BIT
2002, Vol. 42, No. 1, pp. 159–182

0006-3835/02/4201-0159 $16.00

INEXACT RAYLEIGH QUOTIENT-TYPE METHODS FOR EIGENVALUE COMPUTATIONS *

VALERIA SIMONCINI[1] and LARS ELDÉN[2]

[1] *Dipartimento di Matematica, Università di Bologna, and Istituto di Analisi Numerica del CNR, Pavia, Italy. email: val@dragon.ian.pv.cnr.it*

[2] *Department of Mathematics, Linköping University, SE-581 83, Linköping, Sweden. email: laeld@math.liu.se*

Abstract.

We consider the computation of an eigenvalue and corresponding eigenvector of a Hermitian positive definite matrix $A \in \mathbb{C}^{n\times n}$, assuming that good approximations of the wanted eigenpair are already available, as may be the case in applications such as structural mechanics. We analyze efficient implementations of inexact Rayleigh quotient–type methods, which involve the approximate solution of a linear system at each iteration by means of the Conjugate Residuals method. We show that the inexact version of the classical Rayleigh quotient iteration is mathematically equivalent to a Newton approach. New insightful bounds relating the inner and outer recurrences are derived. In particular, we show that even if in the inner iterations the norm of the residual for the linear system decreases very slowly, the eigenvalue residual is reduced substantially. Based on the theoretical results, we examine stopping criteria for the inner iteration. We also discuss and motivate a preconditioning strategy for the inner iteration in order to further accelerate the convergence. Numerical experiments illustrate the analysis.

AMS subject classification: 65F15, 65F10, 65F50.

Key words: Eigenvalue approximation, iterative methods, Newton method, inexact Rayleigh quotient iteration.

1 Introduction.

We consider the problem of computing an eigenvalue and the corresponding eigenvector of a Hermitian positive definite matrix $A \in \mathbb{C}^{n\times n}$. We assume that good approximations of the wanted eigenvalue and eigenvector are already available. A typical application is structural mechanics: the highest or lowest eigenfrequency and corresponding eigenmode need to be recomputed when some material parameters have been changed. We are interested in problems, where the matrix is either so large that it is unfeasible to factorize it (e.g. due to excessive memory requirements for storing the factor), or only available as a subroutine for computing a matrix-vector product.

Standard methods for large and sparse eigenproblems include Lanczos and variants; see e.g. [16, 3]. Recently there has been an increased interest in methods that compute eigenvalue and eigenvector approximations by iteratively solving

*Received January 2000. Communicated by Lothar Reichel.

a sequence of linear systems, often preconditioned. This is a natural approach when the matrix cannot easily be factored and used in a shift-invert method [16]. Examples are the Jacobi–Davidson method [22], Inverse Iteration, and Rayleigh Quotient Iteration [17, 27, 25, 9], and Truncated RQ Iterations [24].

The purpose of the present paper is to discuss in more detail the Rayleigh Quotient iteration for eigenvalue computations when the linear system involved in the recurrence is not solved to high accuracy. We shall refer to this approach as *inexact Rayleigh Quotient–type method*. With a good approximation already available, it is desirable that the convergence of such approach be faster than that obtained by using more powerful but expensive Lanczos approaches [10].

First we describe briefly the Rayleigh Quotient Iteration and a Newton method, formulated on the unit sphere. The latter method is related to the Jacobi–Davidson method [22], and is a special case of Newton's method on the Grassmann manifold presented in [4]. We show that for the case of computing one eigenvalue the Newton method is mathematically equivalent to RQI, both in the case when the linear system is solved exactly, and when an inexact Galerkin–Krylov subspace method is used.

In RQI one needs to solve a linear system shifted by an eigenvalue approximation. Of course, the matrix of this linear system can be extremely ill-conditioned, and one may expect that convergence of an iterative method will be very slow. We analyze this problem, and show that, due to the fact that the right–hand side is close to the eigenvector corresponding to the smallest eigenvalue, *the relevant convergence rate* of inexact RQI is often quite high: We show that during the inner iterations the eigenvalue residual can be reduced substantially while the linear system residual changes only marginally. The important problem of preconditioning the associated linear system is also addressed: a new preconditioning strategy is devised for the inner solver of the RQI method. Our theoretical results concerning the reduction of the eigenvalue residual during the inner iterations are used to derive a stopping criterion for the inner iteration. The theoretical results are illustrated by numerical experiments.

To fix ideas, we shall use (λ_1, y_1) to indicate the sought eigenvalue λ_1 of multiplicity one and eigenvector y_1, regardless of the eigenvalue ordering. Moreover, (θ_i, z_i) will always refer to the current approximation after i Rayleigh Quotient iterations.

Throughout the paper, the 2–norm of a complex vector and the induced norm of a complex matrix are used. Matlab notation is employed; moreover, superscripts are used for iterates of the inner procedure, while subscripts denote iterates of the outer procedure. In order to avoid too much indexing, however, superscripts and subscripts are avoided when the meaning is clear from the context. The vector e_i indicates the ith column of the identity matrix I of given dimension. The following definition of acute angle between two complex vectors x and y (cf. e.g. [16]) is employed whenever angles are involved: $\theta \in [0, \pi/2]$ is the acute angle between x and y if $\cos\theta = |x^*y|/(\|x\|\,\|y\|)$, where x^* denotes the conjugate transpose of x.

2 Rayleigh quotient iteration and a Newton method.

Rayleigh quotient iteration is a classical iterative algorithm; see e.g. [28, 16]. It has a similar structure as the Newton–Grassmann method described in [4]. We first describe the Rayleigh Quotient iteration. Assume that z is an approximation of the wanted eigenvector. The Rayleigh–Ritz approximation of the corresponding eigenvalue is $\theta = z^* A z$, and the new eigenvector approximation w is obtained by solving the (possibly indefinite) linear system of equations

$$(A - \theta I)w = z. \tag{2.1}$$

In Newton's method on the unit sphere [4], the eigenvector approximation z is updated by computing a correction d, satisfying $z^* d = 0$, from the linear system

$$\Pi(A - \theta I)\Pi d = -r, \tag{2.2}$$

where

$$\Pi = I - zz^*, \qquad r = \Pi A z = Az - \theta z.$$

Both methods can be formulated also for the case when a invariant subspace is computed; see [5, 11]. We write the two methods in the pseudo-code below.

Rayleigh Quotient Iteration	Newton–Grassmann Method
Given z_0 with $\|z_0\| = 1$	Given z_0 with $\|z_0\| = 1$
For $i = 0, 1, \dots$	For $i = 0, 1, \dots$
Compute $\theta_i = z_i^* A z_i$	Compute $\theta_i = z_i^* A z_i$
	Set $\Pi = I - z_i z_i^*$ Set $r_i = Az_i - z_i\theta_i$
Solve $(A - \theta_i I)w_i = z_i$	Solve $\Pi(A - \theta_i I)\Pi d_i = -r_i$
Compute $z_{i+1} = w_i / \|w_i\|$	Compute $z_{i+1} = (z_i + d_i)/\|z_i + d_i\|$

The proposed implementation of the Newton–Grassmann method makes the algorithm very similar to classical Newton-type approaches for subspace computation [1]. While in RQI a new *approximation* w is computed from the equation (2.1), in the Newton method one determines a *correction* d by solving the linear equation (2.2). This latter approach is related to deflation techniques [12, 14, 19].

If the matrix dimension is large, an iterative method for solving both systems is preferred. It is shown in [11] that, if z is a close enough approximation to the eigenvector corresponding to one of the extreme eigenvalues, then equation (2.2) is either negative definite (in the case of the largest eigenvalue) or positive definite (smallest eigenvalue). Therefore in such particular case the conjugate gradients (CG) method with preconditioning is ensured to converge. In some cases (2.2) are so well-conditioned that preconditioning is not needed; this occurs e.g. in signal processing contexts [11]. In contrast, the matrix $A - \theta I$ in (2.1) is usually indefinite and very ill-conditioned, in general. Thus, at first sight, it may appear to be much more difficult to solve (2.1) than (2.2) using an iterative method. However, we will show below that the RQ and Newton–Grassmann methods are in fact equivalent in a certain sense.

3 Inexact solution of the inner linear system.

Let w_{RQ} and d_N be the solutions to the linear systems (2.1) and (2.2), respectively. It is straightforward to show that the RQ iteration and the Newton method are equivalent in the sense that the projection $\Pi w_{RQ} = w_{RQ} - z^* w_{RQ} z$, scaled by the factor $1/z^* w_{RQ}$, satisfies (2.2). In other words,

$$w_{RQ} = (z^* w_{RQ})(z + d_N),$$

and, since in both cases we scale the vectors to have length 1, the two procedures produce the same sequence of approximations in exact arithmetic (cf. also [1]). We will now show that the equivalence holds also when (2.2) and (2.1) are solved iteratively using a Galerkin–Krylov subspace method. Part of the proof mimics that of Theorem 4.2 in [25]. We first need the following lemma.

Lemma 3.1 ([2, 25]). *With the notation above,* $\Pi = I - zz^*$ *and* $r_0 = Az - \theta z$,

$$\text{span}\{z, r_0, \Pi A \Pi r_0, \ldots, (\Pi A \Pi)^{m-1} r_0\} = \text{span}\{z, Az, A^2 z, \ldots, A^m z\}.$$

Note that the subspaces in the lemma above are invariant under shift, that is $\text{span}\{z, (A-\theta)z, (A-\theta I)^2 z, \ldots, (A-\theta I)^m z\} = \text{span}\{z, Az, A^2 z, \ldots, A^m z\}$. The latter space is a Krylov subspace, and it is denoted

$$K_{m+1}(A, z) = \text{span}\{z, Az, A^2 z, \ldots, A^m z\}.$$

For the sake of proposition, we recall the definition of a Galerkin–Krylov subspace method. Given a Krylov subspace $K_m(A, z)$, an approximate solution to the linear system $Aw = z$ is constructed as $w^{(m)} \in w^{(0)} + K_m(A, z)$. A Galerkin method is obtained by computing $w^{(m)}$ so that the corresponding residual is orthogonal to the generated Krylov subspace, that is $(z - Aw^{(m)}) \perp K_m(A, z)$ ([18]). For instance, the Conjugate Gradients method belongs to this class of methods.

Proposition 3.2. *Let z be an approximate unit eigenvector of A and let $\theta = z^* Az$. Let $w_{RQ}^{(m+1)}$ and $d_N^{(m)}$ be the approximate solutions to* (2.1) *and* (2.2), *respectively, obtained by applying the same Galerkin–Krylov subspace method to the pair (θ, z). Then there exists $\tau \in \mathbb{C}$ such that*

$$w_{RQ}^{(m+1)} = \tau(z + d_N^{(m)}).$$

Proof. Let $r_0 = Az - \theta z$. The Krylov subspace for the solution to (2.2) is given by $\text{span}\{r_0, \Pi(A-\theta I)\Pi r_0, \ldots, (\Pi(A-\theta I)\Pi)^{m-1} r_0\}$. Let V_m be an orthogonal basis for it and note that $V_m^* z = 0, V_m^* \Pi = V_m^*$. Then $d_N^{(m)} = V_m y_N$ and the Galerkin condition imposes that y_N solves

$$V_m^* \Pi (A - \theta I) \Pi V_m y = -V_m^* r_0,$$

or, equivalently, $V_m^*(A - \theta I)V_m y = -V_m^* Az$ so that

$$y_N = -(V_m^*(A - \theta I)V_m)^{-1} V_m^* Az. \tag{3.1}$$

On the other hand, for Lemma 3.1, let $\tilde{V} = [z, V_m]$ be the orthogonal basis of span$\{z, Az, \ldots, A^m z\}$ generated by the Krylov subspace method. The approximate solution is $w_{RQ}^{(m+1)} = \tau z + V_m t$, with $\tau \in \mathbb{C}$, $t \in \mathbb{C}^m$. The vector $[\tau; t]$ is determined by imposing the Galerkin condition on (2.1):

$$\begin{bmatrix} z^*(A-\theta I)z & z^*(A-\theta I)V_m \\ V_m^*(A-\theta I)z & V_m^*(A-\theta I)V_m \end{bmatrix} \begin{bmatrix} \tau \\ t \end{bmatrix} = e_1.$$

Note that $z^*(A - \theta I)z = 0$. From the second block row we obtain

$$V_m^*(A-\theta I)z\tau + V_m^*(A-\theta I)V_m t = 0,$$

where we have used $V_m^*(A - \theta I)z = V_m^* Az$. Therefore we obtain $t = -(V_m^*(A - \theta I)V_m)^{-1} V_m^* Az\tau$. Combining with (3.1) yields $w_{RQ}^{(m+1)} = \tau(z + V_m y_N)$. □

The equivalence between the use of a Davidson-type eigenvalue solver and the solution of (2.1) by a Galerkin method can be also proved (cf. [25]). However, eigenvalue solvers based on Krylov subspace methods require the explicit storage of the basis vectors, which, in many cases, is undesirable. When using a linear system solver, instead, the approximate solution can be updated at each iteration and only few recurrence vectors need be kept.

Proposition 3.2 shows that solving (2.1) and (2.2) with a Galerkin method when only one eigenvector is sought is mathematically equivalent, as long as the Galerkin solution is defined [17]. In our implementation we opted for the symmetric version of GMRES, the Conjugate Residuals (CR) method, which also minimizes the residual norm at each iteration [18]. We recall that the CR method belongs to the class of Petrov–Galerkin methods, in which the approximate solution $w^{(m)}$ is computed so that the associated residual is orthogonal to $AK_m(A, z)$. In the following sections we shall illustrate the role of the CR solver in the performance of the overall inexact RQ process.

4 On the solution of the linear system $(A - \theta I)w = z$.

Solving the system (2.1) by means of iterative methods may be very slow. Indeed, if (θ, z) is an approximation to an eigenpair of A, then the coefficient matrix $(A - \theta I)$ is almost singular while the right-hand side has a large component along the eigenvector corresponding to the eigenvalue of $(A - \theta I)$ nearest zero. In such a case, if a Krylov subspace method is used as iterative solver, the residual $r^{(m)} = z - (A - \theta I)w^{(m)}$ will not decrease substantially until m is large enough so that the Krylov subspace contains significant spectral information corresponding to the other end of the spectrum. The practical effect is that when the Conjugate Residuals method is used, $\|r^{(m)}\| \approx \|r^{(0)}\|$ possibly until $m = n$, even in exact arithmetic (see Section 7). Nonetheless, our final goal is not that of solving the system, but rather that of obtaining a better approximate eigenvector. The following *eigenvalue* residual norm is what we are interested in during the iteration:

$$\|A\hat{w} - \tau\hat{w}\| \quad \text{where} \quad \hat{w} = w^{(m)}/\|w^{(m)}\|, \quad \tau = \hat{w}^* A\hat{w}, \tag{4.1}$$

where $z_{i+1} := \hat{w}$ at the end of the inner cycle. This residual norm decreases substantially even after very few iterations; this is not surprising, in view of the result in Proposition 3.2, where the equivalence of Galerkin procedures in the Rayleigh Quotient and Newton approaches was proved; see also the discussion in [17] for inverse iteration.

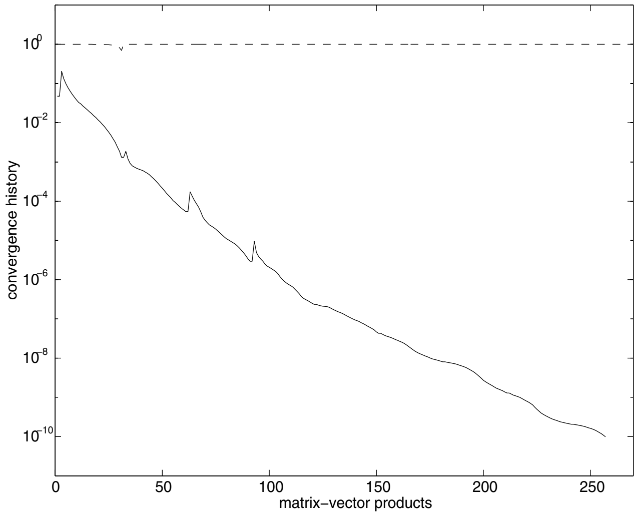

Figure 4.1: Convergence history of inexact RQI using 30 inner iterations per outer cycle. Solid curve: norm of outer eigenvalue residual. Dashed curve: norm of inner system residual.

Example 1. In Figure 4.1 a typical convergence history is depicted. We have considered the problem discussed in [11]. Given the elliptic operator

$$\mathcal{A}(\sigma)u = ((1+\sigma x)u_x)_x + ((1+\sigma y)u_y)_y,$$

the discretization of $\mathcal{A}(\sigma)$ using centered finite differences on the unit square with Dirichlet boundary conditions and 50 nodes in each dimension leads to the matrix $A(\sigma)$ of size 2500. We are interested in approximating the smallest eigenpair of $A(\sigma)$ for $\sigma = 0.15$. Our starting approximation is given by the smallest eigenpair of the matrix $A(0)$, corresponding to the discretization of the Laplacian on the unit square. The four smallest eigenvalues of the two matrices are reported next.

A(0)	0.0076	0.0190	0.0190	0.0303
A(0.15)	0.0081	0.0203	0.0203	0.0326

In order to appreciate the difference between the eigenvalue and system residuals, we considered a fixed number of inner iterations equal to $m = 30$. In Figure 4.1 the solid curve represents the eigenvalue residual norm (4.1) as a function of the number of matrix-vector multiplications. The dashed curve shows the inner system residual behavior throughout the entire RQ process. Clearly, no significant progress is made by the iterative solver in the inner step, in terms of system solution. Nevertheless, the convergence to the wanted eigenpair is more regular. Peaks corresponding to the very first iterations of each inner loop are expected and are not harmful. We shall discuss this issue in Section 7.

In the next theorem we show that a considerable improvement of the approximate eigenvector can be obtained long before the iterative solution of the linear system has converged. We formulate the result for the computation of the smallest (in modulo) eigenvalue and associated eigenvector of a matrix C, which, of course, corresponds to the shifted matrix $A - \theta I$, where $\theta = z^* A z$ is the Rayleigh–Ritz approximation of the eigenvalue.

Theorem 4.1. *Assume that (μ_i, y_i) are the eigenpairs of the Hermitian matrix C, with eigenvalues ordered by increasing modulus $0 < |\mu_1| < |\mu_2| \leq \cdots \leq |\mu_n|$. Further assume that z is a unit norm approximation to y_1. Let $w^{(m)}$ be the Conjugate Residuals approximate solution of $Cw = z$ in $K_m(C, z)$ and let $r^{(m)} = p_m(C)z$ be the associated residual. Let $\psi = \angle(z, y_1)$, $\eta_m = \angle(w^{(m)}, y_1)$. If $p_m(\mu_1) = \varepsilon_m$ with $|\varepsilon_m| < 1$ then*

$$\tan \eta_m \leq \frac{|\mu_1|}{|\mu_2|} \frac{1}{|1 - \varepsilon_m|} \left(1 + \frac{(\|r^{(m)}\|^2 - |\varepsilon_m|^2 \cos^2 \psi)^{1/2}}{\sin \psi} \right) \tan \psi. \tag{4.2}$$

Proof. The polynomial p_m can be written as $p_m(\zeta) = 1 - \zeta q_{m-1}(\zeta)$, with q_{m-1} polynomial of degree at most $m - 1$, so that $w^{(m)} = q_{m-1}(C)z$. Let us write $z = \alpha y_1 + \beta u$, with $u \perp y_1$, $\|u\| = 1$, $|\alpha| = \cos \psi$ and $|\beta| = \sin \psi$. Then $w^{(m)} = q_{m-1}(C)z = q_{m-1}(\mu_1) y_1 \alpha + \|q_{m-1}(C)u\| \hat{u} \beta$, where $\hat{u}$ is the normalized version of $q_{m-1}(C)u$ and $\hat{u} \perp y_1$. Therefore

$$\cos \eta_m = \frac{|q_{m-1}(\mu_1)|}{\|w_m\|} \cos \psi, \qquad \sin \eta_m = \frac{\|q_{m-1}(C)u\|}{\|w_m\|} \sin \psi,$$

from which it follows that

$$\tan \eta_m = \frac{\|q_{m-1}(C)u\|}{|q_{m-1}(\mu_1)|} \tan \psi. \tag{4.3}$$

The condition $p_m(\mu_1) = \varepsilon_m$ corresponds to $q_{m-1}(\mu_1) = (1 - \varepsilon_m)/\mu_1$. Moreover, we have

$$\|q_{m-1}(C)u\| = \|C^{-1}(u - p_m(C)u)\| \leq \frac{1}{|\mu_2|} (1 + \|p_m(C)u\|).$$

Using $\|p_m(C)z\|^2 = \cos^2 \psi |p_m(\mu_1)|^2 + \sin^2 \psi \|p_m(C)u\|^2 \equiv \|r^{(m)}\|^2$ we can write

$$\|p_m(C)u\|^2 = \frac{\|r^{(m)}\|^2 - |\varepsilon_m|^2 \cos^2 \psi}{\sin^2 \psi}.$$

Therefore, collecting all bounds,

$$\frac{\|q_{m-1}(C)u\|}{|q_{m-1}(\mu_1)|} \leq \frac{|\mu_1|}{|\mu_2|} \frac{1}{|1 - \varepsilon_m|} \left(1 + \frac{(\|r_m\|^2 - |\varepsilon_m|^2 \cos^2 \psi)^{1/2}}{\sin \psi} \right).$$

Substituting in (4.3) the final bound follows. □

The theorem shows that it is the size of the linear system residual along the direction of the required eigenvector that matters in the estimate. Since the right–hand side of the linear system has a large component along that eigenvector, the iterative method will reduce the residual along this direction significantly already in the first iterations.

The second term in the parenthesis of the bound looks harmful. However, using the following two bounds (cf. [16, Theorem 11.7.1]) for the eigenvalue λ_1:

$$|\lambda_1 - \theta| \le \frac{\|Az - \theta z\|^2}{\min_{i\neq 1} |\lambda_i - \theta|}, \qquad \|Az - \theta z\| \le \sin\psi \cdot \mathrm{spread}(A),$$

where "spread(A)" is the distance between the extreme eigenvalues of A,

$$\begin{aligned} \frac{|\lambda_1 - \theta|}{\min_{i\neq 1} |\lambda_i - \theta|} \frac{(\|r^{(m)}\|^2 - |\varepsilon_m|^2 \cos^2\psi)^{1/2}}{\sin\psi} &\le \frac{|\lambda_1 - \theta|}{\min_{i\neq 1} |\lambda_i - \theta|} \frac{1}{\sin\psi} \\ &\le \frac{\|Az - \theta z\|}{\min_{i\neq 1} |\lambda_i - \theta|^2}\, \mathrm{spread}(A). \end{aligned}$$

The bound above, in spite of being loose, shows that if the residual of the current eigenpair approximation is small and if the wanted eigenvalue is well separated from the rest of the spectrum, then the iterative solution of the linear system is effective.

The usual bound in the exact inverse iteration only includes the term $|\mu_1|/|\mu_2|$. We next analyze the role of the factor $|1-\varepsilon_m|^{-1}$ in (4.2), which seems to strongly affect the difference between exact and inexact methods. We first note that the residual polynomial plays a role only at μ_1 with the condition $p_m(\mu_1) = \varepsilon_m$, showing that only the behavior of the solver in the neighborhood of the requested eigenvalue is relevant. Since μ_1 is the eigenvalue of $A - \theta I$ closest to zero and the residual polynomial satisfies $p_m(0) = 1$, then in general $p_m(\mu_1) \approx 1$, unless a root of p_m is close to μ_1; we refer to [15] for a deeper analysis of the behavior of residual polynomials at the extreme eigenvalues of a given Hermitian coefficient matrix.

Table 4.1: Example 1: Factors of Theorem 4.1 for the unpreconditioned and preconditioned cases.

system	m	$\|\mu_1\|/\|\mu_2\|$	$\|p_m(\mu_1)\| \equiv \varepsilon_m$	$\frac{\|\mu_1\|}{\|\mu_2\|}\frac{1}{\|1-\varepsilon_m\|}$
$(A-\theta I)w = z$	5	$8.9815 \cdot 10^{-4}$	0.99950	1.7963
	10	$8.9815 \cdot 10^{-4}$	0.99487	0.1751
	20	$8.9815 \cdot 10^{-4}$	0.45662	0.0017
$L^{-*}(A-\theta I)L^{-1}\hat{w} = Lz$	5	$1.0669 \cdot 10^{-3}$	0.87322	0.0084

Fortunately, for the product $|\mu_1|/|\mu_2||1-\varepsilon_m|^{-1}$ to be (much) smaller than one it is sufficient that the ratio $|\mu_1|/|\mu_2|$ is (much) smaller than $|1-\varepsilon_m|$, which is the case when the current Ritz value is a good approximation to the sought eigenvalue. It thus follows that the distance of $p_m(\mu_1)$ from unity may be more significant at an early stage of the RQ process, when $|\mu_1|/|\mu_2|$ may not be small enough. In Table 4.1 we report the values of $|\mu_1|/|\mu_2|$, $|p_m(\mu_1)|$ and $|\mu_1|/|\mu_2||1-\varepsilon_m|^{-1}$ for the first outer iteration on Example 1, in the unpreconditioned and preconditioned cases (see Section 6 for more details on preconditioning). The results in the table show that in practice $|p_m(\mu_1)|$ need not be very small, since the starting approximate eigenvalue ensures that $|\mu_1| \ll |\mu_2|$.

From our results, we also see that for $m = 5$ the factor $|\mu_1|/|\mu_2||1 - \varepsilon_m|^{-1}$ is larger than one in the unpreconditioned system, yielding a supposedly pessimistic bound. It is worth noticing that for such a value of m the inexact RQI procedure does not converge.

5 Eigenvalue residual bounds.

Theorem 4.1 shows that improvements of the approximate eigenvector can be obtained long before the linear system residual has been reduced substantially. In the following theorem we demonstrate that it is the growth of the norm of the approximate system solution that matters in the reduction of the eigenvalue residual norm.

THEOREM 5.1. *Let θ_i, z_i be the current RQI approximation to (λ_1, y_1). Assume that $m > 1$ iterations of some iterative method have been taken and let $w^{(m)} \neq 0$ be the approximate solution to $(A - \theta_i I)w = z_i$ and $r^{(m)} = z_i - (A - \theta_i I)w^{(m)}$ be the corresponding residual. Also let $z_{i+1} = w^{(m)}/\|w^{(m)}\|$ and $\theta_{i+1} = (z_{i+1})^* A z_{i+1}$. Then*

$$\|Az_{i+1} - \theta_{i+1} z_{i+1}\|^2 = \frac{\|z_i - r^{(m)}\|^2}{\|w^{(m)}\|^2} - (\theta_i - \theta_{i+1})^2 \tag{5.1}$$

$$\leq \frac{\|z_i - r^{(m)}\|^2}{\|w^{(m)}\|^2}. \tag{5.2}$$

PROOF. We have

$$\begin{aligned}\|Az_{i+1} - \theta_{i+1} z_{i+1}\|^2 &= \frac{\|\theta_i w^{(m)} + z_i - r^{(m)} - \theta_{i+1} w^{(m)})\|^2}{\|w^{(m)}\|^2} \\ &= \frac{\|(\theta_i - \theta_{i+1}) w^{(m)} + z_i - r^{(m)}\|^2}{\|w^{(m)}\|^2}.\end{aligned}$$

By explicitly writing $\|(\theta_i - \theta_{i+1})w^{(m)} + z_i - r^{(m)}\|^2$ and noticing that

$$(w^{(m)})^*(z_i - r^{(m)}) = (\theta_{i+1} - \theta_i)\|w^{(m)}\|^2,$$

the result follows. □

Note that the bound (5.2) could be directly obtained as

$$\|Az_{i+1} - \theta_{i+1} z_{i+1}\| \leq \|Az_{i+1} - \theta_i z_{i+1}\| = \frac{\|(A - \theta_i I)w^{(m)}\|}{\|w^{(m)}\|} = \frac{\|z_i - r^{(m)}\|}{\|w^{(m)}\|},$$

where the optimality of the Rayleigh quotient has been used in the first bound (see Parlett [16, p.14]). From (5.1) we see that if θ_i is a reasonably good approximation to the sought eigenvalue, then $\|z_i - r^{(m)}\|^2/\|w^{(m)}\|^2$ is a good estimate of the eigenvalue residual norm. The bound above can be specialized if the chosen inner iterative method is taken into account.

Corollary 5.2. *If the Conjugate Residuals method is used as iterative solver with $w^{(0)} = 0$, then, under the hypotheses of Theorem* 5.1,

$$\|z_i - r^{(m)}\|^2 = 1 - \|r^{(m)}\|^2,$$

so that

$$\|Az_{i+1} - \theta_{i+1} z_{i+1}\| \le \frac{\sqrt{1 - \|r^{(m)}\|^2}}{\|w^{(m)}\|}, \quad \textit{for} \quad m > 1. \tag{5.3}$$

Proof. We have $\|z_i - r^{(m)}\|^2 = \|z_i\|^2 - 2z_i^* r^{(m)} + \|r^{(m)}\|^2$. The residual in the CR method satisfies the relation $z_i^* r^{(m)} = \|r^{(m)}\|^2$ (see for instance [21, Proposition 4.1]) from which the result follows. □

The coefficient matrix $(A - \theta_i I)$ is close to singular, and z_i has large component onto the corresponding eigenvector. This implies that the *exact* system solution w has large norm, since $\|w\| \approx \|(A - \theta_i I)^{-1}\|$. The norm of the approximate solution $w^{(m)}$ grows until, for m large enough, it reaches the magnitude of $\|w\|$. In practice, such magnitude is achieved after very few inner iterations; in Section 7 we shall suggest a stopping criterion based on this.

In the following, we give lower bounds for $\|w^{(m)}\|$ which can be used to bound the eigenvalue residual norm. It is known (see, e.g., [16, Theorem 4.8.1]) that if the exact RQI is used, the residual norm $\|Az_i - \theta_i z_i\|$ decreases monotonically and linearly. When using the inexact RQI, we would like to maintain linear convergence in the residual norm. Since $\|r^{(m)}\| \le 1$ for all $m \ge 1$, linear convergence is attained as soon as $\|w^{(m)}\| \ge \|Az_i - \theta_i z_i\|^{-1}$. We next show that, if the approximate eigenvector z_i is not too far from the wanted eigenvector, then linear convergence in the eigenvalue residual norm can be achieved after very few steps of the Conjugate Residuals solver, that is for $|1 - \varepsilon_m| \approx 1$, where ε_m is as in Theorem 4.1.

Note that the following result is in agreement with the eigenvector bound in Theorem 4.1.

Proposition 5.3. *Assume that $m > 1$ iterations of the Conjugate Residuals method have been taken for solving $(A - \theta_i I)w = z_i$ with starting approximation $w^{(0)} = 0$. Let p_m be the associated residual polynomial and let $\varepsilon_m = p_m(\lambda_1 - \theta_i)$. Write the right-hand side*

$$z_i = \alpha_i y_1 + \beta_i u, \tag{5.4}$$

where $y_1^* u = 0$, and $|\alpha_i|^2 + |\beta_i|^2 = 1$. Then

$$\|w^{(m)}\| \ge \frac{|1 - \varepsilon_m||\alpha_i|^3}{|\beta_i|} \frac{1}{\|Az_i - \theta_i z_i\|},$$

and

$$\|Az_{i+1} - \theta_{i+1} z_{i+1}\| \le \frac{|\beta_i|}{|\alpha_i|^3} \frac{\sqrt{1 - \|r^{(m)}\|^2}}{|1 - \varepsilon_m|} \|Az_i - \theta_i z_i\|. \tag{5.5}$$

PROOF. Using the same technique and notation as in Theorem 4.1, and writing $w^{(m)} = q_{m-1}(\lambda_1 - \theta_i)y_1\alpha_i + \|q_{m-1}(A - \theta_i I)u\|\hat{u}\beta$, we first observe that

$$\|w^{(m)}\| \geq \frac{|1 - \varepsilon_m|}{|\lambda_1 - \theta_i|}|\alpha_i|, \tag{5.6}$$

where λ_1 is the eigenvalue approximated by θ_i. Using (5.4) we get $y_1^*(Az_i - \theta_i z_i) = (\lambda_1 - \theta_i)\alpha_i$. Then, since by the definition of θ_i we have $(Az_i - \theta_i z_i)^* z_i = 0$,

$$0 = (Az_i - \theta_i z_i)^* z_i = \alpha_i(Az_i - \theta_i z_i)^* y_1 + \beta(Az_i - \theta_i z_i)^* u,$$

which, by the Cauchy–Schwartz inequality gives the estimate

$$|(Az_i - \theta_i z_i)^* y_1| \leq \frac{|\beta_i|}{|\alpha_i|}\|Az_i - \theta_i z_i\|.$$

Therefore,

$$|(\lambda_1 - \theta_i)| \leq \frac{|\beta_i|}{|\alpha_i|^2}\|Az_i - \theta_i z_i\|.$$

Noticing that $|\alpha_i|^2 + |\beta_i|^2 = 1$, and using (5.6) we obtain the lower bound for $\|w^{(m)}\|$. Substituting that into (5.3) the eigenvalue residual estimate follows. □

Table 5.1: Values of factor $|\beta_i|/|\alpha_i|^3$ during a run of inexact RQI.

ith outer iteration	$\lvert\alpha_i\rvert$	$\lvert\beta_i\rvert/\lvert\alpha_i\rvert^3$
0	0.99955842175	2.97e-02
1	0.99999998728	1.59e-04
2	1.00000000000	5.16e-08*

* In exact arithmetic this value should be zero.

If exact RQI were used, then the bound

$$\|Az_{i+1} - \theta_{i+1} z_{i+1}\| \leq \|Az_i - \theta_i z_i\| \tag{5.7}$$

would hold [16]. In the inexact case, if m is large enough, then

$$\xi_i := \frac{|\beta_i|}{|\alpha_i|^3}\frac{\sqrt{1 - \|r^{(m)}\|^2}}{|1 - \varepsilon_m|} \approx \frac{|\beta_i|}{|\alpha_i|^3} \tag{5.8}$$

and Proposition 5.3 gives a new bound for the eigenvalue residual norm, which is sharper than (5.7) when $|\beta_i|/|\alpha_i|^3 < 1$. This indeed happens whenever the approximate eigenvector is a good approximation to the sought eigenvector, in which case $|\alpha_i|$ is close to unity. To support our argument, we consider the problem in Example 1. In Table 5.1 we report the values of the factor $|\beta_i|/|\alpha_i|^3$ for $i \geq 0$. Since $\alpha_i \approx 1$, in this case $|\beta_i|/|\alpha_i|^3 \approx |\beta_i|$. In order to appreciate the eigenvalue residual bound in Proposition 5.3, in Figure 5.1 we report the values of the factor ξ_i in (5.8) during the whole inexact RQI process, corresponding to 3 outer iterations

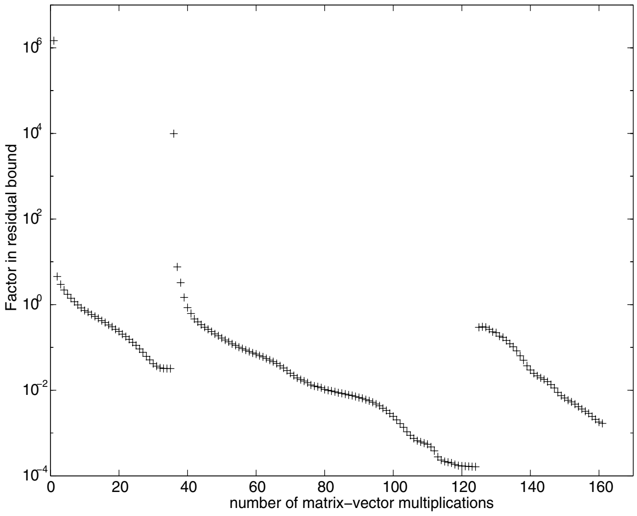

Figure 5.1: Values of the factor ξ_i in (5.8) during a run of inexact RQI.

and 35, 89 and 37 inner iterations, respectively. Note that after very few inner iterations of each outer cycle the factor goes well below 1. The inner stopping criterion used will be discussed in Section 7; here we only notice that the inner tolerance was set to $\epsilon_{inner} = 0.01$.

Our numerical results emphasize the relevant fact that the bound on the eigenvalue residual norm ratio depends on the goodness of the current approximation. More precisely, the factor ξ_i may be much less than one even though the term $\sqrt{1 - \|r^{(m)}\|^2}/|1 - \varepsilon_m|$ is not very small, whenever the approximate eigenvector is close to the exact one. We shall return on this issue in Section 7.

We end this section with the following theorem, which shows that if one were to use classical arguments (as for the proof of (5.3) in [16, Theorem 4.8.1]), then, in place of (5.5), a very pessimistic bound for the eigenvalue residual would be obtained, which is not encountered in practice.

Theorem 5.4. *If the Conjugate Residuals method is used as iterative solver with $w^{(0)} = 0$, then, under the hypotheses of Theorem 5.1,*

$$\|Az_{i+1} - \theta_{i+1} z_{i+1}\| \le \frac{1}{\sqrt{1 - \|r^{(m)}\|^2}} \|Az_i - \theta_i z_i\|. \tag{5.9}$$

Proof. The proof is postponed to Appendix B. □

6 Preconditioning in the inexact RQ iteration.

Preconditioning for the indefinite system (2.1) in RQI need be carefully implemented to be effective. Let $(A - \theta I)w = z$ be the system to be solved, where (θ, z) is an approximate eigenpair of A and consider the incomplete factorization

$P = L^*L$ of A. This yields

$$L^{-*}(A - \theta I)L^{-1}\hat{w} = L^{-*}z \quad \text{and} \quad w = L^{-1}\hat{w}. \tag{6.1}$$

In general, $L^{-*}z$ is no longer an approximate eigenvector of the coefficient matrix $L^{-*}(A - \theta I)L^{-1}$, and therefore, the convenient convergence result of Theorem 4.1 does not apply and convergence can in fact be very slow. We next propose an alternative strategy. Let (δ, x) be an eigenpair of the preconditioned matrix with $\|x\| = 1$, that is

$$L^{-*}(A - \theta I)L^{-1}x = \delta x.$$

Equivalently, we can write

$$(A - \theta I)u = \delta L^* L u \quad \text{with} \quad Lu = x.$$

Let $A = L^*L + E$, where E is the symmetric error matrix associated with the incomplete decomposition of A. Then writing $(A - \theta I)u = \delta(A - E)u$, for $\delta \neq 1$ we obtain

$$Au = \frac{\theta}{1-\delta}u - \frac{\delta}{1-\delta}Eu \tag{6.2}$$

which shows that u is an approximate eigenvector of A with associated eigenvalue $\theta(1-\delta)^{-1}$ and residual norm $\|\delta(1-\delta)^{-1}Eu\|$. The magnitude of the residual norm depends both on $\|E\|$ and on $|\delta|$. In particular, an eigenvalue δ with $|\delta| \ll 1$ appears when $A - \theta I$ is close to singular,[1] that is when θ is a good approximation to an eigenvalue of A. Using a reverse argument, given the approximation z to an eigenvector of A, then the vector Lz approximates an eigenvector x of the preconditioned matrix $L^{-*}(A - \theta I)L^{-1}$ corresponding to the eigenvalue δ closest to zero. In order to justify this reverse argument, we need to show that z is close to $u = L^{-1}x$, where x is the eigenvector mentioned above. This is done in the following proposition.

PROPOSITION 6.1. *Let (θ, z) be an approximation to the eigenpair (λ_1, y_1) of A so that $|\lambda_1 - \theta| \ll |\lambda_i - \theta|$ for $i \neq 1$. Let L^*L be an incomplete Cholesky factorization of A with $A = L^*L + E$, and let (δ, x) be the eigenpair of $L^{-*}(A - \theta I)L^{-1}$ with δ closest to zero. Finally, let $u = L^{-1}x$. Then*

$$\sin\angle(z, u) \leq \sin\angle(u, y_1) + \sin\angle(z, y_1).$$

If, for all $i \neq 1$, we have $\delta \neq (\lambda_i - \theta)/\lambda_i$, then we have the estimate

$$\sin^2\angle(u, y_1) \leq |\delta|^2\|E\|^2 \sum_{i \neq 1} \frac{1}{|\lambda_i - \theta - \lambda_i\delta|^2}. \tag{6.3}$$

PROOF. Let $Y = [y_1, \hat{Y}]$ be the eigenvector matrix of A and let $U = [\hat{u}, \hat{U}]$ be the unitary matrix whose first column is $\hat{u} = u/\|u\|$. We have

$$\|u\|^2 = \|Y^*u\|^2 = \sum_{i=1}^n |y_i^* u|^2 = \|u\|^2 \sum_{i=1}^n \cos^2\angle(y_i, u),$$

[1] We assume A well-conditioned, so that L is also well-conditioned.

therefore, $\sin\angle(u, y_1) = \|\hat{Y}^*\hat{u}\|$. Using the same argument as in [26], it follows that $\|\hat{U}^*y_1\| \leq \sin\angle(u, y_1)$.

Let us write $z = \alpha y_1 + \beta y_\perp$, with $y_1 \perp y_\perp$ and $\|y_\perp\| = 1$. Moreover, $|\alpha| = \cos\angle(z, y_1)$ and $|\beta| = \sin\angle(z, y_1)$. Then $\hat{U}^*z = \alpha\hat{U}^*y_1 + \beta\hat{U}^*y_\perp$ so that we can write

$$\begin{aligned}\sin\angle(z, u) = \|\hat{U}^*z\| &\leq |\alpha|\,\|\hat{U}^*y_1\| + |\beta|\,\|\hat{U}^*y_\perp\| \\ &\leq \|\hat{U}^*y_1\| + |\beta| \leq \sin\angle(y_1, u) + \sin\angle(z, y_1).\end{aligned}$$

This proves the first bound.

Using (6.2) we can write $y_i^*Au = \theta(1-\delta)^{-1}y_i^*u - \delta(1-\delta)^{-1}y_i^*Eu$ for $i = 1, \ldots, n$, or, equivalently

$$(\lambda_i(1-\delta) - \theta)y_i^*u = -\delta\, y_i^*Eu.$$

It follows (under the assumption that $\delta \neq (\lambda_i - \theta)/\lambda_i$ for $i \neq 1$) that

$$\cos\angle(y_i, u) \leq \frac{|\delta|}{|\lambda_i(1-\delta) - \theta|}\|E\|, \quad i \neq 1,$$

and therefore,

$$\sin^2\angle(u, y_1) = \sum_{i\neq 1j}\cos^2\angle(y_i, u) \leq \sum_{i\neq 1}\frac{|\delta|^2}{|\lambda_i(1-\delta) - \theta|^2}\|E\|^2.$$

□

We remark that if θ is a good approximation of λ_1, then $|\delta| \ll 1$ and $|\delta| \approx |\lambda_1 - \theta|$; in particular, $|\delta| \ll |\lambda_i - \theta|$ for $i \neq 1$ so that the bound in (6.3) can be expected to be much less than one. Note further that since δ is small and $\lambda_i - \theta$ is large, the assumption $\delta \neq (\lambda_i - \theta)/\lambda_i$ is no restriction.

In light of the result above, we propose to solve the following preconditioned system:

$$L^{-*}(A - \theta I)L^{-1}\hat{w} = Lz, \tag{6.4}$$

where, in view of the proposition above, Lz is an approximation to the eigenvector of the coefficient matrix corresponding to the eigenvalue closest to zero, so that the result of Theorem 4.1 still applies. The new approximate eigenvector of A is recovered as $w = L^{-1}\hat{w}$.

We emphasize that this strategy completely relies on the fact that the system (6.1) only has to provide a better approximate eigenvector, and we are not interested in accurately solving (6.1). We also notice that the same strategy has been applied in the past to enhance the Lanczos process [20, 13]. However, our motivation and the result of Proposition 6.1 appear to be new.

7 Stopping criteria.

Convergence of inexact RQI is detected when the eigenvalue residual norm associated with the current approximate eigenpair (θ_i, z_i) is less than a given tolerance, that is

$$\frac{\|Az_i - \theta_i z_i\|}{|\theta_i|} \leq \epsilon_{outer}. \tag{7.1}$$

It should be noticed that the value of ϵ_{outer} may vary considerably on different problems, depending on the magnitude of θ_i. If small eigenvalues are sought, then the absolute residual $\|Az_i - \theta_i z_i\|$ may be preferred.

A more delicate issue concerns the stopping criterion for the inner solver, whose analysis has recently received considerable attention; see [9, 8, 23, 6] and references therein. A good combination of stopping strategy with an effective outer iteration may save a considerable amount of computation. In a Rayleigh–Ritz procedure, the outer process is extremely effective, therefore it may be advisable to limit the inner recurrence to the minimum effort and exploit the additional information provided by the Ritz step [16]. As a consequence, the performance of the inexact method should be evaluated by taking into account the overall cost and not only the number of outer iterations. In our experiments we shall consider the cost of matrix–vector multiplications, which is commonly the most time consuming computation. See Appendix A for a practical implementation of the inner–outer process.

Inspired by (5.1), we suggest a stopping criterion based on the variation of the magnitude of $\|w^{(m)}\|$. More precisely, we consider the following condition:

$$\text{stop}_w(m) = \frac{|\,\|w^{(m)}\| - \|w^{(m-1)}\|\,|}{\|w^{(m)}\|} < \epsilon_{inner}, \quad m = 1, 2, \ldots .$$

The use of the magnitude variation in the linear system solution is justified by some practical considerations that conveniently interpret the results of Section 5, when a good approximation to the wanted eigenvector is available. Since in such case $|\alpha_i|$ is very close to unity, the lower bound in (5.6) is rather sharp, so that

$$\|w^{(m)}\| \approx \frac{|1-\varepsilon_m|}{|\lambda_1 - \theta_i|}|\alpha_i| \qquad \text{and} \qquad \frac{\|w^{(m-1)}\|}{\|w^{(m)}\|} \approx \frac{|1-\varepsilon_{m-1}|}{|1-\varepsilon_m|}.$$

Monitoring the magnitude variation provides an estimate of how rapidly the CR polynomial varies at $\mu_1 = \lambda_1 - \theta_i$ as m increases. This estimate turns out to be quite accurate; see e.g. the results in Table 7.1, for the first inner iteration with the matrix $\mathcal{A}(\sigma)$ of Example 1.

Table 7.1: Growth of the solution magnitude $\zeta_w = \|w^{(m+1)}\|\,\|w^{(m)}\|^{-1}$ and $\zeta_\varepsilon = |1-\varepsilon_{m+1}|\,|1-\varepsilon_m|^{-1}$.

m	2	3	4	5	6	7	8	9
ζ_w	3.40408	2.23774	1.85797	1.67654	1.57007	1.49937	1.44799	1.36053
ζ_ε	3.40383	2.23764	1.85791	1.67650	1.57004	1.49935	1.44798	1.36052

A small value of $\text{stop}_w(m)$ highlights stagnation. This may be due either to a final (steady) approximation of the polynomial root (so that $|\varepsilon_m| \approx |\varepsilon_{m+1}| \ll 1$), or to a possibly slow approximation process. This latter event may cause a premature termination of the inner process. In light of (5.3), the fulfillment of the additional condition $\|w^{(m)}\| > \|Az_i - \theta_i z_i\|^{-1}$ ensures that a smaller eigenvalue residual norm is obtained. For this reason, in our algorithm we have adopted the following coupled inner stopping test:

$$\text{(7.2)} \qquad \textbf{if } \ \mathrm{stop}_w(m) < \epsilon_{inner} \ \& \ \|w^{(m)}\| > \frac{1}{\|Az_i - \theta_i z_i\|} \quad \textbf{then} \text{ stop.}$$

The inner recursion also stops whenever the outer stopping test (7.1) is satisfied, as it can be naturally monitored within the inner step.

8 Numerical experiments.

In this section we report some experimental results to support the use of the inner stopping criterion in (7.2). We first analyze the performance when using stop_w, then we study the sensitivity of our coupled stopping strategy as the spectrum of the matrix changes.

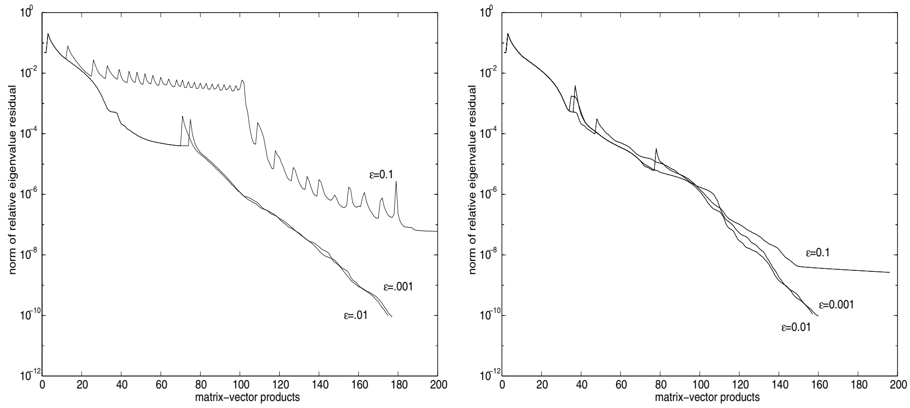

Figure 8.1: Convergence history of inexact RQI using stop_s (left) and stop_w (right) in the inner stopping criterion.

As an alternative to (7.2) one could think of monitoring the variation of the eigenvalue residual norm during the inner iteration; however, we will see that this quantity is less reliable than $\mathrm{stop}_w(m)$. Let stop_s denote the relative variation of the eigenvalue residual norm and consider once more matrix $A(\sigma)$ in Example 1. The plots in Figure 8.1 report the convergence history (eigenvalue residual norm) versus the number of matrix-vector multiplications for various values of ϵ_{inner} when using stop_s (left) and stop_w (right). The unpreconditioned procedure was used. Table 8.1 provides the number of inner iterations per outer iteration.

Analogously, Figure 8.2 and Table 8.2 report on the performance of the method when using the preconditioning strategy developed in Section 6. The preconditioner used was the Matlab `cholinc` function applied on A with fill-in threshold equal to 10^{-2} (see also [5]).

When using stop_w, the method seems to be quite insensitive to the tolerance, as long as the tolerance is not too loose. The performance of the method when using stop_s is instead affected by the tolerance value: see for instance the large improvement given by using $\epsilon_{inner} = 0.2$ as opposed to $\epsilon_{inner} = 0.1$ in the preconditioned case.

Table 8.1: Number of inner iterations when using stop$_s$ (left) and stop$_w$ (right) in the inner stopping criterion; unpreconditioned case.

ith	ϵ_{inner}	ϵ_{inner}	ϵ_{inner}
it.	10^{-1}	10^{-2}	10^{-3}
1	11	69	73
2	14	106	104
3	8		
4	7		

ith	ϵ_{inner}	ϵ_{inner}	ϵ_{inner}
it.	10^{-1}	10^{-2}	10^{-3}
1	33	35	46
2	44	89	112
3	74	37	
4	2		

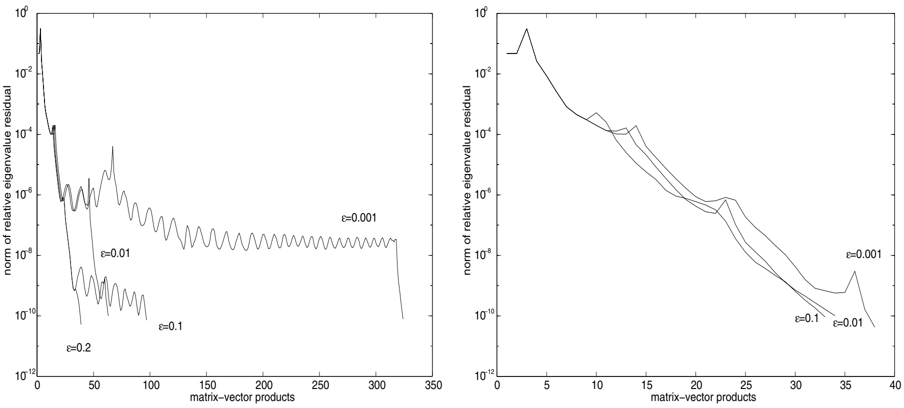

Figure 8.2: Convergence history of inexact RQI with preconditioning, using stop$_s$ (left) and stop$_w$ (right) in the inner stopping criterion.

All figures show periodic peaks in the eigenvalue residual norm. These appear at an early stage of each inner process, during which the solution $w^{(m)}$ does not satisfy the second condition in (7.2). This is expected since in general $w^{(m)}$ does not provide a good eigenvector approximant for very small values (1–3) of m (see Section 5). Already at the third inner iteration, however, the condition $\|w^{(m)}\| > \|Az_i - \theta_i z_i\|^{-1}$ is satisfied in most cases.

Example 2. This is a more contrived example, and it aims at exercising the inner stopping strategy (7.2) in the algorithm. We consider the class of matrices

$$A_\delta = \mathrm{diag}(1, 1+\delta, 3, \ldots, 100) \quad \delta \in (0,1),$$

which only differ for the distance between the smallest and second smallest eigenvalues, 1 and $1+\delta$, respectively. We wish to approximate the eigenpair $(1, e_1)$ and we have considered starting approximate eigenvector z of unit norm with random entries normally distributed. Different selections were studied whose relevant properties are reported in Table 8.3. The starting eigenvalue approximation is given by $zi^* A_\delta zi$; it can be shown that $zi^* A_\delta zi \approx zi^* A_1 zi$ for all $i = 1, 3$ so that only the value of $zi^* A_1 zi$ is reported in the table.

Table 8.2: Number of inner iterations when using stop$_s$ (left) and stop$_w$ (right) in the inner stopping criterion. Preconditioned problem. (*) max # of inner iterations reached.

ith	ϵ_{inner}	ϵ_{inner}	ϵ_{inner}	ϵ_{inner}
it.	$2 \cdot 10^{-1}$	10^{-1}	10^{-2}	10^{-3}
1	12	12	13	14
2	10	10	31	300*
3	13	37	54	10
4	6	6		

ith	ϵ_{inner}	ϵ_{inner}	ϵ_{inner}
it.	10^{-1}	10^{-2}	10^{-3}
1	8	10	12
2	10	11	10
3	16	15	12
4			6

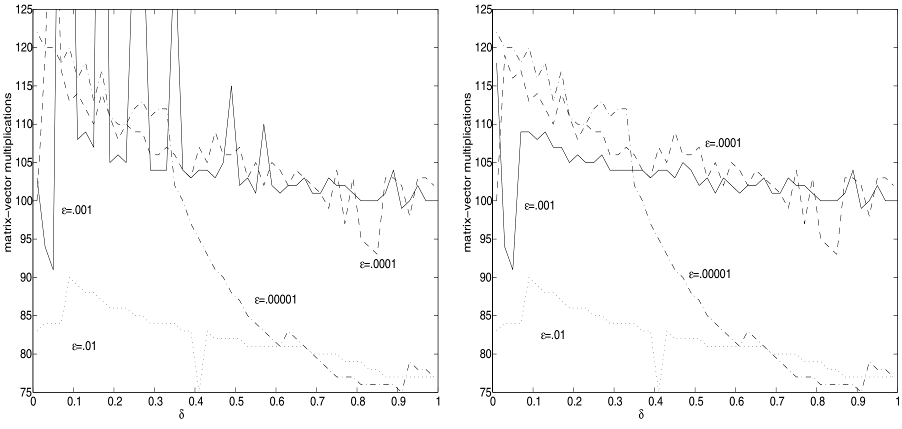

Figure 8.3: Example 2. Number of matrix-vector multiplies as δ varies. Left: Stopping criterion only based on stop$_w$. Right: Coupled stopping criterion (7.2).

Test 1. This experiment shows the effectiveness of the coupled stopping test in (7.2). The outer threshold was set equal to $\epsilon_{outer} = 10^{-10}$ for which the studied behavior was pictorially more pronounced. Figure 8.3 reports the total number of matrix–vector multiplications needed to achieve convergence using $z2$ as starting guess, as δ varies in A_δ. The following different inner tolerance values were used for stop$_w$: $\epsilon_{inner} = 0.01, 0.001, 0.0001, 0.00001$. The left plot reports the convergence history when only using stop$_w(m) < \epsilon_{inner}$. The right plot shows the behavior of the method when adding the condition $\|w^{(m)}\| > \|Az_i - \theta_i z_i\|^{-1}$. The coupled test prevents the inner solver from stopping too early, which causes in these cases more outer iterations to be performed.

This test also shows that for this class of matrices performance is influenced by the inner tolerance ϵ_{inner}: in practice, a lot of computational effort is wasted if the tolerance is too stringent. On the other hand, for the smallest tolerance value, good performance is recovered when the wanted unit eigenvalue is well separated from the eigenvalue $1 + \delta$, that is for $\delta \gg 0$.

Table 8.3: Example 2. Relevant information for different starting approximate eigenvectors.

	$z1$	$z2$	$z3$
$e_1^* z$	0.995	0.9992	0.99996
$z^* A_1 z$	1.4660	1.0804	1.0033

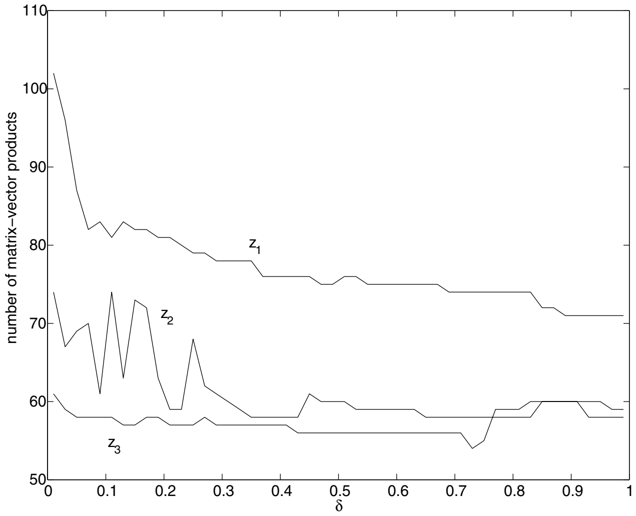

Figure 8.4: Example 2. Number of matrix-vector multiplies as δ varies. Starting approximations $z1, z2$ and $z3$. Threshold values: $\epsilon_{outer} = 10^{-8}$, $\epsilon_{inner} = 10^{-2}$.

Test 2. In this experiment we test the performance of the method as the starting approximation varies from $z1$ to $z3$ with $\epsilon_{inner} = 0.01$. The eigenvalue residual stopping threshold was set equal to $\epsilon_{outer} = 10^{-8}$. The results reported in Figure 8.4 confirm that the performance of the inexact method reflects the closeness of the starting approximate eigenvector. Not surprisingly, the method is also sensitive to the location of the starting approximate eigenvalue (see Table 8.3). Indeed, the total computational cost increases when the distance of the initial Ritz value from the (wanted) unit eigenvalue is not much smaller than the corresponding distance from the second smallest (unwanted) eigenvalue $1 + \delta$.

9 Relations to other stopping criteria.

In this section we show that several inner stopping criteria commonly employed are related to monitoring the quantities we have suggested, whenever the starting eigenvector approximation is close to the sought eigenvector.

Most inner stopping criteria involve the relative residual

$$\frac{\|r^{(m)}\|}{\|r^{(0)}\|} < \epsilon_{inner}, \quad 0 < \epsilon_{inner} < 1 \tag{9.1}$$

for some choice of the starting approximate solution $w^{(0)}$. We recall here that if the Conjugate Residuals method is used, then $\|r^{(m)}\| \le 1$ for all $m > 0$. If $w^{(0)} = 0$, which implies $\|r^{(0)}\| = \|z_i\| = 1$, we have shown that $\|r^{(m)}\|$ is not bound to be small when the inexact RQ iteration is close to convergence, therefore it should only be safely used when the starting approximation is bad and only in the first stage of the RQI process.

The choice $w^{(0)} = z_i\gamma_i$ ([6]), where the real value $\gamma_i > 0$ is the solution norm of the previous inner solution, gives $r^{(0)} = z_i - \gamma_i(A-\theta_i I)z_i$. Since $z_i^*(A-\theta_i I)z_i = 0$, we have

$$\|r^{(0)}\|^2 = 1 + \gamma_i^2\|(A-\theta_i I)z_i\|^2 \ge \gamma_i^2\|(A-\theta_i I)z_i\|^2. \tag{9.2}$$

The last bound becomes quite sharp when $\gamma_i\|(A-\theta_i I)z_i\| \gg 1$, which is usually the case when z_i is a good approximation. For such choice of starting approximate solution $w^{(0)}$, (9.2) yields

$$\frac{\|r^{(m)}\|}{\|r^{(0)}\|} \le \frac{\|r^{(m)}\|}{\gamma_i\|(A-\theta_i I)z_i\|}. \tag{9.3}$$

When close to convergence, the denominator in (9.3) dominates, so that $\|r^{(m)}\|$ need not be small for the criterion (9.1) to be satisfied. The bound (9.3) also shows that it is the magnitude of the solution norm with respect to the eigenvalue residual that plays a crucial role in the determination of the stopping criterion (see also Section 7).

The use of the inner solution magnitude in a stopping criterion is not new. In [17] the authors already proposed such quantity in the inner stopping test for the Conjugate Gradient method, with unfortunately no clear theoretical justification. In [6] the authors also propose the following test for the inexact inverse iteration, which we recall using our notation,

$$\frac{\|r^{(m)}\|}{\|w^{(m)}\|} \le \delta_i \tag{9.4}$$

where i is the number of outer iterations already performed. Adapting the analysis to our notation, in [6] the value of δ_i was suggested to satisfy

$$\delta_i > \left|\frac{\lambda_1 - \theta_{i-1}}{\lambda_2 - \theta_{i-1}}\right|\delta_{i-1}. \tag{9.5}$$

We next give an interpretation of the condition above. If the Conjugate Residuals method is used, then clearly

$$\frac{\|r^{(m)}\|}{\|w^{(m)}\|} \le \frac{1}{\|w^{(m)}\|}$$

and our analysis ensures that the bound will be sharp when the RQI process is close to convergence. In such case, a stopping test on $\|w^{(m)}\|^{-1}$ becomes roughly equivalent to (9.4). The following result shows how the magnitude of $\|w^{(m)}\|$ varies as the RQ approximation takes place.

Proposition 9.1. *With the previous notation and $\psi = \angle(z_i, y_1)$,*

$$\frac{1}{\|w^{(m)}\|} \leq \frac{\text{spread}(A)}{|1-\varepsilon_m|}\tan\psi.$$

Proof. Using [16, Theorem 11.7.1] we have

$$\frac{|\lambda_1-\theta_i|}{|\alpha_i|} \leq \frac{\|Az_i-\theta_i z_i\|}{|\alpha_i|} \leq \frac{|\beta_i|}{|\alpha_i|}\cdot \text{spread}(A).$$

Therefore, from (5.6) we obtain

$$\frac{1}{\|w^{(m)}\|} \leq \frac{|\lambda_1-\theta_i|}{|\alpha_i|}\frac{1}{|1-\varepsilon_m|} \leq \frac{\text{spread}(A)}{|1-\varepsilon_m|}\tan\psi.$$

□

We have shown in Theorem 4.1 that $\tan\psi$ depends on $|\lambda_1-\theta_{i-1}|/|\lambda_2-\theta_{i-1}|$, therefore the condition in (9.4) using (9.5) is justified. We note however that our bound highlights the dependence on another iteration-dependent quantity (namely ε_m), which is expected to influence the performance of the method.

10 Conclusions.

In this paper we have analyzed the behavior of the inexact RQ iteration for computing the eigenpairs of a given large Hermitian matrix when a close approximation is available.

We showed that the inner linear system, arising in the RQI process, can be handled without any special preprocessing with the Conjugate Residuals method. Moreover, such setting allowed us to obtain new insightful relations towards the understanding of the inexact procedure. In particular, we have shown that even though the inner iteration usually does not converge (in the linear system sense), a few iterations with the Conjugate Residuals method produce a good enough approximation for the RQ procedure to be effective. Moreover, we have shown and motivated theoretically how to implement preconditioning for the inner iterations. In light of our theoretical results, we have proposed a new stopping strategy for the inner solver and showed its effectiveness with numerical experiments.

In our future work, we will study the extension of these results to the case when an invariant subspace is sought.

Our analysis focused on the case in which a good approximation is already available. However, our theoretical results can be effectively used to devise stopping strategies in a more general context.

Acknowledgment.

We are indebted to the referees for constructive suggestions.

Appendix A

In this appendix, we write the inexact RQI by explicitly including the CR method in the inner iteration (see e.g. [7]). To simplify the notation, in this section we shall adopt subscription for all inner iterates.

```
Algorithm **Inexact RQI**: Given $A, z$ and a preconditioner $L$
$\theta = z^*Az$, $v_0 = 0$
**while** (not converged)
    $v_1 = (L^*)Lz$,
    $\gamma_0 = \gamma_1 = 1$, $\sigma_0 = \sigma_1 = 0$, $d_0 = d_1 = 0$, $\tilde{d}_0 = \tilde{d}_1 = 0$
    $w_1 = 0$, $\tilde{w}_1 = 0$, $\phi_1 = 1$, $m = 0$
    **while** (not converged **&** $m <$ max_it)
        $m = m + 1$
        $t = L^{-1}((L^*)^{-1}v_m), \quad \tilde{t} = At, \quad \hat{v}_{m+1} = \tilde{t} - \theta t$
        $\alpha = v^*_{m-1}\hat{v}_{m+1}, \quad \hat{v}_{m+1} = \hat{v}_{m+1} - \alpha v_{m-1}$
        $\beta = v^*_m\hat{v}_{m+1}, \quad \hat{v}_{m+1} = \hat{v}_{m+1} - \beta v_m$
        $\chi = \|\hat{v}_{m+1}\|, \quad v_{m+1} = \hat{v}_{m+1}/\chi$
        $\xi_1 = \sigma_{m-1}\alpha, \ \xi_2 = \gamma_m\gamma_{m-1}\alpha + \sigma_m\beta, \ \xi_3 = -\bar{\sigma}_m\gamma_{m-1}\alpha + \gamma_m\beta$
        $[\gamma_{m+1}, \sigma_{m+1}]$ =Givens$(\xi_3, \chi)$
        $\xi_3 = \gamma_{m+1}\xi_3 + \sigma_{m+1}\chi$
        $d_{m+1} = \xi_3^{-1}(t - d_{m-1}\xi_1 - d_m\xi_2) \qquad \tilde{d}_{m+1} = \xi_3^{-1}(\tilde{t} - \tilde{d}_{m-1}\xi_1 - \tilde{d}_m\xi_2)$
        $w_{m+1} = w_m + \gamma_m\phi_m d_{m+1}, \qquad \tilde{w}_{m+1} = \tilde{w}_m + \gamma_m\phi_m\tilde{d}_{m+1}$
        $\phi_{m+1} = -\bar{\sigma}_{m+1}\phi_m$
        Test inner convergence
    **end**
    Update $z$ and $\theta$
**end**
```

Function Givens determines the next Givens rotations (see e.g. [7]). We note that the recurrences for $\tilde{d}, \tilde{w}$ are used to make the matrix vector Aw available, saving a matrix–vector multiply per iteration. Such recurrence overhead becomes precious if one wants to compute the eigenvalue residual within the inner iteration. By setting

$$z = w_{m+1}/\|w_{m+1}\|, \qquad z_A = \tilde{w}_{m+1}/\|w_{m+1}\|, \theta = z^*z_A$$

the outer stopping test (7.1) can be used with $\|\hat{z}_A - \hat{\theta}\hat{z}\|/|\hat{\theta}|$. We also notice that the first matrix-vector multiply of the inner iteration can be obtained for free using the quantity computed during the Ritz step. As a consequence, it can be easily shown that the first CR iteration yields a null approximate solution. The actual approximation takes place starting with the second inner iteration.

Appendix B

In this appendix we report the proof of Theorem 5.4.

Proof. We start by noticing that

$$Az_{i+1} - \theta_i z_{i+1} = \frac{1}{\|w^{(m)}\|}(Aw^{(m)} - \theta_i w^{(m)}) \tag{B.1}$$

and that

$$(A - \theta_i I)w^{(m)} = z_i - r^{(m)}. \tag{B.2}$$

We have

$$\begin{aligned} \|Az_{i+1} - \theta_i z_{i+1}\|^2 &= (Az_{i+1} - \theta_i z_{i+1})^*(Az_{i+1} - \theta_i z_{i+1}) \\ &\overset{(B.1)}{=} \frac{1}{\|w^{(m)}\|}|(Aw^{(m)} - \theta_i w^{(m)})^*(Az_{i+1} - \theta_i z_{i+1})| \\ &\overset{(B.2)}{=} \frac{1}{\|w^{(m)}\|}|(z_i - r^{(m)})^*(Az_{i+1} - \theta_i z_{i+1})|. \end{aligned}$$

The residual in the CR method satisfies the relation $0 = (r^{(m)})^*(A - \theta_i I)w^{(m)} = (r^{(m)})^*(Az_{i+1} - \theta_i z_{i+1})\|w^{(m)}\|$. Therefore,

$$\|Az_{i+1} - \theta_i z_{i+1}\|^2 = \frac{1}{\|w^{(m)}\|}|z_i^*(Az_{i+1} - \theta_i z_{i+1})| \le \frac{1}{\|w^{(m)}\|}\|Az_i - \theta_i z_i\|.$$

From (B.1) we get $\|w^{(m)}\|^{-1} = \|Az_{i+1} - \theta_i z_{i+1}\|/\|(A - \theta_i I)w^{(m)}\|$, and using (B.2) and Corollary 5.2, we can further write $\|(A - \theta_i I)w^{(m)}\| = \|z_i - r^{(m)}\| = \sqrt{1 - \|r^{(m)}\|^2}$, from which it follows that

$$\frac{1}{\|w^{(m)}\|} = \frac{\|Az_{i+1} - \theta_i z_{i+1}\|}{\sqrt{1 - \|r^{(m)}\|^2}}.$$

The result follows from recalling that $\|Az_{i+1} - \theta_{i+1} z_{i+1}\| \le \|Az_{i+1} - \theta_i z_{i+1}\|$. □

REFERENCES

1. F. Chatelin, *Valeurs Propres de Matrices*, Masson, Paris, 1988.
2. J. Cullum and R. A. Willoughby, *Lanczos algorithms for large symmetric eigenvalue computations*, Vol.1, Theory, Vol.2, Program, Birkhäuser, Basel, 1985.
3. J. Dongarra, I. Duff, D. Sorensen, and H. A. van der Vorst, *Numerical Linear Algebra for High-Performance Computers*, SIAM, Philadelphia, PA, 1998.
4. A. Edelman, T. Arias, and S. T. Smith, *The geometry of algorithms with orthogonality constraints*, SIAM J. Matrix Anal. Appl., 20 (1999), pp. 303–353.
5. L. Eldén and V. Simoncini, *Inexact Rayleigh quotient-type methods for subspace tracking*, Tech. Rep. 1172, Istituto di Analisi Numerica del CNR, Dec. 1999.
6. G. H. Golub and Q. Ye, *Inexact inverse iterations for the generalized eigenvalue problems*, BIT, 40 (2000), pp. 671–684.
7. A. Greenbaum, *Iterative methods for solving linear systems*, SIAM, Philadelphia, PA, 1997.

8. Y. Lai, K. Lin, and W. Lin, *An inexact inverse iteration for large sparse eigenvalue problems*, Numer. Linear Algebra Appl., 4 (1997), pp. 425–437.
9. R. Lehoucq and K. Meerbergen, *Using generalized Cayley transformations within an inexact rational Krylov sequence method*, SIAM J. Matrix Anal. Appl., 20 (1999), pp. 131–148.
10. R. B. Lehoucq, D. C. Sorensen, and C. Yang, *ARPACK Users Guide: Solution of Large Scale Eigenvalue Problems by Implicitly Restarted Arnoldi Methods*, SIAM, Philadelphia, PA, 1998.
11. E. Lundström and L. Eldén, *Adaptive eigenvalue computations using Newton's method on the Grassmann manifold*, SIAM J. Matrix Anal. Appl., to appear.
12. L. Mansfield, *On the use of deflation to improve the convergence of conjugate gradient iteration*, Comm. Appl. Numer. Meth., 4 (1988), pp. 151–156.
13. R. B. Morgan and D. S. Scott, *Preconditioning the Lanczos algorithm for sparse symmetric eigenvalue problems*, SIAM J. Sci. Comput., 14 (1993), pp. 585–593.
14. R. A. Nicolaides, *Deflation of conjugate gradients with applications to boundary value problems*, SIAM J. Numer. Anal., 24 (1987), pp. 355–365.
15. C. C. Paige, B. N. Parlett, and H. A. Van der Vorst, *Approximate solutions and eigenvalue bounds from Krylov subspaces*, Numer. Linear Algebra Appl., 2 (1995), pp. 115–134.
16. B. N. Parlett, *The Symmetric Eigenvalue Problem*, SIAM, Philadelphia, PA, 1998.
17. A. Ruhe and T. Wiberg, *The method of conjugate gradients used in inverse iteration*, BIT, 12 (1972), pp. 543–554.
18. Y. Saad, *Iterative Methods for Sparse Linear Systems*, PWS, MA, Boston, 1996.
19. Y. Saad, M. Yeung, J. Erhel, and F. Guyomarc'h, *A deflated version of the conjugate gradient algorithm*, SIAM J. Sci. Comput., 21:5 (2000), pp. 1909–1926.
20. D. S. Scott, *Solving sparse symmetric generalized eigenvalue problems without factorization*, SIAM J. Numer. Anal., 18 (1981), pp. 102–110.
21. V. Simoncini, *On the convergence of restarted Krylov subspace methods*, SIAM J. Matrix Anal. Appl., 22 (2000), pp. 430–452.
22. G. L. G. Sleijpen, A. G. L. Booten, D. R. Fokkema, and H. A. Van der Vorst, *Jacobi–Davidson type methods for generalized eigenproblems and polynomial eigenproblems*, BIT, 36 (1996), pp. 595–633.
23. P. Smit and M. Paardekooper, *The effects of inexact solvers in algorithms for symmetric eigenvalue problems*, Linear Algebra Appl., 287 (1999), pp. 337–357.
24. D. C. Sorensen and C. Yang, *A truncated QR iteration for large scale eigenvalue calculations*, SIAM J. Matrix Anal. Appl., 19 (1998), pp. 1045–1073.
25. A. Stathopoulos and Y. Saad, *Restarting techniques for the (Jacobi–)Davidson symmetric eigenvalue methods*, ETNA, 7 (1998), pp. 163–181.
26. G. W. Stewart, *The convergence of the method of Conjugate Gradients at isolated extreme points of the spectrum*, Numer. Math., 24 (1975), pp. 85–93.
27. D. Szyld, *Criteria for combining inverse iteration and Rayleigh quotient iteration*, SIAM J. Numer. Anal., 25 (1988), pp. 1369–1375.
28. J. H. Wilkinson, *The Algebraic Eigenvalue Problem*, Oxford University Press, Oxford, 1965.