University of Maryland College Park

Institute for Advanced Computer Studies TR–99–24
Department of Computer Science TR–4015

An Analysis of the Rayleigh–Ritz Method for Approximating Eigenspaces*

Zhongxiao Jia† G. W. Stewart‡

May 1999

ABSTRACT

This paper concerns the Rayleigh–Ritz method for computing an approximation to an eigenspace $\mathcal{X}$ of a general matrix A from a subspace $\mathcal{W}$ that contains an approximation to $\mathcal{X}$. The method produces a pair $(N, \tilde{X})$ that purports to approximate a pair (L, X), where X is a basis for $\mathcal{X}$ and $AX = XL$. In this paper we consider the convergence of $(N, \tilde{X})$ as the sine ϵ of the angle between $\mathcal{X}$ and $\mathcal{W}$ approaches zero. It is shown that under a natural hypothesis — called the uniform separation condition — the Ritz pairs $(N, \tilde{X})$ converge to the eigenpair (L, X). When one is concerned with eigenvalues and eigenvectors, one can compute certain refined Ritz vectors whose convergence is guaranteed, even when the uniform separation condition is not satisfied. An attractive feature of the analysis is that it does not assume that A has distinct eigenvalues or is diagonalizable.

*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@dlut.edu.cn`). Work supported by the China State Major Key Project for Basic Researches, the National Natural Science Foundation of China, the Foundation for Excellent Young Scholars of the Ministry of Education and the Doctoral Point Program of the Ministry of Education, 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

An Analysis of the Rayleigh–Ritz Method for Approximating Eigenspaces

Zhongxiao Jia* G. W. Stewart†

Abstract

This paper concerns the Rayleigh–Ritz method for computing an approximation to an eigenspace $\mathcal{X}$ of a general matrix A from a subspace $\mathcal{W}$ that contains an approximation to $\mathcal{X}$. The method produces a pair $(N, \tilde{X})$ that purports to approximate a pair (L, X), where X is a basis for $\mathcal{X}$ and $AX = XL$. In this paper we consider the convergence of $(N, \tilde{X})$ as the sine ϵ of the angle between $\mathcal{X}$ and $\mathcal{W}$ approaches zero. It is shown that under a natural hypothesis — called the uniform separation condition — the Ritz pairs $(N, \tilde{X})$ converge to the eigenpair (L, X). When one is concerned with eigenvalues and eigenvectors, one can compute certain refined Ritz vectors whose convergence is guaranteed, even when the uniform separation condition is not satisfied. An attractive feature of the analysis is that it does not assume that A has distinct eigenvalues or is diagonalizable.

1. Introduction

Many methods for finding eigenvalues and eigenvectors of a large matrix A proceed by generating a sequence of subspaces $\mathcal{W}_k$ containing increasingly accurate approximations to the desired eigenvectors. There are a number of methods for accomplishing this — e.g., the Arnoldi method, the nonsymmetric Lanczos method, subspace iteration, and the Jacobi–Davidson method (for more on these methods see [11, 12]).

A central problem in all these methods is how to extract approximations to the desired eigenvalues and eigenvectors from the subspace $\mathcal{W}_k$. A widely used technique for accomplishing this is called the Rayleigh–Ritz procedure (it is also an example of

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

the more general Galerkin technique). In its simplest form the technique retrieves an approximation to a simple eigenpair (λ, x) as follows (here we drop the iteration subscript).

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

Of course, steps 3 and 4 can be repeated to extract approximations to other eigenpairs.

The matrix B is called a Rayleigh quotient. The number ν 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 (ν, z) of B with ν near λ and Wz near x.

When A is non-Hermitian, one of the authors (Jia [5, 6, 7, 8]) 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 [4, 6, 8]. The refined Ritz vectors have been used in some other cases [9, 10]. Unfortunately, the results just cited were proved under the restrictive hypothesis that the eigenvalues of A are distinct or A is diagonalizable. One of the contributions of this paper is to remove this restriction.

When A has a cluster of very close or multiple eigenvalues, the corresponding eigenvectors are ill-determined, and it makes better sense to approximate the eigenspace $\mathcal{X}$ spanned by the vectors instead of the vectors themselves. The Rayleigh–Ritz procedure can be extended to do this [see (3.1) below]. A second contribution of this paper is to provide an analysis of this extended procedure. In fact, essentially the same results hold for eigenspaces as for eigenvectors, and we will present our results at this level of generality.

Our approach is to derive bounds in terms of the sine ϵ of the angle between $\mathcal{X}$ and the subspace $\mathcal{W}$. (In practice, of course, the size of ϵ must be established from the properties of the underlying algorithm that determines $\mathcal{W}$.) Sections 2 and 3 are devoted to background material and setting the stage for our analysis. In §4 we will consider the convergence of the Ritz values of the Rayleigh quotient. In §5 we will establish the convergence of Ritz pairs under a hypothesis that is computationally verifiable. In §6 we treat the relation of residual vectors to the accuracy of an approximate eigenspace, and in §7, we consider the convergence of the refined Ritz vectors mentioned above. The paper concludes with a brief summary.

2. Preliminaries

Background material for this paper can be found in [2, 13]. The norm $\|\cdot\|$ will denote both the Euclidean vector norm and the subordinate spectral matrix norm. We will denote the spectrum of a matrix A by the multiset $\Lambda(A)$.

A subspace $\mathcal{X}$ is an eigenspace (or invariant subspace) of A if $A\mathcal{X} \subset \mathcal{X}$. If X is a basis for $\mathcal{X}$, then there is a unique matrix L such that

$$AX = XL, \tag{2.1}$$

and conversely if X has linearly independent columns and satisfies (2.1), then the column space of X is an eigenspace of A. In this case, we say that (L, X) is an eigenpair of A, that the matrix X is an eigenbasis of A, and that L is its corresponding eigenblock. The eigenpair is simple if its eigenvalues are distinct from the other eigenvalues of A. It is orthonormal if $X^{\mathrm{H}}X = I$, in which case its eigenblock is given by the Rayleigh quotient $L = X^{\mathrm{H}}AX$. Throughout this paper we will assume that eigenpairs are orthonormal.

We will measure the deviation of a subspace $\mathcal{X}$ from a subspace $\mathcal{W}$ as follows. Let X be an orthonormal basis for $\mathcal{X}$, W an orthonormal basis for $\mathcal{W}$, and $W_\perp$ an orthonormal basis for the orthogonal complement of $\mathcal{W}$. Then we define

$$\sin\angle(\mathcal{X}, \mathcal{W}) = \sin(X, W) = \|W_\perp^{\mathrm{H}} X\|. \tag{2.2}$$

This measure is a metric on any space of subspaces of fixed dimension. If the dimension of $\mathcal{W}$ is greater than that of $\mathcal{X}$, the measure is not symmetric in its arguments; in fact, $\sin\angle(\mathcal{W}, \mathcal{X}) = 1$, although it may happen that $\sin\angle(\mathcal{X}, \mathcal{W}) < 1$.

We will cast our results in terms of a function sep that in some sense measures the distance between the spectra of two matrices. Specifically, let L and M be matrices of order ℓ and m, and define

$$\mathrm{sep}(L, M) = \min_{\|P\|=1} \|PL - MP\|. \tag{2.3}$$

Alternatively, let $\mathbf{S}$ be the Sylvester operator defined by

$$\mathbf{S}P = PL - MP.$$

Then $\mathbf{S}$ is a linear operator whose eigenvalues are

$$\Lambda(\mathbf{S}) = \Lambda(L) - \Lambda(M);$$

i.e., its eigenvalues are the pairwise differences of the eigenvalues of L and M. If

$$\delta \equiv \min|\Lambda(\mathbf{S})| > 0$$

then $\mathbf{S}$ is nonsingular and it follows from (2.3) that $\mathrm{sep}^{-1}(L, M) = \|\mathbf{S}^{-1}\| \geq \delta$. Hence

$$\mathrm{sep}(L, M) = \|\mathbf{S}^{-1}\|^{-1} \leq \delta;$$

i.e., $\mathrm{sep}(L, M)$ is not greater than the physical separation δ of the spectra of L and M. Unfortunately, $\mathrm{sep}(L, M)$ can be much smaller than δ.

The function sep has an important advantage over δ—it is Lipschitz continuous. Specifically

$$\mathrm{sep}(L, M) - \|E\| - \|F\| \leq \mathrm{sep}(L + E, M + F) \leq \mathrm{sep}(L, M) + \|E\| + \|F\|. \tag{2.4}$$

Finally, we note that because $\|\cdot\|$ is unitarily invariant if U and V are unitary then

$$\mathrm{sep}(U^{\mathrm{H}} L U, V^{\mathrm{H}} M V) = \mathrm{sep}(L, M). \tag{2.5}$$

We conclude this section with a theorem of Elsner, which will be used to establish the convergence of Ritz values.

Theorem 2.1. *Let the eigenvalues of A be $\lambda_1, \ldots, \lambda_n$ and let the eigenvalues of $\tilde{A} = A + E$ be $\tilde{\lambda}_1, \ldots, \tilde{\lambda}_n$. Then there is a permutation $j_1, \ldots, j_n$ of the integers $1, \ldots, n$ such that*

$$|\lambda_i - \tilde{\lambda}_{j_i}| \leq n(\|A\| + \|\tilde{A}\|)^{1-\frac{1}{n}} \|E\|^{\frac{1}{n}}, \qquad i = 1, \ldots, n.$$

3. The setting

As we indicated in the introduction, we are concerned with the approximation of a simple eigenpair (L, X) by the Rayleigh–Ritz method applied to a subspace $\mathcal{W}$. We will consider the following generalization of the method (1.1) in the introduction.

$$\begin{array}{ll} 1. & \text{Compute an orthonormal basis } W \text{ for } \mathcal{W}. \\ 2. & \text{Compute } B = W^{\mathrm{H}} A W. \\ 3. & \text{Let } (N, Z) \text{ be an eigenpair of } B \text{, where } \Lambda(N) \cong \Lambda(L). \\ 4. & \text{Take } (N, \tilde{X}) = (N, WZ) \text{ as the approximate eigenpair.} \end{array} \tag{3.1}$$

For definiteness we will suppose that the dimensions of $\mathcal{X}$ and $\mathcal{W}$ are ℓ and p, so that the eigenblock L is of order ℓ and the Rayleigh quotient B is of order p. Denoting the column space of X by $\mathcal{X}$, we will set

$$\epsilon = \sin \angle(\mathcal{X}, \mathcal{W}) \tag{3.2}$$

and examine the behavior of the method as $\epsilon \to 0$.

In the sequel the representation of X in the coordinate system specified by W will play a central role. As above, let the columns of $W_\perp$ form an orthonormal basis for $\mathcal{W}_\perp$. Then X can be written in the form

$$X = WY + W_\perp Y_\perp,$$

where

$$Y = W^{\mathrm{H}} X \quad \text{and} \quad Y_\perp = W_\perp^{\mathrm{H}} X.$$

By definition $\|Y_\perp\| = \epsilon$.

The columns of Y can be orthonormalized by setting

$$\hat{Y} = YQ, \quad \text{where} \quad Q = (Y^{\mathrm{H}} Y)^{-\frac{1}{2}}. \tag{3.3}$$

Since $Y^{\mathrm{H}} Y + Y_\perp^{\mathrm{H}} Y_\perp = I$, it follows that

$$\|I - Q\| = 1 - \frac{1}{\sqrt{1-\epsilon^2}} \cong \frac{1}{2}\epsilon^2 \quad \text{and} \quad \|Q^{-1}\| = \frac{1}{\sqrt{1-\epsilon^2}} \cong 1 + \frac{1}{2}\epsilon^2.$$

Note that since $X - W\hat{Y} = WY_\perp + WY(I - Q)$

$$\|X - W\hat{Y}\| \le \epsilon + O(\epsilon^2). \tag{3.4}$$

4. The convergence of Ritz values

Although we will be primarily concerned with Ritz pairs, the only effective way of choosing a pair from a Rayleigh quotient B is to examine the eigenvalues of B. We therefore need to know when the eigenvalues of a Rayleigh quotient converge.

It is a surprising fact that the hypothesis $\epsilon \to 0$ is by itself sufficient to insure that B in the algorithm (3.1) contains Ritz values that converge to the eigenvalues of L. We will establish this result in two stages. First we will show that if ϵ is small then $\Lambda(L)$ is a subset of the spectrum of a matrix $\tilde{B}$ that is near B. We will then use Elsner's theorem to show that B must have eigenvalues that are near those of L.

Theorem 4.1. *Let B be the Rayleigh quotient in* (3.1)*, and let $\hat{Y}$ and Q be defined by* (3.3)*. Then there is a matrix E satisfying*

$$\|E\| \le \frac{\epsilon}{\sqrt{1-\epsilon^2}}\|A\| \tag{4.1}$$

such that $(Q^{-1}LQ, \hat{Y})$ is an eigenblock of $B + E$.

Proof. From the relation $AX - XL = 0$ we have

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

Equivalently,

$$BY + W^{\mathrm{H}}AW_\perp Y_\perp - YL = 0.$$

Postmultiplying by Q, we have

$$B\hat{Y} + W^{\mathrm{H}}AW_\perp Y_\perp Q - \hat{Y}Q^{-1}LQ = 0. \tag{4.2}$$

Set

$$R = B\hat{Y} - \hat{Y}Q^{-1}LQ. \tag{4.3}$$

Then it follows from (4.2) that

$$\|R\| \le \frac{\epsilon}{\sqrt{1-\epsilon^2}}\|W^{\mathrm{H}}AW_\perp\| \le \frac{\epsilon}{\sqrt{1-\epsilon^2}}\|A\|. \tag{4.4}$$

If we now define

$$E = R\hat{Y}^{\mathrm{H}},$$

then it is easy to verify that E satisfies (4.1) and $(B+E)\hat{Y} = \hat{Y}Q^{-1}LQ$. ■

If we now apply Elsner's theorem, we get the following corollary.

Corollary 4.2. *Let the eigenvalues of B be $\nu_1, \ldots, \nu_p$. Then there are integers $j_1, \ldots, j_\ell$ such that*

$$|\lambda_i - \nu_{j_i}| \le p(2\|A\| + \|E\|)^{1-\frac{1}{p}}\|E\|^{\frac{1}{p}}, \qquad i = 1, \ldots, \ell. \tag{4.5}$$

The right-hand side of (4.5) depends only on $\|A\|$ and ϵ. Hence we may conclude that as $\epsilon \to 0$ there are always Ritz values that converges to the eigenvalues of L. The exponent $\frac{1}{p}$ in (4.5) means that in the worst case the convergence of the Ritz values can be slow. Unfortunately, if the eigenvalues of L are defective, we will indeed observe slow convergence. And even if they are well conditioned, without additional conditions convergence can still be slow. One such condition will emerge in the next section.

5. The convergence of Ritz pairs

Having determined that there are Ritz values that converge to the eigenvalues of our distinguished eigenspace $\mathcal{X}$, we now turn to the convergence of the Ritz pairs. The chief difficulty in establishing convergence is that an eigenspace can have any number of eigenpairs even when the eigenpairs are required to be orthonormal. For if (L, X) is an orthonormal eigenpair of A and U is unitary, then $(U^{\mathrm{H}}LU, XU)$ is also an orthonormal eigenpair corresponding to the same eigenspace. If we are to speak of convergence, therefore, we must find a way of removing the ambiguity in the pairs.

One way is to prove convergence of the Ritz spaces directly, after which we can choose converging bases for the spaces, whose associated Rayleigh quotients will naturally converge. The problem with this approach is that the convergence conditions must be phrased in terms of the eigenblock L, which is unknown before convergence. For this reason, we will take a less direct approach.

We will use the notation and results of Theorem 4.1 and Corollary 4.2. Let (N, Z) be the eigenpair associated with the eigenvalues of B that converge to those of L. Let $(Z\ \ Z_\perp)$ be unitary. From the relation $BZ = ZN$ it follows that

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

Now let E be such that $(Q^{-1}LQ, \hat{Y})$ is an eigenblock of $B + E$. The following theorem, which shows that under appropriate conditions $B + E$ has an eigenpair near (N, Z), is an immediate consequence of Theorem V.2.7 in [13].

Theorem 5.1. *Let*

$$\eta = \frac{2\|E\|}{\mathrm{sep}(N, C) - 2\|E\|} < 1 \tag{5.1}$$

Then there is an orthonormal eigenpair $(\tilde{N}, \tilde{Z})$ and a complementary eigenblock $\tilde{C}$ of $B + E$ such that

$$\tan \angle(\tilde{Z}, Z) \le \eta$$

and

$$\|\tilde{N} - N\|, \|\tilde{C} - C\| \le \frac{(1+\eta)\|E\| + \eta\|B\|}{\sqrt{1-\eta^2}}.$$

The condition (5.1) of Theorem 5.1 is not automatically satisfied. Since the only assumption we have made about $\mathcal{W}$ is that it contains a good approximation to $\mathcal{X}$, the eigenvalues of C can lie almost anywhere within a circle about the origin of radius $\|A\|$. In particular, it could happen that as $\epsilon \to 0$ the matrix C has a rogue eigenvalue that converges to an eigenvalue of L so quickly that $\operatorname{sep}(N, C)$ is always too small for (5.1) to be satisfied. From now on, we will assume that *there is a constant α independent of ϵ such that*

$$\operatorname{sep}(N, C) \geq \alpha > 0. \tag{5.2}$$

We will call this the *uniform separation condition.* It implies the condition (5.1), at least for sufficiently small E. By (2.5) this condition is independent of the orthonormal bases Z and $Z_\perp$ used to define N and C. Note that in principle the uniform separation condition can be monitored computationally by computing $\operatorname{sep}(N, C)$ during the Rayleigh–Ritz procedure.

To see how this theorem implies the convergence of eigenblocks, recall that $B + E$ has the eigenpair $(Q^{-1}LQ, \hat{Y})$. By construction, as $\epsilon \to 0$ the eigenvalues of N converge to those of L, and hence by Elsner's theorem so do the eigenvalues of the eigenblock $\tilde{N}$, which is an increasingly small perturbation of N. But by the the continuity of sep and the assumption that the spectra of N and C are disjoint, the eigenvalues of $\tilde{N}$ are bounded away from those of $\tilde{C}$. Hence the eigenvalues of $\tilde{N}$ are the same as those of $Q^{-1}LQ$; i.e., $\Lambda(\tilde{N}) = \Lambda(L)$. But a simple eigenspace is uniquely determined by its eigenvalues.[1] Hence for some unitary matrix U, $\tilde{Z} = \hat{Y}U$, $\tilde{X} = W\hat{Y}U$, and $\tilde{N} = U^{\mathrm{H}}(Q^{-1}LQ)U$. Since $\|N - \tilde{N}\| \to 0$ and $Q \to I$, we have the following theorem.

Theorem 5.2. *Under the uniform separation condition, there is a unitary matrix U depending on ϵ, such that as $\epsilon \to 0$, the eigenpair $(UNU^{\mathrm{H}}, ZU^{\mathrm{H}})$ approaches $(L, \hat{Y})$. Consequently, by* (3.4) *the Ritz pair $(UNU, \tilde{X}U^{\mathrm{H}})$ approaches the eigenpair (L, X).*

Thus the uniform separation condition implies the convergence of Ritz pairs, up to unitary adjustments. In the sequel we will assume that these adjustments have been made and simply speak of the convergence of $(N, \tilde{X})$ to (L, X).

By combining the error bounds in Theorems 4.1 and 5.1 and the inequality (3.4) we can, after some manipulation, establish the following on the asymptotic rate of convergence of the Ritz pairs.

Corollary 5.3. *If the uniform separation condition holds, then*

$$\sin \angle(X, \tilde{X}), \frac{\|N - L\|}{\|A\|} \leq \left[1 + 2\frac{\|A\|}{\operatorname{sep}(L, C)}\right] \epsilon + O(\epsilon^2). \tag{5.3}$$

Thus the convergence is at a rate proportional to the first power of ϵ. The main factor affecting the convergence constant is the reciprocal of the normalized separation

[1] Although this fact seems obvious, its proof is nontrivial.

$\mathrm{sep}(L, C)/\|A\|$, which by the uniform separation condition is bounded away from zero. The linear bound for the convergence of N to L does not imply eigenvalues of N converge at the same rate; however, their convergence is at worst as the $\frac{1}{\ell}$th power of ϵ, which is better than the rate in Corollary 4.2.

6. Residual analysis

Although the uniform separation condition insures convergence (in the sense of Theorem 5.2) of the Ritz pair $(N, \tilde{X}) = (N, WZ)$ to the eigenpair (L, X), it does not tell us when to stop the iteration. A widely used convergence criterion is to stop when the *residual*

$$R = A\tilde{X} - \tilde{X}N$$

is sufficiently small. Note that this residual is easily computable. For in the course of computing the Rayleigh quotient B, we must compute the matrix $V = AW$. After the pair $(N, \tilde{X})$ has been determined we can compute R in the form $VZ - \tilde{X}N$.

We will be interested in the relation of the residual to the accuracy of $\tilde{X} = WZ$, as an approximation to the eigenbasis X in the eigenpair (X, L). To do this we will need some additional notation. Let $(X \;\; X_\perp)$ unitary. Then

$$\begin{pmatrix} X^{\mathrm{H}} \\ X_\perp^{\mathrm{H}} \end{pmatrix} A (X \;\; X_\perp) = \begin{pmatrix} L & G \\ 0 & M \end{pmatrix}, \tag{6.1}$$

where the $(2,1)$-element of the right hand side is zero because $X_\perp^{\mathrm{H}} AX = X_\perp^{\mathrm{H}} XL = 0$. The following theorem of Ipsen [3] holds for any approximate eigenpair.

Theorem 6.1. *Let $(\tilde{L}, \tilde{X})$ be an approximate eigenpair of A and let*

$$\rho = \|A\tilde{X} - \tilde{X}\tilde{L}\|.$$

Then

$$\sin \angle(\tilde{X}, X) \le \frac{\rho}{\mathrm{sep}(\tilde{L}, M)}.$$

Proof. From (6.1) we have $X_\perp^{\mathrm{H}} A = M X_\perp^{\mathrm{H}}$. Hence if $R = A\tilde{X} - \tilde{X}\tilde{L}$, we have

$$X_\perp^{\mathrm{H}} R = X_\perp^{\mathrm{H}} A\tilde{X} - X_\perp^{\mathrm{H}} \tilde{X}\tilde{L} = M X_\perp^{\mathrm{H}} \tilde{X} - X_\perp^{\mathrm{H}} \tilde{X}\tilde{L}.$$

It follows that

$$\sin \angle(X, \tilde{X}) = \|X_\perp^{\mathrm{H}} \tilde{X}\| \le \frac{\|R\|}{\mathrm{sep}(\tilde{L}, M)}. \quad \blacksquare$$

In our application $\tilde{X} = WZ$ and $\tilde{L} = N$. Hence the accuracy of the space spanned by WZ as an approximation to the space $\mathcal{X}$ is proportional to the size of the residual and inversely proportional to $\operatorname{sep}(N, M)$. If the uniform separation condition holds, then up to unitary similarities $N \to L$, so that by the continuity of sep, the accuracy is effectively inversely proportional to $\operatorname{sep}(L, M)$. Unlike the bounds in the previous sections, these bounds cannot be computed, since M is unknown. Nonetheless they provide us some insight into the attainable accuracy of Ritz approximations to an eigenspace.

7. Refined Ritz vectors

When $\ell = 1$, so that our concern is with approximating an eigenvector x and its eigenvalue λ, Theorem 6.1 has a suggestive implication. From Theorem 2.1, we know that there is a Ritz pair $(\nu, Wz) = (\nu, \tilde{x})$ such that ν converges to λ. Hence by Theorem 6.1, if the residual $A\tilde{x} - \nu\tilde{x}$ approaches zero, $\tilde{x}$ approaches x, independently of whether the uniform separation condition (5.2) holds.

Unfortunately, if the uniform separation condition fails to hold, we will generally be faced with a cluster of Ritz values and their Ritz vectors, of which at most one (and more likely none) is a reasonable approximation to x. Now Theorem 6.1 does not require that $(\nu, \tilde{x})$ be a Ritz pair — only that ν be sufficiently near λ, and that $\tilde{x}$ have a sufficiently small residual. Since the Ritz value ν 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. Thus we take $\hat{x}$ to be the solution of the problem

$$\begin{array}{ll} \text{minimize} & \|(A - \nu I)\hat{x}\| \\ \text{subject to} & \hat{x} \in \mathcal{W}, \|\hat{x}\| = 1. \end{array}$$

Alternatively, $\hat{x} = Wv$, where v is the right singular vector of $(A - \nu I)W$ corresponding to its smallest singular value. 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 7.1. *If*

$$\operatorname{sep}(\nu, M) \geq \operatorname{sep}(\lambda, M) - |\nu - \lambda| > 0, \tag{7.1}$$

then

$$\sin \angle(x, \hat{x}) \leq \frac{\|A - \nu I\|\epsilon + |\lambda - \nu|}{\sqrt{1 - \epsilon^2}(\operatorname{sep}(\lambda, M) - |\lambda - \nu|)}. \tag{7.2}$$

Proof. Let

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

be the normalized projection of x onto $\mathcal{W}$ [cf. (3.3)] and let

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

Then

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

Hence

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

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

$$\|(A-\nu I)\hat{x}\| \le \frac{\|A-\nu I\|\epsilon + |\lambda-\nu|}{\sqrt{1-\epsilon^2}}. \tag{7.3}$$

Since $\|(A-\nu I)\hat{x}\|$ is a residual norm, (7.2) follows directly from Theorem 6.1 and (2.4). ■

It follows immediately from (7.2) that if $\nu \to \lambda$ as $\epsilon \to 0$ then the refined Ritz vector $\hat{x}$ converges to the eigenvector x. In particular, by Corollary 4.2 this will happen if μ is chosen to be the Ritz value.

As they are defined, Ritz vectors are computationally expensive. 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 requires the computation of the singular value decomposition of $(A-\nu I)W$, 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 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 $O(m^3)$ [6, 8]. 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 [10].[2] Thus in some important cases, the computation of refined Ritz vectors is a viable alternative to the Rayleigh–Ritz procedure for eigenvectors.

There is a natural generalization of refined Ritz vectors to higher dimensions. Specifically, given an approximate eigenblock N we can solve the problem

$$\begin{array}{ll} \text{minimize} & \|(A\hat{X} - \hat{X}N)\| \\ \text{subject to} & \mathcal{R}(\hat{X}) \subset \mathcal{W}, \hat{X}^{\mathrm{H}}\hat{X} = I. \end{array}$$

The resulting basis satisfies a theorem analogous to Theorem 7.1. There are, however, two difficulties with this approach. First, there seems to be no reasonably efficient algorithm for computing refined Ritz bases. Second, for the bound on the refined Ritz basis to converge to zero, we must have $N \to L$. However, the only reasonable hypothesis under which $N \to L$ is the uniform separation condition, and if that condition is satisfied the ordinary Ritz bases also converge.

8. Discussion

We have considered the convergence of Ritz pairs generated from a subspace $\mathcal{W}$ to an eigenpair (L, X) associated with an eigenspace $\mathcal{X}$. The results are cast in terms of the quantity $\epsilon = \sin\angle(\mathcal{X}, \mathcal{W})$. An appealing aspect of the analysis is that we do not need to assume that eigenvalues of A are distinct or that A is nondefective.

The first result shows that as $\epsilon \to 0$, there are eigenvalues of the Rayleigh quotient B that converge to the eigenvalues of L. Unfortunately, that is not sufficient for the convergence of the eigenpairs, which requires the uniform separation condition (5.2) to separate the converging eigenvalues from the remaining eigenvalues of B. This condition, which can be monitored during the Rayleigh–Ritz steps, insures that the Ritz pairs $(N, \tilde{X})$ converge (with unitary adjustment) to the pair (L, X). The asymptotic convergence bounds (5.3) show that the convergence is linear in ϵ.

When $\mathcal{X}$ is one dimensional — that is when we are concerned with approximating an eigenpair (λ, x) — the Ritz blocks, which become scalar Ritz values, converge without the uniform separation condition. However, this condition is required for the convergence of the Ritz vectors. Alternatively, we can compute refined Ritz vectors, whose convergence is guaranteed without the uniform separation condition.

[2] If the singular values of $(A - \nu I)W$ are $\sigma_1 \geq \cdots \geq \sigma_m$ and σ_m is small, then the loss of accuracy comes from the fact that the accuracy of the computed singular vector is around $(\sigma_1/\sigma_{m-1})\epsilon_{\mathrm{M}}$, where ϵ_{M} is the rounding unit. If we pass to the cross-product matrix then the ratio σ_1/σ_{m-1} is replaced by the larger value $(\sigma_1/\sigma_{m-1})^2$. If the ratio is near one, then the squaring will have little effect. But if the ratio is fairly large — as it will be when we are attempting to resolve poorly separated eigenvalues — then considerable accuracy can be lost.

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 $V^{\mathrm{H}}W = I$. We expect that the above results will generalize easily to this case provided the product $\|V\|\|W\|$ remains uniformly bounded.

References

[1] B. N. Datta. *Numerical Linear Algebra and Applications.* Brooks/Cole, Pacific Grove, CA, 1995.

[2] G. H. Golub and C. F. Van Loan. *Matrix Computations.* Johns Hopkins University Press, Baltimore, MD, second edition, 1989.

[3] I. C. F. Ipsen. Absolute and relative perturbation bounds for invariant subspaces of matrices. Technical Report TR97-35, Center for Research in Scientific Computation, Mathematics Department, North Carolina State Unversity, 1998.

[4] Z. Jia. *Some Numerical Methods for Large Unsymmetric Eigenproblems.* PhD thesis, University of Bielefeld, 1994.

[5] Z. Jia. The convergence of generalized Lanczos methods for large unsymmetric eigenproblems. *SIAM Journal on Matrix Analysis and Applications*, 16:843–862, 1995.

[6] Z. Jia. Refined iterative algorithms based on Arnoldi's process for large unsymmetric eigenproblems. *Linear Algebra and Its Applications*, 259:1–23, 1997.

[7] Z. Jia. Generalized block Lanczos methods for large unsymmetric eigenproblems. *Numerische Mathematik*, 80: 539–566, 1998.

[8] Z. Jia. A refined algorithm based on the block Arnoldi algorithm for large unsymmetric eigenproblems. *Linear Algebra and Its Applications*, 270:171–189, 1998.

[9] Z. Jia. Polynomial characterizations of the approximate eigenvectors by the refined Arnoldi method and an implicitly restarted refined Arnoldi algorithms. *Linear Algebra and Its Applications*, 287: 191–214, 1999.

[10] Z. Jia. A refined subspace iteration algorithm for large sparse eigenproblems. *Applied Numerical Mathematics*, to appear.

[11] Y. Saad. *Numerical Methods for Large Eigenvalue Problems: Theory and Algorithms.* John Wiley, New York, 1992. Cited in [1].

[12] G. L. G. Sleijpen and H. A. Van der Vorst. A Jacobi–Davidson iteration method for linear eigenvalue problems. *SIAM Journal on Matrix Analysis and Applications*, 17:401–425, 1996.

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