider the effect of setting both $K_{21}$ and $M_{21}$ to 0 on the eigenvalues. Perturbation bounds in terms of the squares of $\|K_{21}\|_2$ and $\|M_{21}\|_2$ are obtained.
The main purpose of this paper, however, is to quantify how far \( \{U, V\} \) is from the pair of exact deflating subspaces. This section sets up the stage for our main result in section 5.

For \( \tilde{K} \) and \( \tilde{M} \) given by (4.1), let

\[
\tilde{X} = \begin{bmatrix} I & -Q^T \\ P & I \end{bmatrix}, \quad \tilde{Y} = \begin{bmatrix} I & -P^T \\ Q & I \end{bmatrix},
\]

(4.2)

where \( P, Q \in \mathbb{R}^{(n-k) \times k} \) are to be determined such that the off-diagonal blocks of

\[
\tilde{X}^T \tilde{K} \tilde{X} = \begin{bmatrix} I & -Q^T \\ P & I \end{bmatrix} \begin{bmatrix} K_1 & K_{21}^T \\ K_{21} & K_2 \end{bmatrix} \begin{bmatrix} I & -Q^T \\ P & I \end{bmatrix} =: \begin{bmatrix} \tilde{K}_1 & \tilde{K}_{21} \\ \tilde{K}_{21} & \tilde{K}_2 \end{bmatrix},
\]

(4.3a)

\[
\tilde{Y}^T \tilde{M} \tilde{Y} = \begin{bmatrix} I & -P^T \\ Q & I \end{bmatrix} \begin{bmatrix} M_1 & M_{21}^T \\ M_{21} & M_2 \end{bmatrix} \begin{bmatrix} I & -P^T \\ Q & I \end{bmatrix} =: \begin{bmatrix} \tilde{M}_1 & \tilde{M}_{21} \\ \tilde{M}_{21} & \tilde{M}_2 \end{bmatrix},
\]

(4.3b)

become 0, i.e., making \( \tilde{K}_{21} = \tilde{M}_{21} = 0 \). Calculations give

\[
\tilde{K}_{21} = K_{21} - QK_1 + K_2P - QK_{21}^T P,
\]

(4.4a)

\[
\tilde{M}_{21} = M_{21} - PM_1 + M_2Q - PM_{21}^T Q.
\]

(4.4b)

Thus to make \( \tilde{K}_{21} = \tilde{M}_{21} = 0 \), we need \( P \) and \( Q \) to satisfy

\[
QK_1 - K_2P = K_{21} - QK_{21}^T P,
\]

(4.4a)

\[
PM_1 - M_2Q = M_{21} - PM_{21}^T Q.
\]

(4.4b)

Suppose we have found such \( P \) and \( Q \) and if also \( \|P^TQ\|_2 < 1 \), then

\[
\tilde{X} \tilde{Y}^T = \tilde{Y}^T \tilde{X} = \begin{bmatrix} I + Q^T P & 0 \\ 0 & I + PQ^T \end{bmatrix} =: D
\]

is nonsingular; so are \( \tilde{X} \) and \( \tilde{Y} \). Now set\(^1\)

\[
X = \tilde{X} D^{-1}, \quad Y = \tilde{Y}.
\]

(4.5)

It can be verified that \( XY^T = \tilde{X} D^{-1} \tilde{Y}^T = I \) and

\[
X^T \tilde{K} X = \begin{bmatrix} \tilde{K}_1 & \tilde{K}_2 \end{bmatrix} =: \tilde{K}, \quad Y^T \tilde{M} Y = \begin{bmatrix} \tilde{M}_1 & \tilde{M}_2 \end{bmatrix} =: \tilde{M},
\]

(4.6a)

where

\[
\tilde{K}_1 = (K_1 + K_{21}^T P)(I + Q^T P)^{-1},
\]

(4.6b)

\[
\tilde{M}_1 = (I + Q^T P)(M_1 + M_{21}^T Q),
\]

(4.6c)

\(^1\)This is only one choice among many. Other choices will also work, e.g.,

\[
X = \tilde{X}, \quad Y = \tilde{Y} D^{-1}, \quad \text{or} \quad X = \tilde{X} D^{-1/2}, \quad Y = \tilde{Y} (D^{-1/2})^T.
\]
\[ \hat{K}_2 = (I + QP^T)^{-1}(K_2 - QK_2^T), \] \[ \hat{M}_2 = (M_2 - PM_2^T)(I + QP^T). \] (4.6d) (4.6e)

These expressions do not look like that they actually come from (4.3). In particular, \( \hat{K}_i \) and \( \hat{M}_i \) do not look like symmetric but they are. This is because we have used (4.4). In fact, we have

\[
\hat{K}_1 = K_1 + P^T K_{21} + K_{21}^T P + P^T K_2 P
\]

(clearly symmetric)

\[
= K_1 + K_{21}^T P + P^T(K_{21} + K_2 P)
\]

(4.4a)

\[
= (I + P^T Q)(K_1 + K_{21}^T P),
\]

\[
\tilde{K}_1 = (I + P^T Q)^{-1} \hat{K}_1 (I + Q^T P)^{-1}
\]

(5.1)

Because of \( XY^T = I \), \( KM \) and \( \tilde{K} \tilde{M} \) have the same eigenvalues and so do \( H \) and

\[
\tilde{H} = \begin{bmatrix} \tilde{K} \\ \tilde{M} \end{bmatrix}.
\]

Our analysis so far suggests the importance of the system (4.4) of equations in \( P \) and \( Q \). In preparation of studying the system, in the next subsection we introduce a linear operator and investigating its invertibility. Much of the development in the next two subsections bears similarity to Stewart [6, 7].

## 5 Error bounds for approximate deflating subspaces

The most important outcome of the analysis in section 4 is that if the equations in (4.4) hold, then \( \{R(U_1 + U_2 P), R(V_1 + V_2 Q)\} \) is an exact pair of deflating subspaces of \( \{K, M\} \). In this section, we will bound the difference between the pair of approximate deflating subspaces \( \{R(U), R(V)\} \) and this exact pair of deflating subspaces. To this end, we first investigate an operator and then use the result of the investigation to devise bound for the difference.

### 5.1 The operator \( \mathcal{L}(P, Q) = (QK_1 - K_2 P, PM_1 - M_2 Q) \)

Let \( K_1, M_1 \in \mathbb{R}^{k \times k} \) and \( K_2, M_2 \in \mathbb{R}^{m \times m} \), not necessarily symmetric. Define linear operator \( \mathcal{L} : \mathbb{R}^{m \times k} \times \mathbb{R}^{m \times k} \to \mathbb{R}^{m \times k} \times \mathbb{R}^{m \times k} \) by

\[
\mathcal{L}(P, Q) = (QK_1 - K_2 P, PM_1 - M_2 Q).
\] (5.1)

In this subsection, we will investigate the non-singularity properties of \( \mathcal{L} \). The results may have independent interests of their own and will be used later in showing the solvability of the system (4.4) of equations for \( P \) and \( Q \) and establishing bounds on \( P \) and \( Q \).

Although in (4.4) \( K_i \) and \( M_i \) are assumed symmetric, as a result of their sources in LREP (1.2), this assumption is not required here. So this subsection is a little bit more general than we will need later. On the other hand, with or without this symmetry assumption does not increase the complexity of our development in this section.

The following theorem succinctly states a non-singularity condition of \( \mathcal{L} \) in terms of the eigenvalues of \( K_1 M_1 \) and \( K_2 M_2 \).
Theorem 5.1. The operator $\mathcal{L}$ is invertible if and only if $\text{eig}(K_1 M_1) \cap \text{eig}(K_2 M_2) = \emptyset$.

Proof. Suppose that $\text{eig}(K_1 M_1) \cap \text{eig}(K_2 M_2) = \emptyset$. We need to show that $\mathcal{L}$ is invertible. Since $\mathcal{L}$ maps $\mathbb{R}^{m \times k} \times \mathbb{R}^{m \times k}$ to itself, it suffices to show that $\mathcal{L}(P, Q) = (0, 0)$ has only the trivial solution $(0, 0)$. Otherwise, let $Q \neq 0$. Post-multiply $QK_1 - K_2 P = 0$ by $M_1$ and pre-multiply $PM_1 - M_2 Q = 0$ by $K_2$ and then add the resulting equations to get $QK_1 M_1 - K_2 M_2 Q = 0$ which can only have the trivial solution $0$ since $\text{eig}(K_1 M_1) \cap \text{eig}(K_2 M_2) = \emptyset$, contradicting $Q \neq 0$.

Conversely, we will prove that $\text{eig}(K_1 M_1) \cap \text{eig}(K_2 M_2) \neq \emptyset$ implies that $\mathcal{L}$ is singular. From $\text{eig}(K_1 M_1) \cap \text{eig}(K_2 M_2) \neq \emptyset$ it follows that there is a nonzero $\tilde{Q} \in \mathbb{R}^{m \times k}$ such that $QK_1 M_1 = K_2 M_2 \tilde{Q}$. For the same reason there is a nonzero $W \in \mathbb{R}^{k \times k}$ so that $W M_1 K_1 = K_1 M_1 W$. Because $\text{eig}(K_1 M_1) \cap \text{eig}(M_1 K_1) \neq \emptyset$. Now set $P = M_2 \tilde{Q} W$ and $Q = Q W M_1$ to get $P M_1 = M_2 Q$ and

$$QK_1 - K_2 P = \tilde{Q} W M_1 K_1 - K_2 M_2 \tilde{Q} W = \tilde{Q} W M_1 K_1 - \tilde{Q} K_1 M_1 W = \tilde{Q} (W M_1 K_1 - K_1 M_1 W) = 0,$$

Therefore $\mathcal{L}$ is singular.

In order to obtain estimates on the solution of the equation $\mathcal{L}(P, Q) = (R, S)$, we define a norm on $\mathbb{R}^{m \times k} \times \mathbb{R}^{m \times k}$ as follows: for given pair $(P, Q) \in \mathbb{R}^{m \times k} \times \mathbb{R}^{m \times k}$,

$$\|(P, Q)\|_F := \|(P, Q)\| = \sqrt{\|P\|_F^2 + \|Q\|_F^2},$$

and accordingly define the dis function, in analogy with the functions sep and dif introduced in Stewart [5, 6],

$$\text{dis}(K_1, M_1; K_2, M_2) := \min \|\mathcal{L}(P, Q)\|_F.$$

Since $\mathcal{L}$ is a linear operator that maps a finite dimensional vector space to itself, by Theorem 5.1, we have

$$\text{dis}(K_1, M_1; K_2, M_2) = 0 \Leftrightarrow \mathcal{L} \text{ is singular, i.e., } \text{eig}(K_1 M_1) \cap \text{eig}(K_2 M_2) \neq \emptyset.$$

If $\mathcal{L}$ is nonsingular, then

$$\|(P, Q)\|_F \leq \frac{\|\mathcal{L}(P, Q)\|_F}{\text{dis}(K_1, M_1; K_2, M_2)}.$$

We remark that the use of min in (5.2) is justifiable. Introduce the vec-operation that turns a matrix into a column vector by appending its columns one by one with the first column followed by the second column and the third one and so on, and let $\otimes$ be the usual matrix Kronecker product. We have

$$\begin{bmatrix} \text{vec}(QK_1 - K_2 P) \\ \text{vec}(PM_1 - M_2 Q) \end{bmatrix} = L \begin{bmatrix} \text{vec}(P) \\ \text{vec}(Q) \end{bmatrix} \quad \text{with} \quad L = \begin{bmatrix} - (I \otimes K_2) & K_1^T \otimes I \\ M_1^T \otimes I & -(I \otimes M_2) \end{bmatrix}. \tag{5.4}$$

---

[2] We thank Dr. Weihong Yang of Fudan University for this “converse” part of proof which is much simpler than our original one.
The big matrix $L$ is thus the matrix representation of the operator $\mathcal{L}$, and

$$\text{dis}(K_1, M_1; K_2, M_2) = \sigma_{\text{min}}(L),$$

the smallest singular value of $L$, attained when $\begin{bmatrix} \text{vec}(P) \\ \text{vec}(Q) \end{bmatrix}$ is the unit right singular vector corresponding to the smallest singular value $\sigma_{\text{min}}(L)$ of $L$.

In what follows, we give a few useful results about this dis($\cdot$) function.

**Lemma 5.1.**

$$\text{dis}(K_1 + E_1, M_1 + F_1; K_2 + E_2, M_2 + F_2) \geq \text{dis}(K_1, M_1; K_2, M_2) - \sqrt{\| (E_1, F_1) \|_F^2 + \| (E_2, F_2) \|_F^2}. \quad (5.5)$$

**Proof.** Suppose $\text{dis}(K_1 + E_1, M_1 + F_1; K_2 + E_2, M_2 + F_2)$ is attended by the pair $(P, Q)$. Then

$$\text{dis}(K_1 + E_1, M_1 + F_1; K_2 + E_2, M_2 + F_2) = \| (Q(K_1 + E_1) - (K_2 + E_2)P, P(M_1 + F_1) - (M_2 + F_2)Q) \|_F$$

$$= \sqrt{\| (Q(K_1 + E_1) - (K_2 + E_2)P, P(M_1 + F_1) - (M_2 + F_2)Q) \|_F^2}$$

$$\geq \sqrt{\| QK_1 - K_2P \|_F^2 + \| PM_1 - M_2Q \|_F^2 - \sqrt{\| QE_1 - E_2P \|_F^2 + \| PF_1 - F_2Q \|_F^2}}$$

$$\geq \text{dis}(K_1, M_1; K_2, M_2) - \sqrt{\| E_1 \|_F^2 + \| E_2 \|_F^2 + \| F_1 \|_F^2 + \| F_2 \|_F^2},$$

as was to be shown. \qed

Stewart [5] introduced the function

$$\text{sep}(K_1, K_2) = \min_{\| P \|_F = 1} \| PK_1 - K_2P \|_F$$

to measure the separation between the operators $K_1$ and $K_2$. It was used in the perturbation analysis for approximate invariant subspaces of a matrix. The function dis($\cdot$) will play an analogous role for deflating subspaces of $\{K, M\}$. Since, when $M = I$, the eigenvalue problem for $KM$ reduces to an ordinary eigenvalue problem, it is natural to look for some connection between $\text{sep}(\cdot)$ and dis($\cdot$).

**Theorem 5.2.** For all square matrices $K_1$ and $K_2$,

$$\frac{1}{\sqrt{2}} \text{sep}(K_1, K_2) \geq \text{dis}(K_1, I; K_2, I). \quad (5.6)$$

If also $\|K_1\|_2 \leq 1$ and $\|K_2\|_2 \leq 1$, then

$$\text{dis}(K_1, I; K_2, I) \geq \frac{1}{2} \text{sep}(K_1, K_2). \quad (5.7)$$

**Proof.** The proof is similar to that of [6, Theorem 4.3]. Notice

$$\text{vec}(PK_1 - K_2P) = (K_1^T \otimes I - I \otimes K_2) \text{vec}(P),$$

and

$$\text{vec}(PK_1 - K_2P) = (K_1^T \otimes I - I \otimes K_2) \text{vec}(P),$$

10
and we let \( \text{vec}(P) \) be the unit singular vector of \( K_1^T \otimes I - I \otimes K_2 \) associated with its smallest singular value. Then
\[
\|P\|_F = 1, \quad \text{sep}(K_1, K_2) = \|PK_1 - K_2P\|_F.
\]
Since \( \|(P, P)\|_F = \sqrt{2} \), we have
\[
\text{dis}(K_1, I; K_2, I) \leq \frac{1}{\sqrt{2}} \|\mathcal{L}(P, P)\|_F
\]
\[
= \frac{1}{\sqrt{2}} \|(PK_1 - K_2P, P - P)\|_F
\]
\[
= \frac{1}{\sqrt{2}} \text{sep}(K_1, K_2).
\]
To prove (5.7), we let \((P, Q)\) be the optimal pair that achieves \(\text{dis}(K_1, I; K_2, I)\), i.e.,
\[
\|((P, Q))\|_F = 1, \quad \text{dis}(K_1, I; K_2, I) = \|\mathcal{L}((P, Q))\|_F. \tag{5.8}
\]
Set \(E = QK_1 - K_2P\) and \(F = P - Q\). The second equation in (5.8) says \(\text{dis}(K_1, I; K_2, I) = \|(E, F)\|_F\). Use \(E = QK_1 - K_2P = QK_1 - K_2Q - K_2F\) to get
\[
\|E\|_F \geq \|QK_1 - K_2Q\|_F - \|K_2F\|_F
\]
\[
\geq \|QK_1 - K_2Q\|_F - \|F\|_F,
\]
\[
\|E\|_F + \|F\|_F \geq \|QK_1 - K_2Q\|_F. \tag{5.9}
\]
Let \(\|P\|_F = \alpha\), and \(\|Q\|_F = \beta\). Then \(\alpha^2 + \beta^2 = 1\) which implies \(\max\{\alpha, \beta\} \geq 1/\sqrt{2}\). It follows from (5.9) that
\[
\|E\|_F + \|F\|_F \geq \|QK_1 - K_2Q\|_F \geq \beta \text{sep}(K_1, K_2), \tag{5.10a}
\]
and similarly
\[
\|E\|_F + \|F\|_F \geq \|PK_1 - K_2P\|_F \geq \alpha \text{sep}(K_1, K_2). \tag{5.10b}
\]
Combining (5.10a) and (5.10b), we have \(\|E\|_F + \|F\|_F \geq \max\{\alpha, \beta\} \text{sep}(K_1, K_2)\), and thus
\[
\text{dis}(K_1, I; K_2, I) = \sqrt{\|E\|_F^2 + \|F\|_F^2}
\]
\[
\geq \frac{1}{\sqrt{2}} \|E\|_F + \|F\|_F
\]
\[
\geq \frac{1}{\sqrt{2}} \max\{\alpha, \beta\} \text{sep}(K_1, K_2)
\]
\[
\geq \frac{1}{2} \text{sep}(K_1, K_2),
\]
as expected.

Finally, we have the following theorem.

**Theorem 5.3.** Let \(L\) be defined by (5.4). Then
\[
\frac{1}{\text{dis}(K_1, M_1; K_2, M_2)} = \|L^{-1}\|_2.
\]

11
5.2 Error bounds

We now return to LREP (1.2) with (4.1) in section 4. We would like to know when the system (4.4) has a solution pair \((P, Q)\) with \(\|P^TQ\|_2 < 1\). In terms of the linear operator \(\mathcal{L}\) defined in (5.1), the equations in (4.4) can be combined into one nonlinear equation:

\[
\mathcal{L}(P, Q) = (K_{21} - QK_{21}^TP, M_{21} - PM_{21}^TQ). \tag{5.11}
\]

**Lemma 5.2.** Let \(K_i, M_i, K_{21} \text{ and } M_{21}\) be given by (4.1), and let

\[
\delta = \|(K_{21}, M_{21})\|_F, \quad \gamma = \text{dis}(K_1, M_1; K_2, M_2). \tag{5.12}
\]

If \(\gamma > 0\) and

\[
\kappa_1 = \frac{\delta^2}{\gamma^2} < \frac{1}{4}, \tag{5.13}
\]

then (5.11) has a solution pair \((P, Q)\) \(\in \mathbb{R}^{(n-k)\times k} \times \mathbb{R}^{(n-k)\times k}\) satisfying

\[
\|(P, Q)\|_F \leq \rho := \frac{\delta}{\gamma} (1 + \kappa) < \frac{2\delta}{\gamma}, \tag{5.14}
\]

where

\[
\kappa = \frac{2\kappa_1}{1 - 2\kappa_1 + \sqrt{1 - 4\kappa_1}} < 1. \tag{5.15}
\]

**Proof.** The proof is largely similar to that of \([6, \text{Theorem 5.1}]\). The idea is to construct an iterative procedure to solve (5.11) for \(P\) and \(Q\). Since \(\gamma > 0\), \(\mathcal{L}\) is invertible. Let

\[
(P_0, Q_0) = \mathcal{L}^{-1}(K_{21}, M_{21}), \tag{5.16}
\]

and recursively define \((P_{i+1}, Q_{i+1})\) by

\[
(P_{i+1}, Q_{i+1}) = \mathcal{L}^{-1}(K_{21} - Q_iK_{21}^TP_i, M_{21} - PM_{21}^TQ_i)
= (P_0, Q_0) - \mathcal{L}^{-1}(Q_iK_{21}^TP_i, P_iM_{21}^TQ_i). \tag{5.17}
\]

We will show that the sequence \(\{(P_i, Q_i)\}\) converges to a solution of (5.11).

The first step is to bound \(\|(P_i, Q_i)\|_F\). From (5.16), we have

\[
\|(P_0, Q_0)\|_F = \|\mathcal{L}^{-1}(K_{21}, M_{21})\|_F \leq \frac{\delta}{\gamma} =: \rho_0 < \frac{1}{2}
\]

by (5.13). Assuming \(\|(P_i, Q_i)\|_F \leq \rho_i\), we have by (5.17)

\[
\|(P_{i+1}, Q_{i+1})\|_F \leq \rho_i + \gamma^{-1} \eta \rho_i^2 =: \rho_{i+1}.
\]

This recursively defines \(\rho_i\) for \(i \geq 1\). Write \(\rho_i = \rho_0(1 + \kappa_i)\) for \(i \geq 1\), which recursively define \(\kappa_i\) by

\[
\kappa_1 = \frac{\delta \rho_0}{\gamma} = \frac{\delta^2}{\gamma^2}, \tag{5.18a}
\]

\[
\kappa_{i+1} = \kappa_1 (1 + \kappa_i)^2 \text{ for } i \geq 1. \tag{5.18b}
\]
If (5.13) is satisfied, we see that \( \kappa_1 < \frac{1}{4} \) and \( \kappa_2 < \frac{1}{4}(1 + \frac{1}{4})^2 = 1. \) From \( \kappa_1 < \frac{1}{4} \) and \( \kappa_2 < 1, \) we have \( \kappa_3 < \frac{1}{4}(1+1)^2 = 1. \) Suppose that \( \kappa_i < 1. \) Then we have \( \kappa_{i+1} = \kappa_i(1+\kappa_i)^2 < \frac{1}{4}(1+1)^2 = 1. \) Thus, by induction, all \( \kappa_i < 1. \) On the other hand, \( \kappa_2 = \kappa_1(1+\kappa_1)^2 > \kappa_1. \) Now if \( \kappa_i > \kappa_{i-1}, \) then \( \kappa_{i+1} = \kappa_i(1+\kappa_i)^2 > \kappa_i(1+\kappa_{i-1})^2 = \kappa_i. \) This shows that the sequence \( \{\kappa_i\} \) is monotonically increasing. Hence it converges and

\[
\kappa_i < \kappa := \lim_{i \to \infty} \kappa_i = \frac{2\kappa_1}{1 - 2\kappa_1 + \sqrt{1 - 4\kappa_1}} < 1.
\]

As a result, the sequence \( \{\rho_i = \rho_0(1 + \kappa_i)\} \) is monotonically increases, too, and

\[
\| (P_i, Q_i) \|_F \leq \rho := \lim_{i \to \infty} \rho_i = \rho_0(1 + \kappa) < 1.
\]

To show that the matrix pair sequence \((P_i, Q_i)\) converge also, we let

\[
\Delta_i = (P_{i+1} - P_i, Q_{i+1} - Q_i) =: (\Delta_{P_i}, \Delta_{Q_i}).
\]

Then

\[
\| \Delta_i \|_F = \| \Sigma^{-1} (Q_i K_{2i}^T P_i - Q_{i-1} K_{2i}^T P_{i-1}, P_i M_{2i}^T Q_i - P_{i-1} M_{2i}^T Q_{i-1}) \|_F \\
\leq \gamma^{-1} \| (\Delta_{Q_{i-1}} K_{2i}^T P_i + Q_{i-1} K_{2i}^T \Delta_{P_{i-1}}, \Delta_{P_{i-1}} M_{2i}^T Q_i + P_{i-1} M_{2i}^T \Delta_{Q_{i-1}}) \|_F \\
\leq 2\gamma^{-1} \delta \rho \| \Delta_{i-1} \|_F.
\]

So the sequence \( \{(P_i, Q_i)\} \) converges provided \( 2\gamma^{-1} \delta \rho < 1, \) which is ensured by (5.13) and \( \rho < 1. \) The limit of \((P_i, Q_i)\) gives a solution to (5.11) and satisfies (5.14).

To measure differences between subspaces, we need the notion of canonical angles between two subspaces \( \mathcal{U} \) and \( \mathcal{U} \) of dimension \( k. \) Let \( \mathcal{U} \) and \( \mathcal{U} \) be the basis matrices of \( \mathcal{U} \) and \( \mathcal{U}, \) respectively. The canonical angles between \( \mathcal{U} \) and \( \mathcal{U} \) are [8, Definition 5.3 on p.43]

\[
\theta_i := \arccos \sigma_i, \quad i = 1, 2, \ldots, k,
\]

where \( \sigma_i \) \((1 \leq i \leq k)\) are the singular values of \((U^T U)^{-1/2} U^T \mathcal{U} (\mathcal{U}^T U)^{-1/2}. \) Furthermore, we define the angle \( \angle(\mathcal{U}, \mathcal{U}) \) and the angle matrix between \( \mathcal{U} \) and \( \mathcal{U} \) be to be

\[
\angle(\mathcal{U}, \mathcal{U}) = \max_i \theta_i, \quad \Theta(\mathcal{U}, \mathcal{U}) = \text{diag}(\theta_1, \theta_2, \ldots, \theta_k).
\]

Note the canonical angles \( \theta_i \) and thus \( \angle(\mathcal{U}, \mathcal{U}) \) and \( \Theta(\mathcal{U}, \mathcal{U}) \) are independent of the choices of basis matrices.

**Theorem 5.4.** For LREP (1.2) with (4.1), where \([U_1, U_2]^T[V_1, V_2] = I_n, \) if (5.13) holds, then there is a matrix pair \((P, Q) \in \mathbb{R}^{(n-k) \times k} \times \mathbb{R}^{(n-k) \times k} \) satisfying (5.14) to yield \( X \) and \( Y \) in (4.5) such that (4.6) holds. In particular,

\[
\{R(U_1 + U_2 P), R(V_1 + V_2 Q)\}
\]

is a pair of deflating subspace for \( \{K, M\}, \) and

\[
eig(H) = \eig(H_1) \cup \eig(H_2), \quad \eig(H_1) \cap \eig(H_2) = \emptyset,
\]

(5.19)
where, with $\tilde{K}_i$ and $\tilde{M}_i$ given by (4.6), $H_i = \begin{bmatrix} 0 & \tilde{K}_i \\ \tilde{M}_i & 0 \end{bmatrix}$ for $i = 1, 2$. Moreover, if
\[
\eta := 2 \max \left\{ \cos \theta_U \frac{\sigma_{\max}(U_2)}{\sigma_{\min}(U_1)}, \cos \theta_V \frac{\sigma_{\max}(V_2)}{\sigma_{\min}(V_1)} \right\} < 1,
\] (5.20)
where $\sigma_{\max}(\cdot)$ and $\sigma_{\min}(\cdot)$ denote the largest and smallest singular value of a matrix, then
\[
\| \tan \Theta_U \|_F \leq \frac{\| P \|_F}{\sigma_{\min}(V_2)\sigma_{\min}(U_1)} \cdot \frac{1}{1 - \eta},
\] (5.21a)
\[
\| \tan \Theta_V \|_F \leq \frac{\| Q \|_F}{\sigma_{\min}(U_2)\sigma_{\min}(V_1)} \cdot \frac{1}{1 - \eta},
\] (5.21b)
where $\| P \|_F$ and $\| Q \|_F$ can be bounded using (5.14), and
\[
\theta_U = \angle(\mathcal{R}(U_1), \mathcal{R}(U_2)), \quad \theta_V = \angle(\mathcal{R}(V_1), \mathcal{R}(V_2)),
\] (5.22)
\[
\Theta_U = \Theta(\mathcal{R}(U_1), \mathcal{R}(U_1 + U_2P)), \quad \Theta_V = \Theta(\mathcal{R}(V_1), \mathcal{R}(V_1 + V_2Q)).
\] (5.23)

Proof. Under the assumption of the theorem, (5.11) has a solution satisfying (5.14). Hence, the column spaces of $U_1 + U_2P$ and $V_1 + V_2Q$ form a pair of deflating subspace for $\{K, M\}$.

Note that $P$ and $Q$ satisfy (4.4). By (4.6), we have
\[
\text{eig}(\tilde{K}_1\tilde{M}_1) = \text{eig}((K_1 + K_{21}^T P)(M_1 + M_{21}^T Q)),
\]
\[
\text{eig}(\tilde{K}_2\tilde{M}_2) = \text{eig}((K_2 - QK_{21}^T (M_2 - PM_{21}^T))).
\]
Hence, the first relation of (5.19), i.e., $\text{eig}(H) = \text{eig}(H_1) \cup \text{eig}(H_2)$ holds as a consequence of (4.6). We now show $\text{eig}(H_1) \cap \text{eig}(H_2) = \varnothing$. Apply Lemma 5.1 to obtain
\[
\text{dis}(K_1 + K_{21}^T P, M_1 + M_{21}^T Q, K_2 - QK_{21}^T, M_2 - PM_{21}^T)
\]
\[
\geq \gamma - \sqrt{\| (K_{21}^T P, M_{21}^T Q) \|_F^2 + \| (K_{21}^T Q, M_{21}^T P) \|_F^2}
\]
\[
\geq \gamma - \sqrt{2(\| K_{21}^T \|_F^2 + \| M_{21}^T \|_F^2)(\| P \|_F^2 + \| Q \|_F^2)}
\]
\[
\geq \gamma - 2\delta \rho > 0,
\]
where $\rho$ is defined in (5.14). Thus apply (5.3) to conclude $\text{eig}(H_1) \cap \text{eig}(H_2) = \varnothing$.

Next we prove (5.21). Let $U_{i0} = U_i(U_i^TU_i)^{-1/2}$, and $V_{i0} = V_i(V_i^TV_i)^{-1/2}$. Then
\[
\mathcal{R}(U_1) = \mathcal{R}(U_{10}), \quad \mathcal{R}(U_1 + U_2P) = \mathcal{R}(U_{10} + U_2P(U_1^TU_1)^{-1/2}).
\]
Note $U_1^TV_2 = 0$ to see $U_{10}U_{10}^T + V_{20}V_{20}^T = I$, and thus
\[
U_{10} + U_2P(U_1^TU_1)^{-1/2} = U_{10} + (U_{10}U_{10}^T + V_{20}V_{20}^T)U_2P(U_1^TU_1)^{-1/2}
\]
\[
= U_{10}[I + U_{10}U_2P(U_1^TU_1)^{-1/2}] + V_{20}V_{20}^T U_2P(U_1^TU_1)^{-1/2}
\]
\[
= U_{10}[I + G] + V_{20}(V_2^TV_2)^{-1/2} P(U_1^TU_1)^{-1/2},
\]
where $G = U_{10}U_2P(U_1^TU_1)^{-1/2} = U_{10}U_{20}(U_2^TU_2)^{1/2} P(U_1^TU_1)^{-1/2}$. We have
\[
\| G \|_2 \leq \cos \theta_U \frac{\sigma_{\max}(U_2)}{\sigma_{\min}(U_1)} \| P \|_2
\]
\[ \leq \cos \theta_U \frac{\sigma_{\text{max}}(U_2)}{\sigma_{\text{min}}(U_1)} \|(P, Q)\|_F \]

\[ \leq 2 \cos \theta_U \frac{\sigma_{\text{max}}(U_2)}{\sigma_{\text{min}}(U_1)} \delta < \eta < 1, \]

where the last inequality holds because of (5.20). Thus

\[ \mathcal{R}(U_1 + U_2P) = \mathcal{R}(U_{10} + U_2P(U_1^T U_1)^{-1/2}) \]

\[ = \mathcal{R}(U_{10} + V_{20}(V_2^T V_2)^{-1/2}P(U_1^T U_1)^{-1/2}(I + G)^{-1}) \]

which gives

\[ \| \tan \Theta_U \|_F = \|(V_2^T V_2)^{-1/2}P(U_1^T U_1)^{-1/2}(I + G)^{-1}\|_F \]

\[ \leq \|(V_2^T V_2)^{-1/2}\|_2 \|P\|_F \|(U_1^T U_1)^{-1/2}\|_2 \|(I + G)^{-1}\|_2 \]

\[ \leq \frac{\|P\|_F}{\sigma_{\text{min}}(V_2) \sigma_{\text{min}}(U_1)} \frac{1}{1 - \eta}. \]

Similarly, we can prove (5.21b). \qed

6 Conclusion

A pair of exact deflating subspaces of \( \{K, M\} \) can be used to decouple LREP (1.2) to two smaller LREP as proved in Theorem 3.1. But a pair of approximate deflating subspaces can only lead to an almost decoupling LREP (1.2) to an extent that is reflected by the thinness of \( \|K_{21}\|_2 \) and \( \|M_{21}\|_2 \) in (4.1). We have obtained in Theorem 5.4 bounds on how far the given pair of approximate deflating subspaces from a nearby pair of exact deflating subspaces of \( \{K, M\} \). The results can be viewed as the extensions of Stewart’s theory [5, 6, 7] for the standard eigenvalue problem of a matrix and the generalize eigenvalue problem of a matrix pencil.

Closely related, in [10], we considered the effect of setting both \( K_{21} \) and \( M_{21} \) to 0 on the eigenvalues. Perturbation bounds in terms of the squares of \( \|K_{21}\|_2 \) and \( \|M_{21}\|_2 \) are obtained.

References


