University of Maryland College Park

Institute for Advanced Computer Studies TR–99–4095
Department of Computer Science TR–99-78

A Generalization of Saad's Theorem on Rayleigh-Ritz Approximations*

G. W. Stewart[†]

December 1999

ABSTRACT

Let (λ, x) be an eigenpair of the Hermitian matrix A of order n and let (μ, u) be a Ritz pair from a subspace $\mathcal{K}$ of $\mathbb{C}^2$. Saad has given a simple inequality bounding $\sin\angle(x, u)$ in terms of $\sin\angle(x, \mathcal{K})$. In this note we show that this inequality can be extended to an equally simple inequality for eigenspaces of non-Hermitian matrices.

*This report is available by anonymous ftp from `thales.cs.umd.edu` in the directory `pub/reports` or on the web at `http://www.cs.umd.edu/ stewart/`.

[†]Department of Computer Science and Institute for Advanced Computer Studies, University of Maryland, College Park, MD 20742. This work was supported in part by the National Science Foundation under Grant No. 970909-8426

A Generalization of Saad's Theorem on Rayleigh-Ritz Approximations

G. W. Stewart

ABSTRACT

Let (λ, x) be an eigenpair of the Hermitian matrix A of order n and let (μ, u) be a Ritz pair from a subspace $\mathcal{K}$ of $\mathbb{C}^n$. Saad has given a simple inequality bounding $\sin \angle(x, u)$ in terms of $\sin \angle(x, \mathcal{K})$. In this note we show that this inequality can be extended to an equally simple inequality for eigenspaces of non-Hermitian matrices.

Let A a be a Hermitian matrix of order n. If n is very large, we will not be able to compute the eigensystem of A and must be content with computing a few of its eigenpairs. The usual methods for doing this — e.g., the Lanzcos and Jacobi–Davidson methods — produce a sequence of subspaces $\mathcal{K}_k$ that contain increasingly accurate approximations to the eigenvectors in question. Approximations to the corresponding eigenpairs of A are extracted from $\mathcal{K}$ by computing *Ritz pairs* (μ, u) defined by the conditions

$$\begin{array}{ll} 1. & u \in \mathcal{K}_k, \\ 2. & Au - \mu u \perp \mathcal{K}_k. \end{array} \tag{1}$$

It is easy to see that if K_k is an orthonormal basis for $\mathcal{K}_k$, then every Ritz pair (μ, u) has the form $K_k y$, where $K_k^{\mathrm{H}} A K_k y = \mu y$.[1]

Saad [2, Theorem 4.6] has proven the following theorem relating eigenpairs of A to Ritz pairs. For now we drop the subscript on $\mathcal{K}_k$.

Theorem 1 (Saad). *Let (λ, x) be an eigenpair of A and (μ, u) be a Ritz pair with respect to the subspace $\mathcal{K}$. Let $P_{\mathcal{K}}$ be the orthogonal projection on $\mathcal{K}$, and for nonzero v define*

$$\angle(v, \mathcal{K}) = \min_{\substack{w \in \mathcal{K} \\ w \neq 0}} \angle(v, w).$$

Then

$$\sin \angle(x, u) \le \sin \angle(x, \mathcal{K}) \sqrt{1 + \frac{\gamma^2}{\delta^2}},$$

[1] Throughout this paper a calligraphic letter will stand for a subspace and the corresponding capital Roman letter will stand for a matrix containing an orthonormal basis for the subspace.

where

$$\gamma = \|P_{\mathcal{K}} A(I - P_{\mathcal{K}})\|_2$$

and δ is the distance between λ and the Ritz values other than μ.

The purpose of this note is to extend this theorem to eigenspaces (invariant subpaces) of non-Hermitian matrices. Specifically, $\mathcal{X}$ is an eigenspace of A if $A\mathcal{X} \subset \mathcal{X}$. If X is an orthonormal basis for $\mathcal{X}$, then

$$AX = XL, \quad \text{where } L = X.$$

We say that (L, X) is an eigenpair of A with eigenbasis X and eigenblock L. In what follows we will assume that all eigenbases are orthonormal.

The definition (1) of Ritz pairs extends to eigenspaces. Specifically the pair (M, U), where U is orthonormal, is a Ritz pair if

1. $\mathcal{U} \subset \mathcal{K}$,
2. $AU - UM \perp \mathcal{K}$.

Again it is easy to see that (M, U) is a Ritz pair if and only if $U = KY$, where (M, Y) is an eigenpair of $K^{\mathrm{H}}AK$.

In what follows we will use canonical angles to measure the distance between subspaces. Specifically, let (L, X) be an eigenpair of A and let Q be an orthonormal basis for the orthogonal complement of $\mathcal{K}$. Then the singular values of $W^{\mathrm{H}}X$ are the sines of the canonical angles between $\mathcal{K}$ and $\mathcal{X}$. For more on canonical angles, see [3].

It will be easier to establish our generalization of Saad's theorem if we change our coordinate system to one in which the matrices bearing the canonical angles appear explicitly. As above, let (M, U) be a Ritz pair lying in $\mathcal{K}$. Let $(U\ V\ W)$ be a unitary matrix in which the columns of V span the orthogonal complement of $\mathcal{U}$ in $\mathcal{K}$. Then we may write

$$\begin{pmatrix} U^{\mathrm{H}} \\ V^{\mathrm{H}} \\ W^{\mathrm{H}} \end{pmatrix} A(U\ \ V\ \ W) = \begin{pmatrix} M & B_{12} & B_{13} \\ B_{21} & N & B_{23} \\ B_{31} & B_{32} & B_{33} \end{pmatrix}$$

It is readily verified that, since (M, U) is a Ritz pair, B_{21} must be zero.

Now let (L, X) be an eigenpair of A and transform X into the UVW coordinate system:

$$\begin{pmatrix} U^{\mathrm{H}} \\ V^{\mathrm{H}} \\ W^{\mathrm{H}} \end{pmatrix} U = \begin{pmatrix} P \\ Q \\ R \end{pmatrix}.$$

Then the singular values of R are the sines of the angles between $\mathcal{X}$ and $\mathcal{K}$, while the singular values of $(Q^{\mathrm{H}}\ \ R^{\mathrm{H}})^{\mathrm{H}}$ are the sines of the angles between $\mathcal{X}$ and $\mathcal{U}$. Thus our problem is to bound

$$\left\| \begin{pmatrix} Q \\ R \end{pmatrix} \right\|$$

in terms of $\|R\|$. In what follows, $\|\cdot\|$ will denote either the 2-norm or the Frobenius norm. Note that for either norm

$$\left\| \begin{pmatrix} Q \\ R \end{pmatrix} \right\|^2 \le \|Q\|^2 + \|R\|^2, \tag{2}$$

with equality holding for the Frobenius norm.

In the UVW coordinate system, the relation $AX = XL$ becomes

$$\begin{pmatrix} M & B_{12} & B_{13} \\ 0 & N & B_{23} \\ B_{31} & B_{32} & B_{33} \end{pmatrix} \begin{pmatrix} P \\ Q \\ R \end{pmatrix} = \begin{pmatrix} P \\ Q \\ R \end{pmatrix} L.$$

Hence

$$\begin{pmatrix} M & B_{12} \\ 0 & N \end{pmatrix} \begin{pmatrix} P \\ Q \end{pmatrix} - \begin{pmatrix} P \\ Q \end{pmatrix} L = - \begin{pmatrix} B_{13} \\ B_{32} \end{pmatrix} R. \tag{3}$$

The right-hand side (3) can be bounded above as follows:

$$\left\| \begin{pmatrix} B_{13} \\ B_{23} \end{pmatrix} R \right\| \le \left\| \begin{pmatrix} B_{13} \\ B_{23} \end{pmatrix} \right\| \|R\| \equiv \eta \|R\|. \tag{4}$$

The left-hand side can be bounded below as follows:

$$\begin{aligned} \left\| \begin{pmatrix} M & B_{12} \\ 0 & N \end{pmatrix} \begin{pmatrix} P \\ Q \end{pmatrix} - \begin{pmatrix} P \\ Q \end{pmatrix} L \right\| &= \left\| \begin{pmatrix} MP - PL + B_{12}Q \\ NQ - QL \end{pmatrix} \right\| \\ &\ge \|NQ - QL\| \\ &\ge \operatorname{sep}(N, L) \|Q\|, \end{aligned} \tag{5}$$

where

$$\operatorname{sep}(N, L) = \inf_{\|Z\|=1} \|NZ - ZL\|.$$

Combining (5) and (4), we have

$$\|Q\| \le \eta \frac{\|R\|}{\operatorname{sep}(N, L)}.$$

It follows from (2) that

$$\left\|\begin{pmatrix} Q \\ R \end{pmatrix}\right\| \leq \|R\|\sqrt{1+\frac{\eta^2}{\mathrm{sep}(N,L)^2}}.$$

Written in terms of angles this inequality becomes

$$\sin\angle(\mathcal{X},\mathcal{U}) \leq \sin\angle(\mathcal{X},\mathcal{K})\sqrt{1+\frac{\eta^2}{\mathrm{sep}(N,L)^2}}\,.$$

We now will transform this inequality back into the original coordinate system. Clearly, $\eta = \|P_{\mathcal{K}}A(I-P_{\mathcal{K}})\|$. The quantity sep appears to depend on the choices of the bases X and V. However, because $\|\cdot\|$ is unitarily invariant, it is easy to show that in fact it depends only on the subspaces spanned by X and V. Hence we will define

$$\mathrm{sep}(\mathcal{V},\mathcal{X}) = \inf_{\|Z\|=1} \|(V^{\mathrm{H}}AV)Z - Z(X^{\mathrm{H}}AX)\|, \tag{6}$$

where V and X are arbitrary orthonormal bases for $\mathcal{V}$ and $\mathcal{X}$.

We are now in a position to state our final result.

Theorem 2. *Let $\mathcal{X}$ be an eigenspace of A. Let $\mathcal{U}$ be a Ritz subspace in $\mathcal{K}$ of the same dimension as $\mathcal{X}$ and let $\mathcal{V}$ be the orthogonal complement of $\mathcal{U}$ in $\mathcal{K}$. Then*

$$\sin\angle(\mathcal{X},\mathcal{U}) \leq \sin\angle(\mathcal{X},\mathcal{K})\sqrt{1+\frac{\eta^2}{\mathrm{sep}(\mathcal{V},\mathcal{X})^2}},$$

where $\eta = \|P_{\mathcal{K}}A(I-P_{\mathcal{K}})\|$ and sep *is defined by* (6).

This theorem is an exact analogue of Theorem 1. The major difference is the replacement of δ by $\mathrm{sep}(\mathcal{V},\mathcal{X})$. This latter quantity is bounded above by the physical separation of the spectra of M and L, but it can be much smaller. If A is Hermitian and we use the Frobenius norm, then it is the physical separation, so that our generalization reduces to Theorem 1. This is not generally true of the 2-norm unless one assumes that the eigenvalues of $X^{\mathrm{H}}AX$ lie inside an interval and the eigenvalues of $V^{\mathrm{T}}AV$ lie outside that interval or vice versa (see [3, Lemma V.3.5]).

We now return to the case where we have a sequence $\mathcal{K}_k$ of subspaces having increasingly accurate approximations to the eigenspace $\mathcal{X}$; i.e., $\sin\angle(\mathcal{X},\mathcal{K}_k) \to 0$. Theorem 2 is not strong enough to formally prove the convergence of Ritz pairs from the $\mathcal{K}_k$. There are two reasons. First, the theorem does not single out any one of the many possible Ritz pairs, so that one must be chosen by ad-hoc methods. Second, the quantity $\mathrm{sep}(\mathcal{V}_k,\mathcal{X})$ may converge to zero, something that we cannot check, since we do not know $\mathcal{X}$.

A treatment of these problems is given in [1]. We should stress, however, that they not of great concern in practice. Typically the person doing the computation will have predetermined the eigenpairs he or she wants — e.g., the eigenpairs with largest real parts — and this will direct the choice of Ritz pairs. Moreover, as the Ritz pairs converge, they can be used to approximate $\text{sep}(\mathcal{V}_k, \mathcal{X})$, which is Lipschitz continuous with constant one.

References

[1] Z. Jia and G. W. Stewart. An analysis of the Rayleigh–Ritz method for approximating eigenspaces. Technical Report TR–4015, Department of Computer Science, University of Maryland, College Park, 1999. To appear in *Mathematics of Computation.*.

[2] Y. Saad. *Numerical Methods for Large Eigenvalue Problems: Theory and Algorithms.* John Wiley, New York, 1992.

[3] G. W. Stewart and J.-G. Sun. *Matrix Perturbation Theory.* Academic Press, New York, 1990.