Electronic Transactions on Numerical Analysis.
Volume 7, 1998, pp. 182-189.

ISSN 1068-9613.

ETNA
Kent State University
etna@mcs.kent.edu

ON THE CONVERGENCE OF A NEW RAYLEIGH QUOTIENT METHOD WITH APPLICATIONS TO LARGE EIGENPROBLEMS*

D. P. OLEARY† AND G. W. STEWART‡

Abstract. In this paper we propose a variant of the Rayleigh quotient method to compute an eigenvalue and corresponding eigenvectors of a matrix. It is based on the observation that eigenvectors of a matrix with eigenvalue zero are also singular vectors corresponding to zero singular values. Instead of computing eigenvector approximations by the inverse power method, we take them to be the singular vectors corresponding to the smallest singular value of the shifted matrix. If these singular vectors are computed exactly the method is quadratically convergent. However, exact singular vectors are not required for convergence, and the resulting method combined with Golub–Kahan–Krylov bidiagonalization looks promising for enhancement/refinement methods for large eigenvalue problems.

Key words. Rayleigh quotient method, singular value decomposition, large eigenproblem.

AMS subject classifications. 15A42, 65F15, 65H15.

1. Introduction.

The starting point for the algorithm analyzed in this paper is the following variant of the Rayleigh quotient method. Let A be of order n, and let λ be a simple eigenvalue of A with right and left eigenvectors x and y^{H}. Let $\tilde{v}$ and $\tilde{w}^{\mathrm{H}}$ be approximations to x and y^{H}, and let τ be an approximation to λ. Then new approximations $\hat{v}$, $\hat{w}^{\mathrm{H}}$, and $\hat{\tau}$ are generated as follows:

$$\begin{array}{ll} 1. & \hat{v} = (A - \tau I)^{-1}\tilde{v} \\ 2. & \hat{w}^{\mathrm{H}} = \tilde{w}^{\mathrm{H}}(A - \tau I)^{-1} \\ 3. & \hat{\tau} = \hat{w}^{\mathrm{H}}A\hat{v}/\hat{w}^{\mathrm{H}}\hat{v}. \end{array} \tag{1.1}$$

This procedure can, of course be iterated. The quantity $\hat{\tau}$ is called the generalized Rayleigh quotient of A at $\hat{v}$ and $\hat{w}^{\mathrm{H}}$. Ostrowski [6] showed that under weak conditions on $\hat{v}$ and $\hat{w}^{\mathrm{H}}$ the shift τ converges cubically to λ provided that the initial shift is sufficiently near λ. There are two reasons for the fast convergence. First, steps 1 and 2 in (1.1) improve earlier approximations to the right and left eigenvectors. Second, this improvement is magnified by the generalized Rayleigh quotient, which is more accurate than an ordinary Rayleigh quotient formed from a single vector.

In this paper we will be concerned with a variant of this method in which the approximations $\tilde{v}$ and $\tilde{w}^{\mathrm{H}}$ are determined in a different way. We begin by noting that if $\tau = \lambda$ then $A - \tau I$ has a zero singular value, with right and left singular vectors x and y^{H}. Consequently, if τ is near λ, the right and left singular vectors v and w corresponding to the smallest singular value σ of $A - \tau I$ should approximate x and y^{H}. (We will make this statement more precise in Theorem 5.1.) For brevity we will call these singular vectors the inferior singular vectors of $A - \tau I$. In practice, we do not compute the inferior singular vectors exactly but instead approximate them. This suggests the following procedure, which can also be iterated.

$$\begin{array}{ll} 1. & \text{Let } \tilde{v} \text{ and } \tilde{w}^{\mathrm{H}} \text{ be approximations to the right and left} \\ & \text{inferior singular vectors of } A - \tau I, \\ 2. & \hat{\tau} = \tilde{w}^{\mathrm{H}}A\tilde{v}/\tilde{w}^{\mathrm{H}}\tilde{v} \end{array} \tag{1.2}$$

*Received October 12, 1997. Accepted for publication July 6, 1998. Recommended by R. Lehoucq. This work was supported by the National Science Foundation under grant CCR-95-03126.

†Department of Computer Science and Institute for Advanced Computer Studies, University of Maryland, College Park, MD 20742.

‡Department of Computer Science and Institute for Advanced Computer Studies, University of Maryland, College Park, MD 20742, (`stewart@cs.umd.edu`).

ETNA
Kent State University
etna@mcs.kent.edu

Because we do not improve on previous vectors in step one, the scheme is slower than (1.1). But, as we will show, it converges quadratically if the singular vectors are exact, and otherwise it can still be fast. We will call the method the singular vector Rayleigh quotient (SVRQ) method.[1]

At first glance the SVRQ method does not seem to have much to recommend it. It is more difficult to compute singular vectors than to solve linear systems, and consequently a SVRQ step (1.2) requires more work than a step of the original algorithm (1.1). And as we have noted, the new method is slower. Nonetheless, the method may be useful in finding eigenpairs of large matrices.

Specifically, over the past decade new algorithms have been developed to solve large eigenvalue problems by building up approximations to the eigenspaces of eigenvalues lying in a neighborhood of the complex plane. These algorithms (e.g., see [5, 7, 1]) generally begin with subspaces $\mathcal{V}$ and $\mathcal{W}$. The space $\mathcal{V}$ approximates a right eigenspace of A (the space $\mathcal{W}$ usually does not approximate a corresponding left eigenspace). In an enhancement step, the spaces $\mathcal{V}$ and $\mathcal{W}$ are expanded in such a way as to improve the approximations they contain. Since storage considerations limit the dimensions of the spaces, enhancement is followed by a refinement step in which unwanted vectors are purged from the spaces.

The enhancement step generally requires the solution of equations involving $A - \tau I$, where τ is a shift chosen during the refinement step.[2] If A is large, these systems cannot be solved directly, and iterative methods such as GMRES must be employed. Unfortunately, these iterative methods are computationally expensive and consume valuable storage. Moreover, although potentially useful information is generated in the course of the iteration, it is not easy to fold it into the algorithm. Consequently, the information is usually discarded and only the approximate solution is retained.

If we regard steps 1 and 2 in the algorithm (1.1) as enhancement steps, and step 3 as a refinement step (the analogies are not at all far-fetched), then the advantage of the new algorithm (1.2) becomes evident. It is true that (1.2) replaces the iterative solution of a large nonsymmetric system with the iterative determination of inferior singular vectors. But there are effective, well-understood Krylov sequence methods for the singular value decomposition. In the present application the Golub–Kahan–Lanczos (GKL) bidiagonalization method is a natural.[3] This method generates two sequences of orthogonal vectors spanning Krylov subspaces defined by

$$\begin{array}{cccc} \hat{v}, & [(A-\tau I)^{\mathrm{H}}(A-\tau I)]\hat{v}, & [(A-\tau I)^{\mathrm{H}}(A-\tau I)]^2\hat{v}, & \ldots \\ (A-\tau I)\hat{v}, & (A-\tau I)[(A-\tau I)^{\mathrm{H}}(A-\tau I)]\hat{v}, & (A-\tau I)[(A-\tau I)^{\mathrm{H}}(A-\tau I)]^2\hat{v}, & \ldots . \end{array}$$

The vectors in the first sequence contain approximations to the right singular vectors, while the vectors in the second contain approximations to the left singular vectors, which makes them natural candidates to add to $\mathcal{V}$ and $\mathcal{W}$. Moreover, since the singular subspaces also contain approximations to eigenvectors corresponding to eigenvalues near λ (see Theorem 5.1), the refinement step will benefit from the fact that we have approximations to both right and left eigenspaces.

The above observations are speculative, and it will be a major undertaking to bring them to fruition. However, the results will depend on the properties of the SVRQ method (1.2),

[1]In a different context Jia has exploited the connection between singular vectors and eigenvectors with small eigenvalues to generate certain "refined Ritz vectors" [3, 4].

[2]The Jacobi–Davidson method works with a projected version of $A - \tau I$.

[3]We use the appellation Golub–Kahan–Lanczos bidiagonalization to stress the fact that the method is based on Krylov sequences and to distinguish it from the Golub–Kahan reduction to bidiagonal form by orthogonal transformations. Actually both methods are due to Golub and Kahan [2].

and there is no point in proceeding if the method does not perform effectively. In this paper, therefore, we give a convergence analysis of the SVRQ method. To anticipate our results, we will show that if the singular vectors are computed exactly, then the method converges quadratically whenever the initial value of τ is sufficiently near λ and that the size of the convergence region is controlled by the condition numbers of λ and x. If the singular vectors are only approximated, then we give conditions under which convergence rate can be maintained.

This paper is organized as follows. In the next section we introduce a decomposition associated with a simple eigenvalue and establish a result on the accuracy of generalized Rayleigh quotients. In §3 we study the convergence of algorithm (1.2). In the final section we discuss the results and draw conclusions. Implicit in the analysis is a relation between the inferior singular vector of a matrix and an eigenvector corresponding to a small eigenvalue. This relation generalizes to clusters of small singular values, and in an appendix we present the generalization. Throughout this paper $\|\cdot\|$ denotes the Euclidean vector norm and the subordinate spectral matrix norm.

2. Accuracy of generalized Rayleigh quotients. In this subsection we introduce a decomposition associated with a simple eigenvalue and use it to assess the accuracy of the generalized Rayleigh quotient in algorithm (1.2). First the decomposition.

THEOREM 2.1. *Let A be of order n. Let λ be a simple eigenvalue of A with right eigenvector x normalized so that $\|x\| = 1$ and left eigenvector y^{H} normalized so that $y^{\mathrm{H}}x = 1$. Then there are $n \times (n-1)$ matrices X and Y with Y orthonormal such that*

$$\begin{pmatrix} y^{\mathrm{H}} \\ Y^{\mathrm{H}} \end{pmatrix} (x \;\; X) = \begin{pmatrix} 1 & 0 \\ 0 & I \end{pmatrix}$$

and

$$\begin{pmatrix} y^{\mathrm{H}} \\ Y^{\mathrm{H}} \end{pmatrix} A(x \;\; X) = \begin{pmatrix} \lambda & 0 \\ 0 & L \end{pmatrix},$$

where

$$L = Y^{\mathrm{H}}AX = Y^{\mathrm{H}}AY.$$

Moreover

$$\|x\| = \|Y^{\mathrm{H}}\| = 1 \quad \text{and} \quad \|y^{\mathrm{H}}\| = \|X\| \equiv \kappa. \tag{2.1}$$

For a proof see [8]. The theorem states that the eigenvalue λ can be uncoupled from the rest of A by a similarity transformation and that the transformation has certain special properties, which we will use in the sequel. Note that there are block versions of this theorem in which x and y^{H} are replaced by matrices spanning left and right eigenspaces of A (see [9, §V.1]).

The number κ, which is never less than one, will appear as a factor in our bounds, and it is worth while to attach a meaning to it. In fact, κ in (2.1) is a condition number for the eigenvalue λ [9, §IV.2.2]. Specifically, for sufficiently small E there is a unique eigenvalue $\tilde{\lambda}$ of $A + E$ such that

$$\tilde{\lambda} = \lambda + y^{\mathrm{H}}Ex + O(\|E\|^2).$$

It follows on taking norms that

$$|\tilde{\lambda} - \lambda| \le \kappa\|E\| + O(\|E\|^2).$$

In other words κ plays the traditional role of a condition number by bounding the effects on the eigenvalue λ of errors in A.

We now consider the accuracy of the generalized Rayleigh quotient $\tilde{w}^{\mathrm{H}}A\tilde{v}/\tilde{w}^{\mathrm{H}}\tilde{v}$. We begin with the observation that in the notation of Theorem 2.1 any vector $\tilde{v}$ can be expressed in the form $\gamma x + Xg$, where $\gamma = y^{\mathrm{H}}\tilde{v}$ and $g = Y^{\mathrm{H}}\tilde{v}$. Likewise $\tilde{w}^{\mathrm{H}} = \eta y^{\mathrm{H}} + h^{\mathrm{H}}Y^{\mathrm{H}}$, where $\eta = \tilde{w}^{\mathrm{H}}x$ and $h^{\mathrm{H}} = \tilde{w}^{\mathrm{H}}X$. These expansions allow us to state the following theorem.

THEOREM 2.2. *In the notation of Theorem 2.1, let*

$$\tilde{v} = \gamma x + Xg \quad \text{and} \quad \tilde{w}^{\mathrm{H}} = \eta y^{\mathrm{H}} + h^{\mathrm{H}}Y^{\mathrm{H}}.$$

If $\tilde{w}^{\mathrm{H}}\tilde{v} \neq 0$*, then*

$$\frac{\tilde{w}^{\mathrm{H}}A\tilde{v}}{\tilde{w}^{\mathrm{H}}\tilde{v}} = \frac{\gamma\eta\lambda + h^{\mathrm{H}}Lg}{\gamma\eta + h^{\mathrm{H}}g}. \tag{2.2}$$

Moreover, if $1 - \|h\| - \kappa\|g\| > 0$*, then*

$$\left|\frac{\tilde{w}^{\mathrm{H}}A\tilde{v}}{\tilde{w}^{\mathrm{H}}\tilde{v}} - \lambda\right| \leq \frac{2\kappa\|A\|\|h\|\|g\|}{1 - \|h\| - \kappa\|g\|}. \tag{2.3}$$

Proof. The expression (2.2) follows immediately from the relations in Theorem 2.1.

To establish (2.3), use (2.2) to write

$$\frac{\tilde{w}^{\mathrm{H}}A\tilde{v}}{\tilde{w}^{\mathrm{H}}\tilde{v}} - \lambda = \frac{h^{\mathrm{H}}Lg - \lambda h^{\mathrm{H}}g}{\gamma\eta + h^{\mathrm{H}}g}. \tag{2.4}$$

Now an upper bound on the numerator of (2.4) is

$$|h^{\mathrm{H}}Lg - \lambda h^{\mathrm{H}}g| \leq (|\lambda| + \|L\|)\|g\|\|h\| \leq 2\|A\|\|g\|\|h\|, \tag{2.5}$$

the last inequality following from (2.1) and the fact that $L = Y^{\mathrm{H}}AY$.

We must now determine a lower bound on the denominator of (2.4). We begin by determining lower bounds on γ and η. Since $\tilde{v} = \gamma x + Xg$ and $\|\tilde{v}\| = 1$, we must have

$$1 = \tilde{v}^{\mathrm{H}}\tilde{v} = |\gamma|^2 + 2\mathrm{Re}(\bar{\gamma}x^{\mathrm{H}}Xg) + \|Xg\|^2$$

(remember $\|x\| = 1$). But

$$\left||\gamma|^2 + 2\mathrm{Re}(\bar{\gamma}x^{\mathrm{H}}Xg) + \|Xg\|^2\right| \leq (|\gamma| + \|X\|\|g\|)^2 = (\gamma + \kappa\|g\|)^2.$$

Hence we must have

$$|\gamma| \geq 1 - \kappa\|g\|.$$

Proceeding analogously, we find that

$$|\eta| \geq \kappa^{-1}(1 - \|h\|).$$

It now follows that a lower bound for the absolute value of the denominator of (2.4) is

$$|\gamma||\eta| - \|g\|\|h\| \geq \kappa^{-1}(1 - \kappa\|g\|)(1 - \|h\|) - \|g\|\|h\| = \kappa^{-1}(1 - \|h\| - \kappa\|g\|). \tag{2.6}$$

The inequality (2.3) now follows on dividing (2.5) by (2.6). $\square$

3. Convergence of the SVRQ iteration. In this section we will consider the convergence of the SVRQ iteration. A single step of algorithm (1.2) ideally consists of computing the right and left inferior singular vectors v and w^{H} of $A - \tau I$ and then computing the Rayleigh quotient $\hat{\tau} = w^{\mathrm{H}}Av/w^{\mathrm{H}}v$ to give a new shift. In practice, though, we do not compute the singular vectors exactly. Instead we obtain $\tilde{v} = v + \delta_v$ and $\tilde{w}^{\mathrm{H}} = w^{\mathrm{H}} + \delta_w^{\mathrm{H}}$, where v and w are the inferior singular vectors and δ_v and δ_w^{H} are the unknown errors. To study the convergence rate of algorithm (1.2), we study the relation between $|\hat{\tau} - \lambda|$ and $|\tau - \lambda|$. From (2.3) it is seen that the crux of the matter is to derive expressions for the vectors g and h^{H}.

We begin by writing the singular value decomposition of $A - \tau I$ in the form

$$\begin{pmatrix} W^{\mathrm{H}} \\ w^{\mathrm{H}} \end{pmatrix} (A - \tau I)(V \;\; v) = \begin{pmatrix} \Sigma & 0 \\ 0 & \sigma \end{pmatrix}.$$

Here $(V \;\; v)$ and $(W \;\; w)$ are unitary. The quantity σ is the inferior singular value of $A - \tau I$, and v and w^{H} are the right and left inferior singular vectors. Although we do not indicate it explicitly, the components of this decomposition are functions of τ.

We will need a lower bound on the smallest singular value of Σ. Since λ is simple, this singular value is nonzero when $\tau = \lambda$. Hence it is bounded below by a positive constant when τ is restricted to a sufficiently small neighborhood of λ. Thus we can let

$$\mu = \left\{ \begin{array}{l} \text{a positive lower bound for the smallest singular value of} \\ \Sigma \ \text{ in some neighborhood of } \ \lambda. \end{array} \right.$$

We now turn to bounding $g = Y^{\mathrm{H}}\tilde{v}$. We begin by expanding x in terms of the right singular vectors:

$$x = (v^{\mathrm{H}}x)v + VV^{\mathrm{H}}x. \tag{3.1}$$

Multiplying this relation by Y^{H} and using the relation $Y^{\mathrm{H}}x = 0$, we find after a little manipulation that

$$g = Y^{\mathrm{H}}\tilde{v} = Y^{\mathrm{H}}v + Y^{\mathrm{H}}\delta_v = -\frac{Y^{\mathrm{H}}VV^{\mathrm{H}}x}{v^{\mathrm{H}}x} + Y^{\mathrm{H}}\delta_v. \tag{3.2}$$

The next step is to derive an expression for $V^{\mathrm{H}}x$. To do this we first exploit the eigendecomposition of A and then the singular value decomposition, as in Theorem 5.1. Specifically, we have

$$(A - \tau I)x = (\lambda - \tau)x.$$

Multiplying this expression by W^{H} and using the relation $W^{\mathrm{H}}(A - \tau I) = \Sigma V^{\mathrm{H}}$ we get $\Sigma V^{\mathrm{H}}x = (\lambda - \tau)W^{\mathrm{H}}x$ or

$$V^{\mathrm{H}}x = (\lambda - \tau)\Sigma^{-1}W^{\mathrm{H}}x. \tag{3.3}$$

We can now derive a bound on g. Taking norms in (3.3), we get

$$\|V^{\mathrm{H}}x\| \le \frac{\epsilon}{\mu}, \tag{3.4}$$

where we have set

$$\epsilon = |\lambda - \tau|.$$

Since (3.1) is a decomposition of x into orthogonal components and $\|x\| = 1$, it follows that

$$|v^{\mathrm{H}}x| \geq \sqrt{1-(\epsilon/\mu)^2}.$$

Hence from (3.2)

$$\|g\| \leq \frac{\epsilon/\mu}{\sqrt{1-(\epsilon/\mu)^2}} + \|Y^{\mathrm{H}}\delta_v\| \leq \frac{\epsilon/\mu}{\sqrt{1-(\epsilon/\mu)^2}} + \|\delta_v\|.$$

The derivation of a bound for $h^{\mathrm{H}} = w^{\mathrm{H}}X$ is similar, and we only reproduce the result:

$$\|h^{\mathrm{H}}\| \leq \frac{\kappa\epsilon/\mu}{\sqrt{1-(\epsilon/\mu)^2}} + \kappa\|\delta_w^{\mathrm{H}}\|.$$

The additional factor κ comes from the fact that we work with the eigenvector y^{H} and the matrix X, whose norms are κ, instead of working with x and Y, whose norms are one.

If we now substitute these bounds in (2.3) we obtain after some manipulations the following theorem.

THEOREM 3.1. *In the notation of algorithm* (1.2) *and Theorem* 2.1*, if* $\epsilon = |\lambda - \tau|$ *is sufficiently small, there is a a constant* μ *such that*

$$\hat{\epsilon} \equiv |\lambda - \hat{\tau}| \leq C\big((\epsilon/\mu)^2 + (\epsilon/\mu)(\|\delta_v\| + \|\delta_w\|) + \|\delta_v\|\|\delta_w\|\big), \tag{3.5}$$

where

$$C = \frac{2\kappa^2\|A\|}{\sqrt{1-(\epsilon/\mu)^2} - \kappa\big(2(\epsilon/\mu) + \|\delta_v\| + \|\delta_w\|\big)}. \tag{3.6}$$

4. Discussion. In most applications, the quantities $\kappa(\epsilon)/\mu$, $\kappa\|\delta_v\|$, and $\kappa\|\delta_w\|$ will be reasonably small, so that the "constant" C will be essentially $2\kappa^2\|A\|$.

The inequality (3.5) shows that if $\delta_v = \delta_w = 0$ then the iteration is locally quadratically convergent. If δ_v and δ_w are nonzero, we can maintain the local quadratic convergence by computing $\tilde{v}$ and $\tilde{w}$ to an accuracy of $O(\epsilon)$. If we compute $\tilde{v}$ and $\tilde{w}$ to fixed accuracy, then the iteration cannot converge, but the limiting accuracy is the *product* $\|\delta_v\|\|\delta_w\|$. Thus if C and μ are near one, computation of the vectors to an accuracy of 10^{-8} should give eigenvalues of accuracy 10^{-16}.

The bound suggests that we can obtain satisfactory convergence when the error of one of the vectors $\tilde{v}$ or $\tilde{w}$ is actually growing. Suppose, for example we compute $\tilde{v}$ to full accuracy, say 10^{-16}, but compute $\tilde{w}$ by the formula $\tilde{w} = (A - \tau I)\tilde{v}/\|(A-\tau I)\tilde{v}\|$. Initially, $\tilde{w}$ will be reasonably accurate. But as $\tau \to \lambda$, the inferior singular value of $A - \tau I$ will approach zero and $\tilde{w}$ will be computed with increasing cancellation. (In fact, if $\|A\| = 1$, the relative accuracy of $\tilde{w}$ will be about $10^{-16}/\sigma_{\min}$, where $\sigma_{\min}$ is the smallest singular value of $A - \tau I$.) However, the bound (3.5) suggests that convergence will continue until $C\|\delta_w\|/\mu \geq 1$. In fact, the following example shows that the convergence in this case can be quite fast.

EXAMPLE 4.1. *A matrix A of standard normal deviates was generated and normalized to one. One of its eigenvalues*

$$\lambda = -0.35815874795571$$

was chosen and the iteration described above was performed from a starting value of $\tau_0 = 1.3\lambda$. *The following table lists the smallest singular value of* $A - \tau_{k-1}I$ *and* $|\lambda_k - \tau_k|$.

k	$\sigma_{\min}$	$\lvert\lambda_k - \tau_k\rvert$
1	8.6864e−02	4.4909e−03
2	3.4432e−03	1.3098e−05
3	1.0013e−05	1.1433e−10
4	8.7395e−11	5.5511e−17

Although the accuracy of $\tilde{w}_k$ *deteriorates as* $\tau_k \to \lambda$, *the deterioration does not prevent essentially quadratic convergence until the fourth iteration–after which* τ *approximates* λ *to working accuracy.*

We have established the local superlinear convergence of the SVRQ iteration to a simple eigenvalue, as long as the approximate singular vectors are accurate enough. In this case, the vectors $\tilde{v}$ still converge to x, and we have an upper bound on the sine of the angle between $\tilde{v}$ and x, namely

$$\|g\| \leq \frac{\epsilon/\mu}{\sqrt{1-(\epsilon/\mu)^2}} + \|\delta_v\|.$$

The bound (3.5) depends on κ and μ. We have already seen that κ is the condition number of the eigenvalue λ. The quantity μ is related to the condition of the eigenvectors. For it can be shown that when $\tau = \lambda$

$$\mu^{-1} = \|\Sigma^{-1}\| \leq \|(L - \lambda I)^{-1}\|.$$

The quantity $\|(L - \lambda I)^{-1}\|^{-1}$ is written $\mathrm{sep}(\lambda, L)$, and its reciprocal governs the sensitivity of the eigenvectors corresponding to λ [9, §V.2].

If λ is a nondefective multiple eigenvalue of A, then $A - \lambda I$ has a zero singular value of multiplicity at least two. It this case, Σ must have a zero singular value, and our analysis fails because the required positive lower bound μ does not exist. The common sense of this situation is that perturbations of $A - \lambda I$ may cause the right and left singular vectors to move independently in subspaces of dimension at least two. This raises the possibility of generating orthogonal right and left inferior vectors, for which the Rayleigh quotient does not exist.[4] Fortunately, this problem should not affect our intended application to subspace methods for large eigenvalue problems, provided the subspaces $\mathcal{V}$ and $\mathcal{W}$ mentioned in the introduction are large enough to accommodate the multiplicity of the eigenvalue.

5. Appendix: Singular subspaces and eigenspaces. In the derivation of the bound (3.5) we used the fact that if a simple eigenvalue of a matrix is small then its eigenvector must approximate the inferior singular vector of a matrix. This fact can be generalized to eigenspaces and singular spaces.

THEOREM 5.1. *Let* A *be of order* n. *Let* $X \in \mathbf{C}^{n\times p}$ *have orthonormal columns and satisfy*

$$AX = XE, \tag{5.1}$$

where $E = X^{\mathrm{H}}AX$. *Let* A *have the singular value decomposition*

$$\begin{pmatrix} W_1^{\mathrm{H}} \\ W_2^{\mathrm{H}} \end{pmatrix} A(V_1 \;\; V_2) = \begin{pmatrix} \Sigma_1 & 0 \\ 0 & \Sigma_2 \end{pmatrix},$$

[4]Except for the case of Hermitian A, the generalized Rayleigh quotient algorithm (1.1) has an analogous problem.

where Σ_1 is nonsingular of order p and the singular values are in descending order. If we denote by $\Theta(V_1, X)$ the diagonal matrix of canonical angles between the column spaces of V_2 and X, then

$$\|\sin\Theta(V_2, X)\| \leq \frac{\|E\|}{\sigma_p}. \tag{5.2}$$

Proof. The sines of the canonical angles between X and V_2 are the singular values of $V_1^{\mathrm{H}}X$ (see [9, §I.5.2]). Multiplying (5.1) by W_1^{H} and using the fact that $W_1^{\mathrm{H}}A = \Sigma_1 V_1^{\mathrm{H}}$, we find that

$$W_1^{\mathrm{H}}XE = W_1^{\mathrm{H}}AX = \Sigma_1 V_1^{\mathrm{H}}X.$$

The inequality (5.2) now follows on multiplying by Σ_1^{-1} and taking norms. □

A related result holds in which the spectral norm in (5.2) is replaced by the Frobenius norm. In plain words, the theorem says that if an invariant subspace of A has a small spectrum and the rest of the spectrum is well behaved in the sense that σ_p is larger than $\|E\|$, then as E approaches zero the invariant subspace and the corresponding singular subspace approach one another.

REFERENCES

[1] D. R. FOKKEMA, G. L. G. SLEIJPEN, AND H. A. V. DER VORST, *Jacobi–Davidson style QR and QZ algorithms for the reduction of matrix pencils*, Preprint 941, Department of Mathematics, Universiteit Utrecht, 1996.

[2] G. H. GOLUB AND W. KAHAN, *Calculating the singular values and pseudo-inverse of a matrix*, SIAM J. Numer. Anal., 2 (1965), pp. 205–224.

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

[4] ———, *A refined iterative algorithm based on the block Arnoldi process for large unsymmetric eigenproblems*, Linear Algebra Appl., 270 (1997), pp. 171–189.

[5] R. B. MORGAN, *On restarting the Arnoldi method for large nonsymmetric eigenvalue problems*, Math. Comp., 65 (1996), pp. 1213–1230.

[6] A. M. OSTROWSKI, *On the convergence of the Rayleigh quotient iteration for the computation of the characteristic roots and vectors. III (generalized Rayleigh quotient and characteristic roots with linear elementary divisors)*, Arch. Rational Mech. Anal., 3 (1959), pp. 325–240.

[7] D. C. SORENSEN, *Implicit application of polynomial filters in a k-step Arnoldi method*, SIAM J. Matrix Anal. Appl., 13 (1992), pp. 357–385.

[8] G. W. STEWART, *Computable error bounds for aggregated Markov chains*, J. ACM, 30 (1983), pp. 271–285.

[9] G. W. STEWART AND J.-G. SUN, *Matrix Perturbation Theory*, Academic Press, New York, 1990.