University of Maryland College Park

Institute for Advanced Computer Studies TR–99–08
Department of Computer Science TR–3986

On the Convergence of Ritz Values, Ritz Vectors, and Refined Ritz Vectors[*]

Zhongxiao Jai[†]
G. W. Stewart[‡]

January, 1999

ABSTRACT

This paper concerns the Rayleigh–Ritz method for computing an approximation to an eigenpair (λ, x) of a non-Hermitian matrix A. Given a subspace $\mathcal{W}$ that contains an approximation to x, this method returns an approximation $(\mu, \tilde{x})$ to (λ, x). We establish four convergence results that hold as the deviation ϵ of x from $\mathcal{W}$ approaches zero. First, the Ritz value μ converges to λ. Second, if the residual $A\tilde{x} - \mu\tilde{x}$ approaches zero, then the Ritz vector $\tilde{x}$ converges to x. Third, we give a condition on the eigenvalues of the Rayleigh quotient from which the Ritz pair is computed that insures convergence of the Ritz vector. Finally, we show that certain refined Ritz vectors, introduced by the first author, converge unconditionally.

[*]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 Applied Mathematics, Dalian University of Technology, Dalian 116024, P.R. China, (`zxjia@gingko.dlut.edu.cn`). Work supported by the China State Major Key Project for Basic Researches and the Doctoral Point Program of the Ministry of Education, P. R. China.

[†]Department of Computer Science and Institute for Advanced Computer Studies, University of Maryland, College Park, MD 20742. Work supported by the National Science Foundation under Grant No. 970909-8562

On the Convergence of Ritz Values, Ritz Vectors, and Refined Ritz Vectors

Zhongxiao Jia* G. W. Stewart†

Abstract

This paper concerns the Rayleigh–Ritz method for computing an approximation to an eigenpair (λ, x) of a non-Hermitian matrix A. Given a subspace $\mathcal{W}$ that contains an approximation to x, this method returns an approximation $(\mu, \tilde{x})$ to (λ, x). We establish four convergence results that hold as the deviation ϵ of x from $\mathcal{W}$ approaches zero. First, the Ritz value μ converges to λ. Second, if the residual $A\tilde{x} - \mu\tilde{x}$ approaches zero, then the Ritz vector $\tilde{x}$ converges to x. Third, we give a condition on the eigenvalues of the Rayleigh quotient from which the Ritz pair is computed that insures convergence of the Ritz vector. Finally, we show that certain refined Ritz vectors, introduced by the first author, converge unconditionally.

1. Introduction

This paper is concerned with computing a simple eigenpair (λ, x) of a matrix A. If A is large and sparse a general strategy for approximating (λ, x) is to produce a sequence of subspaces $\mathcal{W}_k$ that contain increasingly accurate approximations to x. There are a number of methods for accomplishing this — e.g., the Arnoldi method, the nonsymmetric Lanczos method, subspace iteration method, the Jacobi–Davidson method (for more on these methods see [**?**, 11]).

A central problem in all these methods is how to extract approximations to λ and x from the subspaces $\mathcal{W}_k$. A widely used technique for accomplishing this is called the Rayleigh–Ritz procedure (it is also an example of the more general Galerkin technique).

*Department of Applied Mathematics, Dalian University of Technology, Dalian 116024, P.R. China, (`zxjia@gingko.dlut.edu.cn`). Work supported by the China State Major Key Project for Basic Researches and the Doctoral Point Program of the Ministry of Education, P. R. China.

†Department of Computer Science and Institute for Advanced Computer Studies, University of Maryland, College Park, MD 20742, USA (`stewart@cs.umd.edu`). Work supported by the National Science Foundation under Grant No. 970909-8562

In its simplest form the technique goes as follows (here we drop the iteration subscript).

$$\begin{array}{ll} 1. & \text{Compute an orthonormal basis } W \text{ for } \mathcal{W}. \\ 2. & \text{Compute } B = W^{\mathrm{H}}AW. \\ 3. & \text{Let } (\mu, z) \text{ be an eigenpair of } B, \text{ where } \mu \cong \lambda. \\ 4. & \text{Take } (\mu, \tilde{x}) = (\mu, Wz) \text{ as the approximate eigenpair.} \end{array} \tag{1.1}$$

The matrix B is called a Rayleigh quotient. The number μ is called a Ritz value and the vector $\tilde{x} = Wz$ is called a Ritz vector. The informal justification for the method is that if $x \in \mathcal{W}$ then there is an eigenpair (λ, z) of B with $x = Wz$. Continuity suggests that if x is nearly in $\mathcal{W}$ then there should be an eigenpair (μ, z) of B with μ near λ and Wz near x.

When A non-Hermitian, one of the authors (Jia [4, 5, 6]) has established a priori error bounds for Ritz values and Ritz vectors in terms of the deviation of x from $\mathcal{W}$. The results show that Ritz values converge. The Ritz vectors, on the other hand, behave more erratically and may even fail to converge. This led the first author to introduce certain refined Ritz vectors for which the continuity argument is valid [3, 5]. Unfortunately, the results just cited are proved under the restrictive hypothesis that the eigenvalues of A are distinct. One of the contributions of this paper is to remove this restriction.

Our general approach is to derive bounds in terms of the deviation ϵ of x from the subspace $\mathcal{W}$ defined below in (1.3). (In practice, of course, the size of this deviation must be established from the properties of the underlying algorithm that determines $\mathcal{W}$.) We will establish four convergence results. First, as $\epsilon \to 0$, the Rayleigh quotient B contains a Ritz value μ that converges to the eigenvalue λ. Second, although the corresponding Ritz vector $\tilde{x}$ need not converge, if the residual $A\tilde{x} - \mu\tilde{x}$ converges to zero, then the Ritz vector also converges. Thus we have a practical way of telling when the Ritz vector corresponding to a converging Ritz value is itself converging. Third, it is shown that if μ remains well separated in an appropriate sense from the other eigenvalues of B, then the Ritz vector converges. Finally, the refined Ritz vector converges unconditionally as $\epsilon \to 0$. Thus refined Ritz vectors provide a useful alternative for approximating eigenvectors.

Background material for this paper can be found in [2, 12]. The norm $\|\cdot\|$ will denote both the Euclidean vector norm and the subordinate spectral matrix norm. Throughout, we will assume that eigenvectors, Ritz vectors and refined Ritz vectors have been normalized to have norm one.

We will measure the deviation of an a normalized vector y from a subspace $\mathcal{X}$ as follows. Let X be an orthonormal basis for the space $\mathcal{X}$ and let $X_{\perp}$ be an orthonormal basis for the orthogonal complement of $\mathcal{X}$. We will measure the deviation of y from $\mathcal{X}$

by the quantity

$$\sin \angle(y, \mathcal{X}) = \|X_\perp^{\mathrm{H}} y\| = \|(I - P_{\mathcal{X}})y\|, \tag{1.2}$$

where $P_{\mathcal{X}}$ is the orthogonal projection onto $\mathcal{X}$. As the notation indicates, this number is sine of the angle between y and $\mathcal{X}$. For brevity, we will denote the sine of the angle between our fixed eigenvector x and our subspace $\mathcal{W}$ by

$$\epsilon = \sin \angle(x, \mathcal{W}). \tag{1.3}$$

We will need some notation from the perturbation theory of eigenspaces. Let $X_\perp$ be an orthonormal basis for the orthogonal complement of the space spanned by x, so that $(x \;\, X_\perp)$ is unitary. Then it follows from the relation $Ax = \lambda x$ that

$$\begin{pmatrix} x^{\mathrm{H}} \\ X_\perp^{\mathrm{H}} \end{pmatrix} A(x \;\, X_\perp) = \begin{pmatrix} \lambda & h^{\mathrm{H}} \\ 0 & L \end{pmatrix}, \tag{1.4}$$

where $h^{\mathrm{H}} = x^{\mathrm{H}} A X_\perp$ and $L = X_\perp^{\mathrm{H}} A X_\perp$. Because the right-hand side of (1.4) is block triangular, the eigenvalues of A consist of λ and the eigenvalues of L. Since λ is assumed to be simple, the eigenvalues of L are not equal to λ and hence $L - \lambda I$ is nonsingular. Define

$$\mathrm{sep}(\lambda, L) = \|(L - \lambda I)^{-1}\|^{-1}, \tag{1.5}$$

where $\|\cdot\|$ is the spectral matrix norm. The quantity sep, which approaches zero as λ approaches an eigenvalue of L, will play a central role in our theory. For more on the properties of sep see [12].

In the next section we will consider the convergence of Ritz values. In §3 we will treat the convergence of Ritz vectors. In §4 we will establish the convergence of refined Ritz vectors. The paper concludes with a discussion of our results.

2. Convergence of Ritz values

It is a surprising fact that the hypothesis $\epsilon \to 0$ is sufficient to insure that B contains a Ritz value that converges to λ. We will establish this result in two stages. First we will show that if ϵ is small then λ is an exact eigenvalue of a matrix $\tilde{B}$ that is near B. We will then use a theorem of Elsner to show that B must have an eigenvalue that is near λ.

Theorem 2.1. *Let B be the Rayleigh quotient in* (1.1). *Then there is a matrix E satisfying*

$$\|E\| \le \frac{\epsilon}{\sqrt{1 - \epsilon^2}} \|A\| \tag{2.1}$$

such that λ is an eigenvalue of $B + E$.

Proof. Let $y = W^{\mathrm{H}}x$ and $y_\perp = W_\perp^{\mathrm{H}}x$. By (1.2) and (1.3), we have $\|y_\perp\| = \epsilon$. Hence $\|y\| = \sqrt{1-\epsilon^2}$.

From the relation $Ax - \lambda x = 0$ we have

$$W^{\mathrm{H}}A(W\ \ W_\perp)\begin{pmatrix} W^{\mathrm{H}} \\ W_\perp^{\mathrm{H}} \end{pmatrix} x - \lambda W^{\mathrm{H}}x = 0.$$

Equivalently,

$$By + W^{\mathrm{H}}AW_\perp y_\perp - \lambda y = 0. \tag{2.2}$$

If we normalize y by setting $\hat{y} = y/\sqrt{1-\epsilon^2}$ and set

$$r = B\hat{y} - \lambda\hat{y}, \tag{2.3}$$

then it follows from (2.2) that

$$\|r\| \le \frac{\epsilon}{\sqrt{1-\epsilon^2}}\|W^{\mathrm{H}}AW_\perp\|. \tag{2.4}$$

If we now define

$$E = -r\hat{y}^{\mathrm{H}},$$

then it is easy to verify that E satisfies (2.1) and $(B+E)\hat{y} = \lambda\hat{y}$ because W and $W_\perp$ have orthonormal columns that lead to $\|W^{\mathrm{H}}AW_\perp\| \le \|A\|$. ■

Since E goes to zero along with ϵ, it would appear that we can now deduce the convergence of the Ritz values from the continuity of eigenvalues of a matrix. However, there is a subtle point here. For the number ϵ to change, the space $\mathcal{W}$ and hence the Rayleigh quotient B must also change. Thus we must allow for the possibility that as $\epsilon \to 0$ the eigenvalue λ of $\tilde{B} = B + E$ becomes ill conditioned so swiftly that no eigenvalue of $B = \tilde{B} - E$ is near λ—in spite of the decrease in $\|E\|$.

Fortunately, a theorem of Elsner [1] (also see [12, p. 168]) implies that this cannot happen. Specifically, Elsner's theorem says that given matrices A and $\tilde{A}$ of order n for any eigenvalue λ of A there is an eigenvalue $\tilde{\lambda}$ of $\tilde{A}$ satisfying

$$|\lambda - \tilde{\lambda}| \le (\|A\| + \|\tilde{A}\|)^{1-\frac{1}{n}}\|A - \tilde{A}\|^{\frac{1}{n}}.$$

Hence we have the following corollary.

Corollary 2.2. *There is an eigenvalue μ of B such that*

$$|\lambda - \mu| \le (2\|A\| + \|E\|)^{1-\frac{1}{m}}\|E\|^{\frac{1}{m}}, \tag{2.5}$$

where m is the order of B.

The right-hand side of (2.5) depends only on $\|A\|$ and ϵ. Hence we may conclude that as $\epsilon \to 0$ there is always a Ritz value that converges to λ. The bound (2.5) will in general be a gross overestimate. If, as usually happens in practice, the condition number of μ is bounded, the convergence will be linear in ϵ.

3. Convergence of Ritz vectors

Elsner's theorem says in essence that no eigenvalue of a matrix can be infinitely ill conditioned in a global sense. Given an error, its effects on the eigenvalues are bounded, and the smaller the error the smaller the bound. The same is decidedly not true of eigenvectors. For example, as two simple eigenvalues approach one another, their eigenvectors become less and less stable. In fact, there are arbitrarily small perturbations that can be made to have arbitrarily large effects simply by bringing the eigenvalues close enough.

Fortunately, the following theorem, which applies to arbitrary pairs, shows that there is an easily checked condition that insures that the Ritz vector converges.

Theorem 3.1. *Let the (not necessarily Ritz) pair* $(\mu, \tilde{x})$ *be given and let*

$$\rho = \|A\tilde{x} - \mu\tilde{x}\|.$$

Let $\mathrm{sep}(\lambda, L)$ *be defined by* (1.5). *If*

$$\mathrm{sep}(\lambda, L) - |\mu - \lambda| > 0, \tag{3.1}$$

then

$$\sin\angle(x, \tilde{x}) \le \frac{\rho}{\mathrm{sep}(\mu, L)} \le \frac{\rho}{\mathrm{sep}(\lambda, L) - |\mu - \lambda|}. \tag{3.2}$$

Proof. Let

$$\begin{pmatrix} \eta \\ y \end{pmatrix} = \begin{pmatrix} x^{\mathrm{H}} \\ X_{\perp}^{\mathrm{H}} \end{pmatrix} \tilde{x} \quad \text{and} \quad \begin{pmatrix} \sigma \\ s \end{pmatrix} = \begin{pmatrix} x^{\mathrm{H}} \\ X_{\perp}^{\mathrm{H}} \end{pmatrix} (A\tilde{x} - \mu\tilde{x})$$

Then $\|y\| = \sin\angle(x, \tilde{x})$. Moreover, by the unitary invariance of $\|\cdot\|$,

$$\rho = \left\| \begin{pmatrix} \sigma \\ s \end{pmatrix} \right\| = \left\| \begin{pmatrix} \lambda & h^{\mathrm{H}} \\ 0 & L \end{pmatrix} \begin{pmatrix} \eta \\ y \end{pmatrix} - \mu \begin{pmatrix} \eta \\ y \end{pmatrix} \right\| = \left\| \begin{pmatrix} (\lambda - \mu)\eta + h^{\mathrm{H}} y \\ (L - \mu I) y \end{pmatrix} \right\|.$$

Now it is easily verified that

$$\mathrm{sep}(\mu, L) \ge \mathrm{sep}(\lambda, L) - |\mu - \lambda|.$$

By (3.1), $\mathrm{sep}(\mu, L) > 0$. Hence $L - \mu I$ is nonsingular and

$$\begin{aligned} \sin\angle(x, \tilde{x}) &= \|y\| \\ &= \|(L - \mu I)^{-1} s\| \\ &\le \|(L - \mu I)^{-1}\| \rho \\ &= \frac{\rho}{\mathrm{sep}(\mu, L)} \\ &\le \frac{\rho}{\mathrm{sep}(\lambda, L) - |\mu - \lambda|}. \quad \blacksquare \end{aligned}$$

We have seen that $\mu \to \lambda$ as $\epsilon \to 0$. If, in addition, the residual norm $\rho = \|A\tilde{x} - \mu\tilde{x}\|$ approaches zero, the Ritz vector converges to an eigenvector. Thus a converging Ritz value and vanishing residual imply a converging Ritz vector. These convergence conditions can be easily checked in the course of the Rayleigh-Ritz procedures.

It is instructive to see how the Ritz vector can fail to converge. Let μ be the converging Ritz value and let z be the corresponding eigenvector of B. Let $(z\ Z_\perp)$ be unitary. From the relation $Bz = \mu z$ it follows that

$$\begin{pmatrix} z^{\mathrm{H}} \\ Z_\perp^{\mathrm{H}} \end{pmatrix} B(z\ Z_\perp) = \begin{pmatrix} \mu & g^{\mathrm{H}} \\ 0 & C \end{pmatrix}.$$

Since the only assumption we have made about the subspace $\mathcal{W}$ is that it contains an approximation to the eigenvector x, the eigenvalues of C need not be near the eigenvalues of A other than λ. In particular, an eigenvalue of C could happen to be equal μ. If this double eigenvalue is not defective, its eigevectors will span a two-dimensional subspace, and it will be impossible to tell which is the one that reproduces an approximation to x. When the eigenvalue of C is near μ, there will be a unique eigenvector associated with μ, but there is no guarantee that it will reproduce an approximation to x.

The foregoing suggests that the Ritz vector will converge provided the eigenvalue μ of B remains well separated from those of C. The following theorem, which generalize a result in [4], shows that this is indeed the case.

Theorem 3.2. *Under the above assumptions, if*

$$\mathrm{sep}(\lambda, C) > 0,$$

then

$$\sin \angle(x, \tilde{x}) \le \left(1 + \frac{\|A\|}{\sqrt{1-\epsilon^2}\,\mathrm{sep}(\lambda, C)}\right)\epsilon. \tag{3.3}$$

Proof. As in the proof of Theorem 2.1, let (μ, z) be the Ritz pair corresponding to x. Let $\hat{y}$ be the normalized value of $W^{\mathrm{H}}x$. Then if $r = B\hat{y} - \lambda\hat{y}$, we have

$$\|r\| \le \frac{\epsilon}{\sqrt{1-\epsilon^2}}\|A\|$$

[see (2.4)].

From Theorem 3.1 it follows that

$$\sin \angle(z, \hat{y}) \le \frac{\|r\|}{\mathrm{sep}(\lambda, C)}.$$

Hence

$$\sin \angle(z, \hat{y}) \leq \frac{\|A\|}{\operatorname{sep}(\lambda, C)} \frac{\epsilon}{\sqrt{1-\epsilon^2}}$$

Since W is orthonormal, we have from the definition of $\tilde{x}$ and $\hat{y}$ that

$$\sin \angle(z, \hat{y}) = \sin \angle(Wz, W\hat{y}) = \sin \angle(\tilde{x}, WW^{\mathrm{H}}x) = \sin \angle(\tilde{x}, P_{\mathcal{W}}x).$$

Now

$$\angle(x, \tilde{x}) \leq \angle(x, P_{\mathcal{W}}x) + \angle(P_{\mathcal{W}}x, \tilde{x}) = \angle(x, \mathcal{W}) + \angle(\tilde{x}, P_{\mathcal{W}}x),$$

where the equality comes from the relation $\angle(x, P_{\mathcal{W}}x) = \angle(x, \mathcal{W})$. Hence

$$\sin \angle(x, \tilde{x}) \leq \sin \angle(x, \mathcal{W}) + \sin \angle(\tilde{x}, P_{\mathcal{W}}x)$$

Since $\sin \angle(x, \mathcal{W}) = \|(I - P_{\mathcal{W}})x\| = \epsilon$ and $\sin \angle(\tilde{x}, P_{\mathcal{W}}x) = \sin \angle(z, \hat{y})$, we have

$$\begin{aligned} \sin \angle(x, \tilde{x}) &\leq \epsilon + \tfrac{\|A\|}{\operatorname{sep}(\lambda,C)} \tfrac{\epsilon}{\sqrt{1-\epsilon^2}} \\ &\leq \left(1 + \tfrac{\|A\|}{\sqrt{1-\epsilon^2}\operatorname{sep}(\lambda,C)}\right)\epsilon \quad \blacksquare \end{aligned}$$

Since $\operatorname{sep}(\lambda, C) \geq \operatorname{sep}(\mu, C) - |\mu - \lambda|$, we have

$$\sin \angle(x, \tilde{x}) \leq \left(1 + \frac{\|A\|}{\sqrt{1-\epsilon^2}(\operatorname{sep}(\mu, C) - \mid \mu - \lambda \mid)}\right)\epsilon.$$

Since μ approaches λ as $\epsilon \to 0$, we see that a sufficient condition for the convergence of the Ritz vector is that $\operatorname{sep}(\mu, C)$ be uniformly bounded away from 0. This condition can also be checked during the computation of the Ritz vector.

4. Convergence of refined Ritz vectors

Theorem 3.1 does not require that $(\mu, \tilde{x})$ be a Ritz pair — only that μ be sufficiently near λ, and that $\tilde{x}$ have a sufficiently small residual. Since the Ritz value μ is known to converge to λ, this suggests that we can deal with the problem of nonconverging Ritz vectors by retaining the Ritz value and replacing the Ritz vector with a vector $\hat{x} \in \mathcal{W}$ having a suitably small residual. It is natural to choose the best such vector (as proposed in [5, 7, 8, 9]). Thus we take $\hat{x}$ to be the solution of the problem

$$\begin{array}{ll} \text{minimize} & \|(A - \mu I)\hat{x}\| \\ \text{subject to} & \hat{x} \in \mathcal{W}. \end{array}$$

Alternatively, $\hat{x} = Wv$, where v is the right singular vector of $(A - \mu I)W$ corresponding to its smallest singular value. Whatever the definition, we will call such a vector a refined Ritz vector.

The following theorem shows that the refined Ritz vectors converge as $\epsilon \to 0$.

Theorem 4.1. *If*

$$\operatorname{sep}(\mu, L) \geq \operatorname{sep}(\lambda, L) - |\mu - \lambda| > 0, \tag{4.1}$$

then

$$\sin \angle(x, \hat{x}) \leq \frac{\|A - \mu I\|\epsilon + |\lambda - \mu|}{\sqrt{1-\epsilon^2}(\operatorname{sep}(\lambda, L) - |\mu - \lambda|)}. \tag{4.2}$$

Proof. Let

$$y = \frac{P_{\mathcal{W}} x}{\sqrt{1-\epsilon^2}}$$

be the normalized projection of x onto $\mathcal{W}$ and let

$$e = (I - P_{\mathcal{W}})x.$$

Then

$$\begin{aligned}(A - \mu I)y &= \frac{(A-\mu I)P_{\mathcal{W}}x}{\sqrt{1-\epsilon^2}} \\ &= \frac{(A-\mu I)(x-e)}{\sqrt{1-\epsilon^2}} \\ &= \frac{(\lambda-\mu)x - (A-\mu I)e}{\sqrt{1-\epsilon^2}}.\end{aligned}$$

Hence

$$\|(A-\mu I)y\| \leq \frac{\|A-\mu I\|\epsilon + |\mu - \lambda|}{\sqrt{1-\epsilon^2}}.$$

By the minimality of $\hat{x}$ we have

$$\|(A-\mu I)\hat{x}\| \leq \frac{\|A-\mu I\|\epsilon + |\mu - \lambda|}{\sqrt{1-\epsilon^2}}. \tag{4.3}$$

Since $\|(A-\mu I)\hat{x}\|$ is a residual norm, (4.2) follows directly from Theorem 3.1. ■

It follows immediately from (4.2) that if $\mu \to \lambda$ as $\epsilon \to 0$ then the refined Ritz vector $\hat{x}$ converges to the eigenvector x. In particular, by Corollary 2.2 this will happen if μ is chosen to be the Ritz value.

5. Discussion

The results of the last three sections hang together nicely. Suppose we have a method that produces subspaces $\mathcal{W}$ containing increasingly accurate approximations to our distinguished eigenvector x. Corollary 2.2 says that the Rayleigh quotient B of (1.1) contains an increasingly accurate approximation μ to our distinguished eigenvalue λ. Theorem 3.1 says that if the residual $A\tilde{x} - \mu\tilde{x}$ of the Ritz vector $\tilde{x}$ converges to zero, then the Ritz vector itself converges to x. However, Theorem 3.2 shows that the Ritz vector $\tilde{x}$ can fail to converge to x. If the convergence of the residual stagnates, we can switch to a refined Ritz vector, in which case Theorem 4.1 assures its convergence to x.

Computationally, it is important to be able to detect converging Ritz vectors. If A is of order n and the dimension of $\mathcal{W}$ is m, a Rayleigh–Ritz procedure requires $O(nm^2)$ operations to produce a complete set of m Ritz vectors. On the other hand, to compute a refined Ritz vector to requires the computation of the singular value decomposition of $(A - \mu I)W$, which also requires $O(nm^2)$ operations. Thus in general the computation of a single refined Ritz vector requires the same order of work as the an entire Rayleigh-Ritz procedure. Fortunately, if $\mathcal{W}$ is determined by a Krylov sequence, as in the Arnoldi method, this work can be reduced to $O(m^3)$ [5]. Moreover, if we are willing to sacrifice some accuracy, we can compute the refined vectors from the cross product matrices $W^{\mathrm{H}}A^{\mathrm{H}}AW$ and $W^{\mathrm{H}}AW$ with $O(m^3)$ work [9]. Thus in some important cases, the computation of refined Ritz vectors is a viable alternative to the Rayleigh-Ritz procedure.

A nice feature of our bounds is that λ is linked to the other eigenvalues of A only by the quantity $\mathrm{sep}(\lambda, L)$, and μ is linked to the other eigenvalues of the Rayleigh quotient B only by the quantity $\mathrm{sep}(\mu, C)$. In particular, this implies that we do not have to assume that the other eigenvalues are simple — i.e., that the matrix in question is diagonalizable.

An unsavory aspect of Corollary 2.2 is that it depends on mth root of E, which goes to zero slowly. As we pointed out, the bound is unrealistic in practice, and a more realistic bound would depend on the condition of the eigenvalue μ in B. We can, if we wish, compute a condition number for μ in the course of the Rayleigh–Ritz procedure.

The bound (3.2) in Theorem 3.1 is called a residual bound. The method used to establish it is quite different from that of other residual bounds that have appeared in the literature.

We have analyzed only the the simplest version of Rayleigh–Ritz procedure. In some forms of the method, the Rayleigh quotient is defined by $V^{\mathrm{H}}AW$, where $V^{\mathrm{H}}W = 1$. We expect that the above results will generalize easily to this case provided the product $\|V\|\|W\|$ remains uniformly bounded.

It may happen that the matrix A has a cluster of close eigenvalues, in which case it does not make sense to compute individual eigenvectors. Instead one should compute an eigenspace corresponding to these eigenvalues. In the context of this paper, this amounts to replacing λ and x with matrices. Since we have made free use of the fact that λ is a scalar in deriving our results, we feel that it may be difficult to extend them to the computation of eigenspaces. The attempt, however, is worth making.

References

[1] L. Elsner, *On the variations of the spectra of matrices*, Linear Algebra Appl., 47 (1982), pp. 127–138.

[2] G. H. Golub and C. F. van Loan, *Matrix Computations*, the 3rd ed., The Johns Hopkins University Press, Baltimore and London, 1996.

[3] Z. Jia, *Some numerical methods for large unsymmetric eigenproblems*, Ph.D. thesis, University of Bielefeld, Feb., 1994.

[4] Z. Jia, *The convergence of generalized Lanczos methods for large unsymmetric eigenproblems*, SIAM J. Matrix Anal. Appl., 16 (1995), pp. 843–862.

[5] Z. Jia, *Refined iterative algorithms based on Arnoldi's process for large unsymmetric eigenproblems*, Linear Algebra Appl., 259 (1997), pp. 1–23.

[6] Z. Jia, *Generalized block Lanczos methods for large unsymmetric eigenproblems*, Numer. Math., 80 (1998), pp. 239–266.

[7] Z. Jia, *A refined iterative algorithm based on the block Arnoldi process for large unsymmetric eigenproblems*, Linear Algebra Appl., 270 (1998), pp. 171–189.

[8] Z. Jia, *Polynomial characterizations of the approximate eigenvectors by the refined Arnoldi method and an implicitly restarted refined Arnoldi algorithm*, Linear Algebra Appl., 287 (1999), pp. 191–214.

[9] Z. Jia, *A refined subspace iteration algorithm for large sparse eigenproblems*, submitted.

[10] Y. Saad, *Numerical Methods for Large Eigenvalue Problems*, Manchester University Press, 1992.

[11] G. L. G. Sleijpen and H. A. van der Vorst, *A Jacobi-Davidson iteration method for linear eigenvalue problems*, SIAM J. Matrix Anal. Appl., 17 (1996), pp. 401–425.

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