Universiteit Utrecht

Department of Mathematics

Two-sided and Alternating Jacobi–Davidson

by

Michiel E. Hochstenbach and Gerard L. G. Sleijpen

Preprint

nr. 1196

June, 2001

TWO-SIDED AND ALTERNATING JACOBI–DAVIDSON

MICHIEL E. HOCHSTENBACH* AND GERARD L. G. SLEIJPEN*

Abstract. We discuss two variants of a two-sided Jacobi–Davidson method, which have asymptotically cubic convergence for nonnormal matrices and aim to find both right and left eigenvectors. These methods can be seen as Jacobi–Davidson analogs of Ostrowski's two-sided Rayleigh Quotient Iteration. Some relations between (exact and inexact) two-sided Jacobi–Davidson and (exact and inexact) two-sided Rayleigh Quotient Iteration are given, together with convergence rates.

Furthermore, we introduce an alternating Jacobi–Davidson process, that can be seen as the Jacobi–Davidson analog of Parlett's alternating Rayleigh Quotient Iteration. The methods are extended to the generalized and polynomial eigenproblem. Advantages of the methods are illustrated by numerical examples.

Key words. Jacobi–Davidson, Rayleigh Quotient Iteration, Ostrowski's two-sided Rayleigh Quotient Iteration, Parlett's alternating Rayleigh Quotient Iteration, two-sided Lanczos, correction equation, nonnormal matrix, inexact accelerated Newton method, rate of convergence, generalized eigenproblem, polynomial eigenproblem.

AMS subject classifications. 65F15.

1. Introduction. We are interested in the computation of one or more eigenvalues and the corresponding left and right eigenvectors of the (possibly nonnormal) matrix A. It is well-known that when Rayleigh Quotient Iteration (RQI) converges to a simple eigenvalue of a normal matrix, the asymptotic convergence rate is cubic (see for example [5, p. 683], [6, p. 77]). For a nonnormal eigenvalue of a nonnormal matrix, RQI has locally quadratic convergence at best [5, p. 688].

Ostrowski's two-sided RQI [3] works with the two-sided (or generalized) Rayleigh Quotient

$$\theta(u,v) := \frac{v^*Au}{v^*u},$$

where u and v are an approximate right and left eigenvector. It can be shown that when two-sided RQI converges to a simple eigenvalue, the local convergence is cubic (see [5, p. 689] and §2). In (two-sided) RQI, one has to solve linear systems of the form $(A-\theta I)\widetilde{u} = u$. For large sparse matrices, these computations, and therefore (two-sided) RQI as a whole, may be less attractive.

Jacobi–Davidson (JD) [8] is an efficient method to compute a few eigenvalues and corresponding (right) eigenvectors of A. The essence of JD is its correction equation, where the shifted operator $A-\theta I$ is restricted to the subspace orthogonal to the current approximation to an eigenvector. When we solve this equation exactly, then JD can be considered as accelerated RQI (see [8] and §2).

Because of this, it is of interest to investigate Jacobi–Davidson analogs of two-sided RQI, leading to an acceleration of two-sided RQI. The idea of a two-sided Jacobi–Davidson is already, though somewhat hidden, present in [7], in particular Remark 3.5 and Section 5.1.3. We will see that two-sided JD has two search spaces, one for the right and one for the left eigenvector. When the correction equations are solved exactly, the method has locally cubic convergence.

*Mathematical Institute, Utrecht University, P.O. Box 80.010, NL-3508 TA Utrecht, The Netherlands.
`www.math.uu.nl/people/[hochstenbach,sleijpen]`
Email: `[hochstenbach,sleijpen]@math.uu.nl` Revised version, June 1, 2001

In practice, it is neither necessary nor advisable to solve the correction equation in the Jacobi–Davidson method accurately. Instead, we may solve it only approximately, for instance to a certain precision. This principle can also be applied to the two-sided processes, leading to inexact two-sided JD and inexact two-sided RQI. At the price of slower convergence, the methods thus become computationally more attractive. An attempt to merge the two search spaces of two-sided JD gives rise to alternating Jacobi–Davidson, which can be viewed as an acceleration of Parlett's alternating RQI [5].

It is not quite clear yet for which class of matrices the new methods brings an advantage, although our experiments indicate that it might be attractive for moderately nonnormal matrices. The methods are attractive in applications where one wishes to have estimates for the condition number of an eigenvalue during the iteration process (for instance for deciding for further refinement, or for pseudospectra computations).

This paper has been organized as follows. Section 2 introduces some notations and definitions and gives a presentation of Jacobi–Davidson. In Section 3 we review Ostrowski's two-sided RQI, and in Section 4 we consider two flavors of two-sided Jacobi–Davidson. Inexact variants of these two-sided methods, some relations between them, as well as convergence rates, can be found in Section 5. Section 6 proposes alternating Jacobi–Davidson, and Section 7 translates the two-sided methods to the generalized and polynomial eigenvalue problem. In Section 8 we discuss various aspects of the methods. Numerical experiments are presented in Section 9, and a discussion and some conclusions can be found in Section 10.

2. Jacobi–Davidson and Rayleigh Quotient Iteration. Let us first introduce some notations. Throughout the paper, λ denotes a simple eigenvalue of the $n \times n$-matrix A, $n > 1$, with x and y as its normalized right and left eigenvector. The (finite) condition of λ is equal to $\kappa(\lambda) := |y^*x|^{-1}$. Approximations to the eigentriple are indicated by θ for the eigenvalue and u, v for the right and left eigenvector. We assume that θ is not equal to an eigenvalue of A, which is equivalent to the assumption that $A - \theta I$ is invertible. By $\|\cdot\|$ we denote the Euclidean norm. To avoid confusion, we remark that throughout the paper, the word "right" is used as the opposite of "left" (e.g., right eigenvector versus left eigenvector), and does not have the meaning of "correct".

Since $x \not\perp y$, $(A - \lambda I)|_{y^\perp} : y^\perp \to y^\perp$ is invertible; in particular it has a finite condition number, denoted by $\kappa((A - \lambda I)|_{y^\perp})$. Later in the paper, we use the following (related) definition.

Definition 2.1. (Cf. [12, p. 145]) We define the *effective condition number* of a nonzero matrix C as

$$\kappa_e(C) := \|C\|\,\|C^+\| = \sigma_{\max} \,/\, \min_{\sigma_j \neq 0} \sigma_j,$$

where C^+ is the pseudo-inverse of C, and the σs are the singular values of C.

Next, we give a presentation of 'standard' JD, such that two-sided JD will follow as a natural generalization in Section 4. The Jacobi–Davidson method [8] consists of two ingredients. The first part, the well-known *Rayleigh-Ritz* approach, deals with the question: having a k-dimensional search space $\mathcal{U}$ (where one should think of the typical situation $k \ll n$), how do we get an approximate eigenpair (θ, u), where $u \in \mathcal{U}$? Let the columns of U form an orthonormal basis for $\mathcal{U}$, and define the *residual* r by

$$r := Au - \theta u.$$

Imposing the *Ritz-Galerkin condition* on the residual

$$r = Au - \theta u \perp \mathcal{U},$$

and writing $u = Uc$ (where c is a k-dimensional vector), we find that (θ, c) should be a solution of the low dimensional *projected eigenproblem*

$$U^* AUc = \theta c,$$

so a *Ritz pair* $(\theta, u) = (\theta, Uc)$ is a backtransformed eigenpair of the *projected matrix* U^*AU. In particular, if (θ, u) is a Ritz pair, we have

$$\theta = \theta(u) := \frac{u^* Au}{u^* u} \qquad \text{and} \qquad r \perp u,$$

that is, θ is the *Rayleigh Quotient* of u and the corresponding residual is orthogonal to u.

The second ingredient of JD gives an answer to the question: having an approximate eigenpair (θ, u) to (λ, x), how do we expand the search space $\mathcal{U}$ to get an even better approximation? For this, Jacobi–Davidson looks for an orthogonal correction $s \perp u$ such that

$$A(u + s) = \lambda(u + s),$$

i.e., such that $u + s$ is a multiple of the eigenvector x. This equation can be rewritten to obtain

$$(A - \lambda I)s = -(A - \theta I)u + (\lambda - \theta)u = -r + (\lambda - \theta)u.$$

During the process, λ and hence also $(\lambda - \theta)u$ are unknown. Therefore it is interesting to consider the projection of this equation that maps u to 0 and keeps r fixed. Because $r \perp u$, this projection is $I - uu^*$, the orthogonal projection onto the orthogonal complement of u. The result of projecting the previous equation is

$$(I - uu^*)(A - \lambda I)s = -r.$$

In this equation JD replaces λ by θ, this may be considered as 'throwing away second order terms' (both $\lambda - \theta$ and s will be asymptotically small). This suggests that JD is in fact a Newton method, which is true indeed [9]. Using

$$(I - uu^*)s = s,$$

we derive the *JD correction equation*:

$$(I - uu^*)(A - \theta I)(I - uu^*)s = -r \qquad \text{where } s \perp u, \tag{2.1}$$

from which we see that the shifted operator $A - \theta I$ is restricted to the orthogonal complement of u. In practice, (2.1) is often solved only *approximately* (or *inexactly*), for example by an iterative method, e.g., a few steps of (preconditioned) GMRES. The approximate solution is used to expand the search space $\mathcal{U}$, this is called *subspace acceleration*. JD can therefore be viewed as an *accelerated inexact Newton method* for the eigenvalue problem [9].

However, when we solve (2.1) exactly, then we find (see [8])

$$s = -(A - \theta I)^{-1} r + \alpha (A - \theta I)^{-1} u = -u + \alpha (A - \theta I)^{-1} u,$$

where $\alpha = \left(u^* (A-\theta I)^{-1} u\right)^{-1}$ is such that $s \perp u$. JD uses s to expand the search space $\mathcal{U}$. Since already $u \in \mathcal{U}$, we get the same subspace expansion using $\widetilde{s} := (A-\theta I)^{-1} u$. Here we recognize a step of RQI, and we conclude that *exact* JD (i.e. JD where we solve the correction equation exactly) can also be seen as accelerated RQI.

In RQI, when the approximations (θ_k, u_k) converge, they converge asymptotically cubically for normal matrices:

THEOREM 2.2 (Constant of cubic convergence of RQI). *If A is normal and $u_k \to x$ as $k \to \infty$, then*

$$\lim_{k\to\infty} \|u_{k+1} - x\| \,/\, \|u_k - x\|^3 \leq 1.$$

Proof. See [5, p. 683]. □

The underlying reason for the cubic convergence is the following property of the Rayleigh Quotient for normal matrices [5, p. 681]:

$$\theta(u) = \frac{u^* A u}{u^* u} \text{ is stationary } \iff u \text{ is an eigenvector of } A. \tag{2.2}$$

(Recall that stationary means that all directional derivatives are zero.) We have already seen that exact JD can be considered as accelerated RQI. Because JD uses subspace acceleration, it will trivially converge in a finite number of steps. However, in many practical situations an eigenpair will be well approximated long before the dimension k of the search space is of the order of the size of the matrix. When speaking of asymptotic convergence, following [14], we think of this situation. In other words, by the word 'asymptotically' we mean the situation where we have a (very) good approximation to an eigenpair, rather than the situation where $k \to \infty$. This justifies speaking of the asymptotic convergence of JD. When we neglect the effect of the subspace acceleration on the asymptotic convergence, JD 'inherits' the asymptotic convergence of RQI. This explains the statement that "Jacobi–Davidson has asymptotically cubic convergence for normal matrices."

3. Two-sided Rayleigh Quotient Iteration.

If A is nonnormal, property (2.2) is lost for nonnormal eigenvalues. This implies that RQI (and therefore also exact JD) converges asymptotically at best (only) quadratically to a nonnormal eigenpair (λ, x) [5, p. 688]. But instead of (2.2), we have the following property for the *two-sided Rayleigh Quotient* $\theta(u, v)$:

$$\theta(u,v) := \frac{v^* A u}{v^* u} \text{ is stationary } \iff \begin{array}{l} u \text{ and } v \text{ are right and left eigenvector} \\ \text{of } A \text{ with eigenvalue } \theta \text{ and } v^* u \neq 0. \end{array} \tag{3.1}$$

Because of this property, one may expect cubic convergence for simple eigenvalues of nonnormal matrices when we approximate the left and the right eigenvector simultaneously. For this reason Ostrowski proposes a *two-sided Rayleigh Quotient Iteration* [3]. In every step of this method, we solve the two equations

$$(A - \theta_k I) u_{k+1} = u_k \qquad \text{and} \qquad (A - \theta_k I)^* v_{k+1} = v_k. \tag{3.2}$$

This leads to Algorithm 3.1.

In [5, p. 689] it is shown that when this two-sided RQI converges to a simple eigenvalue, it has locally cubic convergence. However, the following theorem states

Pick initial vectors u_1 and v_1 with unit norm, such that $v_1^* u_1 \neq 0$
for $k = 1, 2, \ldots$
1. Compute $\theta_k := \theta_k(u_k, v_k) = \frac{v_k^* A u_k}{v_k^* u_k}$
2. If $A - \theta_k I$ is singular, solve $(A - \theta_k I)x = 0$ and $(A^* - \overline{\theta}_k I)y = 0$, stop
3. Solve $(A - \theta_k I)u_{k+1} = u_k$ and normalize u_{k+1}
4. Solve $(A^* - \overline{\theta}_k I)v_{k+1} = v_k$ and normalize v_{k+1}
5. If $v_{k+1}^* u_{k+1} = 0$ then method fails

Alg. 3.1: Two-sided Rayleigh Quotient Iteration [3].

that the speed of the cubic convergence might be significantly slower in the nonnormal case:

THEOREM 3.1 (Locally cubic convergence of two-sided RQI). *Suppose that u_k and v_k converge to x and y, respectively, as $k \to \infty$. Then $\theta_k \to \lambda$, and we can write*

$$u_k = \alpha_k(x + \delta_k d_k) \qquad \text{and} \qquad v_k = \beta_k(y + \varepsilon_k e_k), \tag{3.3}$$

where $\delta_k, \varepsilon_k \geq 0$, $d_k \perp y$, $e_k \perp x$, and u_k, v_k, x, y, d_k, and e_k all have unit norm. Then

$$\delta_{k+1} \leq \gamma \delta_k^2 \varepsilon_k + \text{h.o.t.} \qquad \text{and} \qquad \varepsilon_{k+1} \leq \gamma \delta_k \varepsilon_k^2 + \text{h.o.t.},$$

where

$$\gamma = \kappa(\lambda)\kappa((A - \lambda I)|_{y^\perp}),$$

and 'h.o.t.' stands for higher order terms in δ_k and ε_k (i.e., in the statement above h.o.t. *stands for terms of order $\mathcal{O}(\delta_k^i \varepsilon_k^j)$, where $i + j > 3$).*

Proof. This is a slight extension on a result in [5, p. 689], where Parlett shows that (in our notation) there exist nonzero $\alpha_{k+1}, \beta_{k+1}$ such that

$$\begin{aligned} u_{k+1} &= \alpha_{k+1}(x + \delta_k(\lambda - \theta_k)(A - \theta_k I)^{-1} d_k), \\ v_{k+1} &= \beta_{k+1}(y + \varepsilon_k(\lambda - \theta_k)^*(A - \theta_k I)^{-*} e_k), \end{aligned}$$

where

$$\theta_k - \lambda = \delta_k \varepsilon_k \frac{e_k^*(A - \lambda I)d_k}{y^* x + \delta_k \varepsilon_k e_k^* d_k}.$$

Hence

$$|\lambda - \theta_k| = \delta_k \varepsilon_k \kappa(\lambda)|e_k^*(A - \lambda I)d_k| + \text{h.o.t.}. \tag{3.4}$$

Since $(A - \lambda I)^{-1}$ exists on $y^\perp$, and $(A - \lambda I)^{-*}$ exists on $x^\perp$, we have for $k \to \infty$

$$\begin{aligned} \|(A - \theta_k I)^{-1} d_k\| &\leq \|\left((A - \lambda I)|_{y^\perp}\right)^{-1}\| + \text{h.o.t.}, \\ \|(A - \theta_k I)^{-*} e_k\| &\leq \|((A - \lambda I)|_{x^\perp})^{-*}\| + \text{h.o.t.}. \end{aligned}$$

We can conclude that

$$\begin{aligned} \delta_{k+1} &\leq \delta_k^2 \varepsilon_k \kappa(\lambda)\kappa((A - \lambda I)|_{y^\perp}) + \text{h.o.t.}, \\ \varepsilon_{k+1} &\leq \delta_k \varepsilon_k^2 \kappa(\lambda)\kappa((A - \lambda I)^*|_{x^\perp}) + \text{h.o.t.}. \end{aligned}$$

The proof is completed by the observation

$$\begin{aligned}\kappa((A-\lambda I)|_{y^\perp}) &= \kappa_e\left(\left(I-\tfrac{xy^*}{y^*x}\right)(A-\lambda I)\left(I-\tfrac{xy^*}{y^*x}\right)\right)\\ &= \kappa_e\left(\left(I-\tfrac{yx^*}{x^*y}\right)(A-\lambda I)^*\left(I-\tfrac{yx^*}{x^*y}\right)\right)\\ &= \kappa((A-\lambda I)^*|_{x^\perp}).\end{aligned}$$

□

Comparing Theorems 2.2 and 3.1, one may observe two differences. First, Theorem 2.2 can also be expressed in the angle $\angle(u_k, x)$ (see [6, Th. 4.7.1]), but in the nonnormal case of Theorem 3.1 this is not obvious. Second, because of the possibly large constant of Theorem 3.1, the cubic convergence may have less significance in practice.

4. Two-sided Jacobi–Davidson. Inspired by two-sided RQI, we design a two-sided Jacobi–Davidson method. We work with two search spaces, $\mathcal{U}$ for the right, and $\mathcal{V}$ for the left eigenvector. Suppose that we have k-dimensional search spaces $\mathcal{U}$ and $\mathcal{V}$, and approximations $u \in \mathcal{U}$ and $v \in \mathcal{V}$ to the right and left eigenvectors, $u \not\perp v$. We now would like to take

$$\theta = \theta(u, v) = \frac{v^*Au}{v^*u} \tag{4.1}$$

as approximation to the eigenvalue. Note that (4.1) holds if and only if $Au - \theta u \perp v$ and $A^*v - \overline{\theta}v \perp u$. This suggests to impose *Petrov-Galerkin conditions* on the *right residual* r_u and *left residual* r_v to determine approximate eigenvectors u and v:

$$r_u := Au - \theta u \perp \mathcal{V} \qquad \text{and} \qquad r_v := A^*v - \overline{\theta}v \perp \mathcal{U}.$$

Now write $u = Uc$ and $v = Vd$, where the columns of U and V form bases for $\mathcal{U}$ and $\mathcal{V}$ (not necessarily orthogonal, see §§4.1 and 4.2), and c and d are k-dimensional vectors. We see that the desired c and d are the right and left eigenvector corresponding to the eigenvalue θ of the projected (generalized) eigensystem

$$V^*AUc = \theta\, V^*Uc \qquad \text{and} \qquad U^*A^*Vd = \overline{\theta}\, U^*Vd. \tag{4.2}$$

To expand the search spaces $\mathcal{U}$ and $\mathcal{V}$, the two-sided JD method looks for corrections s and t (not necessarily orthogonal to u and v, respectively, see §§4.1 and 4.2) such that

$$\left\{\begin{array}{rcl} A(u+s) &=& \lambda(u+s)\\ A^*(v+t) &=& \overline{\lambda}(v+t).\end{array}\right.$$

For the *right correction equation* this means

$$(A-\lambda I)s = -r_u + (\lambda-\theta)u. \tag{4.3}$$

As in Section 2, we consider the projection of this equation that maps u to 0 and fixes r_u. In this situation $r_u \perp v$, so the sought (oblique) projector is given by $P = I - \frac{uv^*}{v^*u}$. P is an approximation to the spectral projector, just as $I - uu^*$ is in the normal case (see §2). The projection and the (second order) replacement of λ by θ yields:

$$\left(I - \frac{uv^*}{v^*u}\right)(A-\theta I)s = -r_u. \tag{4.4}$$

In a similar way we get for the *left correction equation*

$$\left(I - \frac{vu^*}{u^*v}\right)(A^* - \overline{\theta}I)t = -r_v. \tag{4.5}$$

We now discuss two variants of the two-sided JD approach: one where the pair (U, V) is bi-orthogonal, and one where both U and V have orthogonal columns.

4.1. The pair (U, V) bi-orthogonal. For the first variant of two-sided JD, we want the columns of U and V to be bi-orthogonal, that is, V^*U should be a diagonal matrix. This is a natural idea, because the right eigenvector corresponding to a particular eigenvalue is orthogonal to the left eigenvector corresponding to a different eigenvalue. This choice has the advantage that the projected eigenproblem (4.2) is easily transformed into a standard eigenproblem. Since in this variant we look for bi-orthogonal corrections $s \perp v$ and $t \perp u$, the correction equations (4.4) and (4.5) can be written as

$$\left\{\begin{array}{lcll} \left(I - \frac{uv^*}{v^*u}\right)(A - \theta I)\left(I - \frac{uv^*}{v^*u}\right)s & = & -r_u & \qquad (s \perp v) \\ \left(I - \frac{vu^*}{u^*v}\right)(A^* - \overline{\theta}I)\left(I - \frac{vu^*}{u^*v}\right)t & = & -r_v & \qquad (t \perp u). \end{array}\right.$$

The operator in the first equation is the conjugate transpose of the operator in the second equation, so these equations may be solved simultaneously by Bi-Conjugate Gradients (BiCG). Note that BiCG tries to solve two equations; but often only one approximate solution is used, the other solution solves a shadow equation and has no practical interest. In this situation we do use both approximate solutions from BiCG; r_v takes the role of the shadow residual. Of course, we can also deal with the correction equations separately, for instance we may try to solve each of them by a few steps of (preconditioned) GMRES (see the numerical experiments).

The resulting algorithm for the computation of the eigenvalue with the largest magnitude is shown in Algorithm 4.2. If one is interested in other eigenvalues, one should change the choice in step 4 of the algorithm accordingly (possibly using harmonic Ritz values, see for instance [8]). Also recall that $V_k^*U_k$ is a diagonal matrix. In step 2 of the algorithm, Bi-MGS stands for Bi-Modified Gram–Schmidt, used to make U_k and V_k bi-orthogonal.

Note that if the algorithm terminates, we have in general found only one eigenvector, say the right eigenvector, to the prescribed tolerance. Often we will also have a good approximation to the left eigenvector (see also the numerical experiments), but this is not necessarily the case. In any case, it is not sensible to continue with the algorithm, for we would then perform superfluous calculations for one of the eigenvectors. If we want to have both eigenvectors accurately, then, at the end of Algorithm 4.2, it suffices to (reasonably accurately) solve $t \perp v$ from the system

$$(I - vv^*)(A - \theta I)^*(I - vv^*)t = -r_v, \tag{4.6}$$

where v is the (often good) approximate left eigenvector from Algorithm 4.2. Solving one such system will in general be enough, since θ is a very good approximation to λ. Instead of (4.6), we may also solve a correction equation with oblique projections, but we prefer (4.6) for stability reasons.

Input: a device to compute Ax and A^*x or arbitrary x,
starting vectors u_1 and v_1 ($v_1^* u_1 \neq 0$), and a tolerance ε.
Output: approximations (θ, u, v) to the largest eigenvalue of A and its left and right eigenvector satisfying $\min\{\|Au - \theta u\|, \|A^*v - \overline{\theta} v\|\} \leq \varepsilon$.

1. $s = u_1, \ t = v_1$
   for $k = 1, 2, \ldots$
2. $(U_k, V_k) = $ Bi-MGS (U_{k-1}, V_{k-1}, s, t)
3. Compute kth column of $W_k = AU_k$
   Compute kth row and column of $H_k = V_k^* W_k$
4. Compute the 'largest' eigentriple (θ, c, d) of the pencil $(V_k^* A U_k, V_k^* U_k)$
5. $u = U_k c / \|U_k c\|$, $v = V_k d / \|V_k d\|$, $(\theta = \frac{v^* A u}{v^* u})$
6. $r_u = (A - \theta I) u = W_k c - \theta u$
   $r_v = (A^* - \overline{\theta} I) v$
7. Stop if $\min\{\|r_u\|, \|r_v\|\} \leq \varepsilon$ (and compute second vector at will)
8. Solve (approximately) $s \perp v$, $t \perp u$ from
   $\left(I - \frac{uv^*}{v^*u}\right)(A - \theta I)\left(I - \frac{uv^*}{v^*u}\right) s = -r_u$
   $\left(I - \frac{vu^*}{u^*v}\right)(A^* - \overline{\theta} I)\left(I - \frac{vu^*}{u^*v}\right) t = -r_v$

Alg. 4.2: The bi-orthogonal two-sided JD algorithm for the computation of the eigenvalue with the largest magnitude.

4.2. Both U and V orthogonal. Another obvious idea is to keep the columns of both U and V orthogonal. Because in this variant we look for updates $s \perp u$, $t \perp v$, the two correction equations now take the form

$$\left\{ \begin{array}{rcll} \left(I - \frac{uv^*}{v^*u}\right)(A - \theta I)(I - uu^*)\, s & = & -r_u & \qquad (s \perp u) \\ \left(I - \frac{vu^*}{u^*v}\right)(A^* - \overline{\theta} I)(I - vv^*)\, t & = & -r_v & \qquad (t \perp v). \end{array} \right. \tag{4.7}$$

This leads to Algorithm 4.3 for the computation of the eigenvalue with the largest magnitude. In step 2 of the algorithm, MGS stands for Modified Gram–Schmidt, used to make U_k and V_k orthogonal. A problem in this variant is that the operator in the first equation in (4.7) maps $u^\perp$ onto $v^\perp$, while the operator in the second equation maps $v^\perp$ onto $u^\perp$. As also observed in [7, § 3.3], it is unnatural to repeat such an operator, so it seems unattractive to solve the equations in (4.7) by a Krylov solver. However, we can fix this by working with preconditioners of the form $\left(I - \frac{au^*}{u^*a}\right) M_1^{-1} \left(I - \frac{uv^*}{v^*u}\right)$ for the first equation ($a \not\perp u$), which maps $v^\perp$ back to $u^\perp$, and $\left(I - \frac{bv^*}{v^*b}\right) M_2^{-1} \left(I - \frac{vu^*}{u^*v}\right)$ for the second equation ($b \not\perp v$), which maps $u^\perp$ back to $v^\perp$. Let us consider the simplest case, $M_1 = M_2 = I$, for a moment. With $a = u$ and $v = b$ we get

$$\left\{ \begin{array}{rcl} (I - uu^*)(A - \theta I)(I - uu^*)\, s & = & -(I - uu^*)\, r_u \\ (I - vv^*)(A^* - \overline{\theta} I)(I - vv^*)\, t & = & -(I - vv^*)\, r_v. \end{array} \right. \tag{4.8}$$

Here we have used $(I - uu^*)\left(I - \frac{uv^*}{v^*u}\right) = I - uu^*$ and $(I - vv^*)\left(I - \frac{vu^*}{u^*v}\right) = I - vv^*$. We recognize these equations as the correction equations of standard JD applied to A

and A^*; so for this special case we get a version of two-sided JD, where the correction equations are the same as in standard JD. Another natural choice is $a = v$, $b = u$, then on the right hand side, a multiple of v is added to r_u. Because $r_u \perp v$, all information of r_u (and similarly all information of r_v) stays intact. Although this approach seems to be more natural, numerical experiments (not reported here) indicate that (4.8) leads to faster convergence.

Input: a device to compute Ax and A^*x or arbitrary x,
starting vectors u_1 and v_1 ($v_1^* u_1 \neq 0$), and a tolerance ε.
Output: approximations (θ, u, v) to the largest eigenvalue of A and its
left and right eigenvector satisfying $\min\{\|Au - \theta u\|, \|A^*v - \overline{\theta}v\|\} \leq \varepsilon$.

1. $s = u_1, t = v_1$
   for $k = 1, 2, \ldots$
2. U_k = MGS (U_{k-1}, s)
   V_k = MGS (V_{k-1}, t)
3. Compute kth column of $W_k = AU_k$
   Compute kth row and column of $H = V_k^* W_k$
4. Compute the 'largest' eigentriple (θ, c, d) of the pencil $(V_k^* A U_k,\ V_k^* U_k)$
5. $u = U_k c$, $v = V_k d$, $(\theta = \frac{v^* A u}{v^* u})$
6. $r_u = (A - \theta I) u = W_k c - \theta u$
   $r_v = (A^* - \overline{\theta} I) v$
7. Stop if $\min\{\|r_u\|, \|r_v\|\} \leq \varepsilon$ (and compute second vector at will)
8. Solve (approximately) $s \perp u$, $t \perp v$ from
   $\left(I - \frac{uv^*}{v^*u}\right)(A - \theta I)\,(I - uu^*)\,s = -r_u$
   $\left(I - \frac{vu^*}{u^*v}\right)(A^* - \overline{\theta} I)\,(I - vv^*)\,t = -r_v$

Alg. 4.3: The orthogonal two-sided JD algorithm for the computation of the eigenvalue with the largest magnitude.

The following theorem states that exact two-sided Jacobi–Davidson, as two-sided RQI, has locally cubic convergence.

THEOREM 4.1. *If the two correction equations* (4.4) *and* (4.5) *are solved exactly, both the bi-orthogonal and the orthogonal variant of the two-sided Jacobi–Davidson process converge asymptotically cubically to an eigenvalue, if that eigenvalue is simple.*

Proof. The solution of (4.4) satisfies

$$s = -u + \zeta\,(A - \theta I)^{-1}\,u.$$

In the bi-orthogonal variant $s \perp v$, so $\zeta = \frac{v^* u}{v^* (A - \theta I)^{-1} u}$. In the orthogonal variant $s \perp u$, then $\zeta = \left(u^*\,(A - \theta I)^{-1}\,u\right)^{-1}$. We recognize the updated vector $u + s$ as a multiple of the one from two-sided RQI. For the left correction equation we have a similar expression. Now apply Theorem 3.1. ∎

5. Inexact two-sided RQI and Jacobi–Davidson. In the introduction we have already mentioned that two-sided JD and two-sided RQI are in practice often very expensive when we solve the linear systems, occurring in the methods ((3.2) and (4.4), (4.5)), accurately. In this section, we therefore consider inexact variants. In §§5.1 and 5.2, we investigate two-sided RQI and two-sided JD when the linear systems are solved to a certain precision (minimal residual approach). In §5.3, a relation between two-sided RQI and two-sided JD is established, when the linear systems are solved by a number of BiCG-steps (bi-orthogonal residual approach).

5.1. Inexact two-sided RQI. In [10] and [13], the authors study inexact RQI for Hermitian matrices. They show that the asymptotic convergence rate under certain assumptions is quadratic. Here we give a generalization for inexact two-sided RQI.

Consider the situation where we solve the two equations (3.2) of the two-sided RQI method inexactly, by which we mean that we are contented with u_{k+1}, v_{k+1} satisfying

$$\|(A-\theta I)u_{k+1}-u_k\| \le \xi_1 < 1 \qquad \text{and} \qquad \|(A-\theta I)^* v_{k+1}-v_k\| \le \xi_2 < 1. \tag{5.1}$$

Note that if we have nonsingular preconditioners $M_1 \approx A-\theta I$ and $M_2 \approx (A-\theta I)^*$, such that

$$\|(A-\theta I)M_1^{-1}-I\| \le \xi_1 \qquad \text{and} \qquad \|(A-\theta I)^* M_2^{-1}-I\| \le \xi_2,$$

then only one action with each preconditioner (that is, take $u_{k+1} := M_1^{-1}u_k$ and $v_{k+1} := M_2^{-1}v_k$) is enough to satisfy (5.1). However, since $A-\theta I$ is almost singular if $\theta \approx \lambda$, it is not a realistic assumption to have such preconditioners to our disposal.

To study the convergence rate of inexact two-sided RQI, the following lemma is useful.

LEMMA 5.1. *Let $v = \beta(y+\varepsilon e)$, $e \perp x$ (cf. (3.3)). The following statements are true:*

(a) $\|I - \frac{xy^*}{y^*x}\| = \kappa(\lambda)$

(b) $\|\frac{xy^*}{y^*x}\| = \kappa(\lambda)$

(c) $\|(I - \frac{xy^*}{y^*x})|_{v^\perp}\| \le \sqrt{1+\varepsilon^2\kappa(\lambda)^2}$

(d) $\|\frac{xy^*}{y^*x}|_{v^\perp}\| \le \varepsilon\kappa(\lambda)$.

Proof. Define $Q := I - \frac{xy^*}{y^*x}$. By examining the eigenpairs of Q^*Q we see that all singular values of Q restricted to the space $\text{span}\{x,y\}^\perp$ are equal to one. Likewise, the singular values of $I-Q$ restricted to the space $y^\perp$ are zero. Therefore, one may check that, up to a normalizing constant, $\text{argmax}_{z\neq 0}\frac{\|Qz\|}{\|z\|} = y-(x^*y)x$, and $\text{argmax}_{z\neq 0}\frac{\|(I-Q)z\|}{\|z\|} = y$, both with maximum $|y^*x|^{-1}$. This proves (a) and (b). Using $y = \beta^{-1}v - \varepsilon e$, we get that $I - \frac{xy^*}{y^*x} = I + \varepsilon\frac{xe^*}{y^*x}$ on the subspace $v^\perp$. Similar to the proof of (a) and (b), it is only of interest to consider the singular values of this operator on the subspace $\text{span}\{x,e\}^\perp$. Now $\|Qe\|^2 = \|e+\varepsilon\,(y^*x)^{-1}\,x\|^2 = 1+\varepsilon^2\kappa(\lambda)^2$, because $e \perp x$. Because in general $e \not\perp v$, (c) follows. Finally, (d) can be proved by noting that $\frac{xy^*}{y^*x} = -\varepsilon\frac{xe^*}{y^*x}$ on $v^\perp$. □

From the proof of part (a) and (b), we see that, when $\kappa(\lambda)$ is very large, $\text{argmax}(I - \frac{xy^*}{y^*x}) \approx \text{argmax}(\frac{xy^*}{y^*x})$. This implies that when the right eigenvector x and the corresponding left eigenvector y are nearly orthogonal, the decomposition $v = \eta_1 x + \eta_2 d$, $d \perp y$, has (almost equally) large η_1 and η_2 components. We are now ready to state the following result. As in Theorem 2.2, 'h.o.t.' stands for 'higher order terms in δ_k and ε_k'.

THEOREM 5.2 (Locally quadratic convergence of inexact two-sided RQI, generalization of [10, Cor. 4.3] and [13, Prop. 2.2]). *Suppose that* $\max\{\xi_1, \xi_2\} \cdot \kappa(\lambda) < 1$. *For one step of inexact two-sided* RQI, *where the equations are solved inexactly according to* (5.1), *we have (using the notation in* (3.3)*)*

$$\delta_{k+1} \le \gamma_1 \delta_k \varepsilon_k + \text{h.o.t.} \qquad \textit{and} \qquad \varepsilon_{k+1} \le \gamma_2 \delta_k \varepsilon_k + \text{h.o.t.}.$$

Here

$$\gamma_i = \kappa(\lambda)\kappa((A - \lambda I)|_{y^\perp}) \frac{\xi_i \kappa(\lambda)}{1 - \xi_i \kappa(\lambda)} \qquad (i = 1, 2).$$

Proof. From the first equation of (5.1) we know that there exists a $\widetilde{\xi}$, $0 \le \widetilde{\xi} \le \xi_1$, and a unit vector f such that

$$(A - \theta_k I)u_{k+1} = u_k + \widetilde{\xi} f.$$

Decomposing f in an x-component and a component orthogonal to y, we get with Lemma 5.1(a,b) that

$$(A - \theta_k I)u_{k+1} = \widetilde{\alpha} x + \widetilde{\delta}\widetilde{d},$$

where $\widetilde{d} \perp y$, $|\widetilde{\alpha}| \ge |\alpha_k| - \xi_1\kappa(\lambda)$, and $\widetilde{\delta} \le |\alpha_k|\delta_k + \xi_1\kappa(\lambda)$. Moreover, we have the estimates

$$|\alpha_k| = 1 + \text{h.o.t.}, \qquad \text{and} \qquad |\beta_k| = 1 + \text{h.o.t.}. \tag{5.2}$$

The value of γ_1 now follows, analogous to the proof of Theorem 3.1, from bounding $|\widetilde{\delta}/\widetilde{\alpha}|$, and γ_2 is derived in a similar manner. ∎

When we have preconditioners to our disposal, we may also try to solve the (left) preconditioned equations to a certain precision, e.g.

$$\|M_1^{-1}((A - \theta I)u_{k+1} - u_k)\| \le \xi_1,$$

Just as for Theorem 5.2, one can prove that this yields locally quadratic convergence, now with constants

$$\gamma_i = \kappa(\lambda)\kappa((A - \lambda I)|_{y^\perp}) \frac{\xi_i \kappa(\lambda)\|M_i\|}{1 - \xi_i \kappa(\lambda)\|M_i\|} \qquad (i = 1, 2).$$

As $\theta \approx \lambda$, the condition number of the matrix $A - \theta_k I$ increases. Therefore, it may get more and more expensive to solve (5.1) to a certain tolerance. This forms a motivation to study inexact two-sided JD.

5.2. Inexact two-sided Jacobi–Davidson. In [13], the author studies inexact JD for Hermitian matrices. He shows that the asymptotic convergence rate is linear under certain assumptions. Here we give a generalization for two-sided JD.

Consider the situation where we solve the two equations (4.4) and (4.5) of the two-sided JD method inexactly, by which we mean that we are satisfied with $\widetilde{s} \perp v$ and $\widetilde{t} \perp u$ (bi-orthogonal variant) or $\widetilde{s} \perp u$ and $\widetilde{t} \perp v$ (orthogonal variant) where

$$\left\| \left(I - \frac{uv^*}{v^*u} \right) (A - \theta I)\widetilde{s} + r_u \right\| \le \xi_1 \|r_u\| \tag{5.3}$$

and

$$\left\| \left(I - \frac{vu^*}{u^*v} \right) (A - \theta I)^* \widetilde{t} + r_v \right\| \le \xi_2 \, \|r_v\|, \tag{5.4}$$

for some $0 < \xi_1, \xi_2 < 1$. The next theorem states that the resulting local convergence is linear.

THEOREM 5.3 (Locally linear convergence of inexact two-sided Jacobi–Davidson, generalization of [13, Th. 4.1]). *For one step of inexact bi-orthogonal two-sided Jacobi–Davidson, when the equations are solved inexactly according to* (5.3) *and* (5.4), *we have (using the notation in* (3.3)*)*

$$\delta_{k+1} \le \gamma_1 \delta_k + \text{h.o.t.} \qquad \textit{and} \qquad \varepsilon_{k+1} \le \gamma_2 \varepsilon_k + \text{h.o.t.}.$$

Here

$$\gamma_i = \xi_i \kappa((A - \lambda I)|_{y^\perp}) \qquad (i = 1, 2).$$

The orthogonal variant of two-sided Jacobi–Davidson has locally linear convergence as well.

Proof. For clarity, we leave out the index k. Write $P = I - \frac{uv^*}{v^*u}$. Let us first consider the bi-orthogonal variant of Section 4.1, where $\widetilde{s} \perp v$. From (5.3) we know that there exists a $\widetilde{\xi}$, $0 \le \widetilde{\xi} \le \xi_1$, and a unit vector $f \perp v$ such that

$$P(A - \theta I)P\widetilde{s} = -r_u + \widetilde{\xi} \, \|r_u\| f. \tag{5.5}$$

From (4.3) we know that the 'real update' $s \perp v$ satisfies

$$P(A - \lambda I)Ps = -r_u,$$

hence

$$P(A - \theta I)Ps = -r_u + (\lambda - \theta)s. \tag{5.6}$$

Both $u + s$ ($s \perp v$) and $u - \alpha\delta d$ ($d \perp y$) are multiples of the eigenvector x, and (5.2) and Lemma 5.1(c,d) give that $\|s\| = \delta + \text{h.o.t.}$. Subtracting (5.6) from (5.5) gives

$$P(A - \theta I)P(\widetilde{s} - s) = \widetilde{\xi} \, \|r_u\| f - (\lambda - \theta)s.$$

The operator $P(A - \theta I)P$ is a bijection from $v^\perp$ to $v^\perp$ and

$$\| \left((P(A - \theta I)P)|_{v^\perp}\right)^{-1} \| = \| \left((A - \lambda I)|_{y^\perp}\right)^{-1} \| + \text{h.o.t.}.$$

For the norm of the residual we have by (3.4)

$$\begin{aligned} \|r_u\| &= \|\alpha((\lambda - \theta)x + \delta(A - \theta I)d)\| \\ &= \delta \, \|(A - \theta I)d\| + \text{h.o.t.} \\ &\le \delta \, \|(A - \theta I)|_{y^\perp}\| + \text{h.o.t.} \\ &= \delta \, \|(A - \lambda I)|_{y^\perp}\| + \text{h.o.t.}. \end{aligned}$$

So

$$\|\widetilde{s} - s\| \le \delta \xi_1 \kappa((A - \lambda I)|_{y^\perp}) + \text{h.o.t.}.$$

The term $(\widetilde{s} - s) \perp v$ represents the error in the updated vector $u + \widetilde{s}$. Again using Lemma 5.1(c,d), we get $\delta_{k+1} = \|\widetilde{s} - s\| + $ h.o.t.. This proves the statement for the bi-orthogonal variant. Now consider the orthogonal variant of Section 4.2. The essential difference is that $\widetilde{s} \perp u$. Along the same lines, it can be shown that $\|r_u\| \le \delta \, \|(A - \lambda I)|_{x^\perp}\| + $ h.o.t.. In the same way as in the proof for the bi-orthogonal case, we get

$$\|\widetilde{s} - s\| \le \xi_1 \kappa((A - \lambda I)|_{x^\perp}) \|s\| + \mathcal{O}(\|s\|^2),$$

where $(A - \lambda I)|_{x^\perp}$ is interpreted as operator from $x^\perp$ to $y^\perp$. This estimate means locally linear convergence. $\square$

Comparing inexact two-sided RQI with inexact two-sided JD, we remark that it is by no means possible to conclude from Theorems 5.2 and 5.3 that "inexact two-sided RQI is faster than inexact two-sided JD". Firstly, the theorems only say something about the local, not the global, rate of convergence. In fact, it can happen that inexact two-sided RQI does not converge at all (see the numerical experiments, Fig.9.1(b)), while inexact two-sided JD trivially converges in a finite number of steps. Secondly and more importantly, the theorems do not tell how much effort it takes to solve the equations in question ((5.1) versus (5.3) and (5.4)) to a certain precision. In the proof of the previous theorem we have seen that the effective condition number of $\left(I - \frac{uv^*}{v^*u}\right)(A - \theta I)\left(I - \frac{uv^*}{v^*u}\right)$ approaches $\kappa((A - \lambda I)|_{y^\perp})$ as $\theta \to \lambda$, $u \to x$, $v \to y$, while the condition number of $A - \theta I$ is unbounded as $\theta \to \lambda$. Therefore, it may be much more difficult to solve (5.1). Thirdly, in the next section we show that the solutions to the linear systems are the same if the systems are solved by unpreconditioned BiCG with a fixed number of steps. This result seems to confirm our second observation.

5.3. Relation between inexact two-sided Jacobi–Davidson and inexact two-sided Rayleigh Quotient Iteration. We have already seen that two-sided Jacobi–Davidson is equivalent to accelerated two-sided RQI if all linear systems ((3.2) and (4.4), (4.5)) are solved exactly. In [1], the somewhat surprising result is proved that for Hermitian matrices, standard Jacobi–Davidson is equivalent to accelerated RQI, when all linear equations are solved by a certain number of steps of Conjugate Gradients. We generalize this result, and show that two-sided Jacobi–Davidson for nonnormal matrices is also equivalent to accelerated two-sided RQI, if the linear systems are solved by a certain number of steps of BiCG. In the next lemma, C plays the role of $A - \theta I$.

Lemma 5.4 (Generalization of [11, Lemma 4.1]). *Let $P = I - \frac{uv^*}{v^*u}$ and $r = Cu$. Then for all $m \ge 1$,*

$$\mathit{span}\{u, r, (PCP)r, \ldots, (PCP)^{m-1}r\} = \mathit{span}\{u, r, Cr, \ldots, C^{m-1}r\}.$$

Proof. Let $\mathcal{K}_m = \operatorname{span}\{u, r, Cr, \ldots, C^{m-1}r\}$ and $\mathcal{L}_m = \operatorname{span}\{u, r, (PCP)r, \ldots, (PCP)^{m-1}r\}$. The proof is by induction. For $m = 1$, the claim is evidently true. Now assume that $\mathcal{K}_j = \mathcal{L}_j$ for all $j < m$. If $a \in \mathcal{L}_m$, then there exist $b \in \mathcal{L}_1 = \mathcal{K}_1$, and $c \in \mathcal{L}_{m-1} = \mathcal{K}_{m-1}$ such that $a = b + (PCP)c$. Writing out the projection P, we get

$$a = b + \left(I - \frac{uv^*}{v^*u}\right) C \left(I - \frac{uv^*}{v^*u}\right) c = b + Cc - \frac{v^*c}{v^*u} Cu + \left(\frac{(v^*Cu)(v^*c)}{(v^*u)^2} - \frac{v^*Cc}{v^*u}\right) u.$$

Now $Cc \in \mathcal{K}_m$, and all other terms are in $\mathcal{K}_1$, so $a \in \mathcal{K}_m$ and $\mathcal{L}_m \subset \mathcal{K}_m$.

If $\mathcal{L}_m$ is of full rank, then the lemma is proved. Otherwise, let j be the largest index such that $\mathcal{L}_j$ is full rank, then $\mathcal{L}_{j+1} = \mathcal{L}_j = \mathcal{K}_j$. Let now $c \in \mathcal{K}_j$, then we deduce that also $PCPc \in \mathcal{K}_j$. From an equation similar to the one displayed above, we see that $Cc \in \mathcal{K}_j$, so $\mathcal{K}_{j+1} = \mathcal{K}_j$. By induction we have $\mathcal{L}_m = \mathcal{L}_j = \mathcal{K}_j = \mathcal{K}_m$ for all $m \geq j$. □

PROPOSITION 5.5 (Generalization of [1, Prop. 3.2]). *Let u and v be approximate eigenvectors. Let $\widetilde{s}_m, \widetilde{t}_m$ be the approximate solutions to the right and left* JD *correction equation (equations* (4.4) *and* (4.5)*), respectively, obtained by m steps of the* BiCG *method, without suffering from a breakdown. Let $\widetilde{u}_{m+1}$ and $\widetilde{v}_{m+1}$ be the approximate solutions to the two-sided* RQI *equations (equation* (3.2)*), obtained by $m+1$ steps of* BiCG. *Then there exist μ_1, μ_2 such that*

$$\widetilde{u}_{m+1} = \mu_1(u + \widetilde{s}_m), \qquad \textit{and} \qquad \widetilde{v}_{m+1} = \mu_2(v + \widetilde{t}_m).$$

Proof. Let the columns of W_m and Z_m be bi-orthogonal bases for $\text{span}\{r_u, (P(A-\theta I)P)r_u, \ldots, (P(A-\theta I)P)^{m-1}r_u\}$ and $\text{span}\{r_v, (P^*(A-\theta I)^*P^*)r_v, \ldots, (P^*(A-\theta I)^*P^*)^{m-1}r_v\}$, respectively. Apply BiCG to the JD correction equations, then $\widetilde{s}_m$ and $\widetilde{t}_m$ are of the form $\widetilde{s}_m = W_m w$, $\widetilde{t}_m = Z_m z$, where w, z are solutions of

$$Z_m^* P(A-\theta I)PW_m w = -Z_m^* r_u \qquad \text{and} \qquad W_m^* P^*(A-\theta I)^* P^* Z_m z = -W_m^* r_v.$$

Now note that $v \perp W_m$ and $u \perp Z_m$, so $PW_m = W_m$ and $P^* Z_m = Z_m$. Hence w, z solve

$$Z_m^*(A-\theta I)W_m w = -Z_m^* r_u \qquad \text{and} \qquad W_m^*(A-\theta I)^* Z_m z = -W_m^* r_v.$$

On the other hand, according to Lemma 5.4, the columns of $[u\ W_m]$ and $[v\ Z_m]$ are bi-orthogonal bases for $\text{span}\{u, (A-\theta I)u, \ldots, (A-\theta I)^{m-1}u\}$ and $\text{span}\{v, (A-\theta I)^*v, \ldots, ((A-\theta I)^*)^{m-1}v\}$, respectively. Hence, BiCG applied to the two-sided RQI equations gives approximations $\widetilde{u}_{m+1}$ and $\widetilde{v}_{m+1}$ of the form $\widetilde{u}_{m+1} = W_m p + \mu_1 u$, $\widetilde{v}_{m+1} = Z_m q + \mu_2 v$, where p, q, μ_1, μ_2 are determined by the Petrov-Galerkin conditions

$$(5.7) \qquad \widehat{A}\begin{bmatrix} \mu_1 \\ p \end{bmatrix} = \begin{bmatrix} v^*u \\ Z_m^*u \end{bmatrix} \qquad \text{and} \qquad \widehat{A}^*\begin{bmatrix} \mu_2 \\ q \end{bmatrix} = \begin{bmatrix} u^*v \\ W_m^*v \end{bmatrix},$$

where

$$\widehat{A} := \begin{bmatrix} v^*(A-\theta I)u & v^*(A-\theta I)W_m \\ Z_m^*(A-\theta I)u & Z_m^*(A-\theta I)W_m \end{bmatrix}.$$

The terms Z_m^*u and W_m^*v in (5.7) vanish. From the (twice) last $n-1$ equations in (5.7) we get

$$\begin{aligned} Z_m^*(A-\theta I)W_m p &= -\mu_1 Z_m^*(A-\theta I)u \\ W_m^*(A-\theta I)^* Z_m q &= -\mu_2 W_m^*(A-\theta I)^* v. \end{aligned}$$

Because of the assumption that no breakdown is encountered, $Z_m^*(A-\theta I)W_m$ is invertible and the proposition is proved. □

Commenting on the number of BiCG-steps in the previous proposition, note that it is natural that two-sided RQI needs one step more ($m+1$ versus m), because two-sided JD already uses two matrix-vector multiplications to compute the residuals.

Based on this proposition, it is tempting to conclude that two-sided JD and two-sided RQI are just the same. But the proposition only gives a statement for the situation when the linear systems ((3.2) and (4.4), (4.5)) are solved by unpreconditioned BiCG; and even then JD uses subspace acceleration, and RQI does not. Preconditioning can be included in JD and in RQI in such a way that Proposition 5.5 still holds for the preconditioned methods. However, this preconditioning for JD will not be the one described in [8, 7] and seems to be less effective.

6. Alternating Jacobi–Davidson. Theoretically, RQI does not need to converge globally. Therefore, Parlett [5] proposes a different generalization of RQI to ensure global convergence, which he calls *alternating Rayleigh Quotient Iteration*; see Algorithm 6.4.

Pick an initial vector u_1 with unit norm
for $k = 1, 3, \ldots$
1. Compute $\theta_k = \theta(u_k) = u_k^* A u_k$
2. Solve $(A - \theta_k I) u_{k+1} = u_k$ and normalize u_{k+1}
3. Compute $\theta_{k+1} = \theta(u_{k+1}) = u_{k+1}^* A u_{k+1}$
4. Solve $(A^* - \overline{\theta}_{k+1} I) u_{k+2} = u_{k+1}$ and normalize u_{k+2}

If $A - \theta_k I$ or $A^* - \overline{\theta}_{k+1} I$ happen to be singular, solve for the eigenvectors.

Alg. 6.4: Parlett's alternating Rayleigh Quotient Iteration [5].

This method is somewhat counter-intuitive, because $(A - \theta I)^{-1}$ and $(A - \theta I)^{-*}$ are used alternatingly on the iterates. For fixed θ, every two steps of the algorithm result in one action with $((A - \theta I)^* (A - \theta I))^{-1}$. This method could therefore be interpreted as an attempt to find the smallest singular value and corresponding singular vectors of $A - \theta I$. As such, it can be considered as a method for the singular value problem, rather than one for the eigenvalue problem. But because a matrix has a zero eigenvalue if and only if it has a zero singular value, the method can asymptotically (that is, for $\theta \approx \lambda$) also be regarded as an eigenvalue method. Parlett shows that the (only) advantage of this process is that it converges for all starting vectors, while it has a big drawback: the asymptotic convergence is in general only linear with factor close to one $(1 - \kappa(\lambda)^{-2})$ when applied to nonnormal matrices [5, p. 692].

Alternating RQI gives us inspiration for a new Jacobi–Davidson variant, which we call *alternating Jacobi–Davidson*. The idea is to accelerate Parlett's process, building up one (orthogonal) search space for *both* the left *and* the right eigenvector. Every odd step focuses on approximating the right eigenvector, every even step on approximating the left eigenvector, see Algorithm 6.5.

Because of the subspace acceleration, the convergence behavior of alternating JD is much better than that of alternating RQI. For nonnormal matrices, the odd or even steps alone guarantee us quadratic convergence (when the correction equations are solved exactly). For normal matrices, one can check that alternating JD does exactly the same as standard JD, so with the same amount of work we get cubic convergence. Our hope is that alternating JD will have fast convergence for (slightly) nonnormal matrices with only a modest amount of extra work. Numerical experiments show that alternating JD can even be faster that standard JD (see §9).

7. Generalizations.

Input: a device to compute Ax and A^*x or arbitrary x,
a starting vector y_1, and a tolerance ε.
Output: approximations (θ, u, v) to the largest eigenvalue of A and its
left and right eigenvector satisfying $\min\{\|Au - \theta u\|, \|A^*v - \overline{\theta}v\|\} \le \varepsilon$.

1. $s = y_1$
   for $k = 1, 2, \ldots$
2. $Y_k = \text{MGS}\ (Y_{k-1}, s)$
3. Compute kth column of $Z_k = AY_k$
   Compute kth row and column of $M_k = Y_k^* Z_k$
4. Compute the 'largest' eigenpair
   (θ, c) of $M_k = Y_k^* A Y_k$ (**k even**)
   $(\overline{\theta}, c)$ of $M_k^* = Y_k^* A^* Y_k$ (**k odd**)
5. $y = Y_k c$
6. $r = Ay - \theta y = Z_k x - \theta y$ (**k even**)
   $r = A^* y - \overline{\theta} y$ (**k odd**)
7. Stop if $\|r\| \le \varepsilon$ (and compute second vector at will)
8. Solve (approximately) $s \perp u$ from
   $(I - uu^*)(A - \theta I)(I - uu^*)s = -r$ (**k even**)
   $(I - uu^*)(A^* - \overline{\theta} I)(I - uu^*)s = -r$ (**k odd**)

Alg. 6.5: The alternating JD algorithm for the computation of the eigenvalue with the largest magnitude.

7.1. The generalized eigenproblem. Two-sided and alternating Jacobi–Davidson can easily be generalized. Let us, for instance, examine the adaptations to apply two-sided JD to the generalized eigenproblem $Ax = \lambda Bx$. The Galerkin conditions $(A - \theta B)u \perp \mathcal{V}$ and $(A - \theta B)^* v \perp \mathcal{U}$ lead to the approximations

$$\theta = \frac{v^* A u}{v^* B u},$$

where u and v are the backtransformed right and left eigenvectors of the projected pencil $(V^*AU,\, V^*BU)$. For bi-orthogonal two-sided JD, the right correction equation now becomes

$$\left(I - \frac{Buv^*}{v^*Bu}\right)(A - \theta B)\left(I - \frac{Buv^*}{v^*Bu}\right) s = -(A - \theta B)u, \quad (s \perp v). \tag{7.1}$$

If we solve this correction equation exactly, then we get

$$s = -u + \zeta\,(A - \theta B)^{-1} Bu,$$

so that exact two-sided JD can also in this case be viewed as accelerated "generalized two-sided RQI" (see e.g., [6, Th. 15.9.3] for the symmetric case), leading to cubic convergence:

Proposition 7.1. *Let B be nonsingular, and let λ be a simple eigenvalue of $B^{-1}A$. Then exact two-sided Jacobi–Davidson converges locally cubically.*

Proof. First, observe that λ is a simple eigenvalue of $B^{-1}A$ if and only if λ is a simple eigenvalue of AB^{-1}. Now, note that $(A - \theta B)^{-1} Bu = \left(B^{-1}A - \theta I\right)^{-1} u$ and apply Theorem 3.1. $\square$

Note that, we get cubic convergence using $\alpha Au + \beta Bu$ instead of Bu as well, because Au and Bu are linearly dependent in the limit.

Observe that, in contrast to the correction equations for the standard eigenvalue problem, the operators in the right and in the left correction equation (cf., (7.1)) are not complex conjugated, which obstructs a direct application of BiCG.

7.2. The polynomial eigenproblem. We now derive the right correction equation of two-sided JD for the polynomial eigenproblem

$$p(\lambda)x = 0, \tag{7.2}$$

where

$$p(\lambda) = \lambda^l A_l + \lambda^{l-1} A_{l-1} + \cdots + \lambda A_1 + A_0.$$

Suppose that we have approximate right and left eigenvector $u \in \mathcal{U}$ and $v \in \mathcal{V}$, where $\mathcal{U}$ and $\mathcal{V}$ are, as before, the right and left search spaces. The Petrov-Galerkin condition $p(\theta)u \perp \mathcal{V}$ implies that $\theta = \theta(u, v)$ satisfies

$$\sum_l (v^* A_l u)\, \theta^l = 0. \tag{7.3}$$

To derive Newton's method for (7.2), consider

$$p(\theta)(u + h) - p(\theta)(u) = p(\theta)(h) + \sum_l l\, \theta^{l-1} A_l u \frac{\partial \theta}{\partial u} h + \mathcal{O}(\|h\|^2).$$

Differentiating (7.3) with respect to u gives

$$\sum_l l\, (v^* A_l u)\, \theta^{l-1} \frac{\partial \theta}{\partial u} + \sum_l \theta^l v^* A_l = 0,$$

so with the notation $z := p'(\theta)u = \sum_l l\, \theta^{l-1} A_l u$ we find, if $v^* z \neq 0$,

$$\frac{\partial \theta}{\partial u} = -\left(\sum_l l\, (v^* A_l u) \theta^{l-1} \right)^{-1} v^* \left(\sum_l \theta^l A_l \right) = -\,(v^* z)^{-1}\, v^* p(\theta).$$

Hence, the Jacobian $\frac{\partial p(\theta)}{\partial u}$ is equal to $\left(I - \frac{z v^*}{v^* z}\right) p(\theta)$, and a Newton step solves s from

$$\left(I - \frac{z v^*}{v^* z}\right) p(\theta) s = -p(\theta) u.$$

When we work with bi-orthogonal bases, we search for $s \perp v$ and the equation is equivalent to

$$\left(I - \frac{z v^*}{v^* z}\right) p(\theta) \left(I - \frac{z v^*}{v^* z}\right) s = -p(\theta) u.$$

This is the right correction equation for the polynomial eigenvalue problem, see also [7, Eq. (8.4)], where the result is stated without derivation. For the special case $l = 1$ we recognize equation (7.1), and if, in addition, $A_1 = -I$, we recognize (4.4). Because it is a Newton method, we expect locally quadratic convergence. This is, under some conditions, true indeed [2, Th. 2].

8. Various issues.

8.1. Deflation. If we have found one or more eigentriples of A, and we want to find another, we can deflate to avoid finding the same value again. Suppose that we have already found the right eigenvectors x_i and corresponding left eigenvectors y_i. Then it can be verified that, if we found the exact vectors,

$$\widetilde{A} = \prod_i \left(I - \frac{x_i y_i^*}{y_i^* x_i} \right) \cdot A \cdot \prod_i \left(I - \frac{x_i y_i^*}{y_i^* x_i} \right)$$

has the same eigentriples as A, except that the found eigenvalues are transformed to zeros.

8.2. Comparison with two-sided Lanczos. Suppose that we do not solve the corrections equations (4.4) and (4.5), but just take $\widetilde{s} = r_u = Au - \theta u$ and $\widetilde{t} = r_v = A^* v - \overline{\theta} v$. Because of the orthogonalization at step 2 of Algorithm 4.2, this is equivalent to taking $\widetilde{s} = Au$ and $\widetilde{t} = A^* v$, which is the subspace expansion of two-sided Lanczos. Therefore two-sided JD may, besides as a generalization of standard JD, also be regarded as a generalization of two-sided Lanczos.

8.3. Breakdown. As two-sided Lanczos, two-sided JD may suffer from a breakdown, but in two-sided JD this can easily be overcome. First, BiCG (which we may use to solve the correction equations) may break down. Second, in the bi-orthogonal variant, the computed updates $\widetilde{s}$ and $\widetilde{t}$ may be (nearly) orthogonal. Realizing that our aim is to compute an eigenvalue and not to solve the correction equation accurately, we see that these breakdowns are not an intrinsic problem. In both cases, we can simply restart the method, or take different (e.g., random) approximate solutions to the correction equation.

9. Numerical experiments. Our experiments are coded in MATLAB and executed on a SUN workstation. We have already seen that JD has different convergence behavior for normal (cubic convergence) and nonnormal matrices (quadratic convergence); this in contrary to two-sided JD. The following lemma implies that two-sided JD does 'feel' a difference, but this is only noticeable in the norm of the residuals, and not in the approximations to the eigenvalue.

LEMMA 9.1. *Let $A = U_* \Lambda V_*^*$ be a diagonalizable matrix (so $V_*^* = U_*^{-1}$). If there are no rounding errors, and two-sided* JD*'s correction equations* (4.4) *and* (4.5) *in step k are solved by m_k steps of a Krylov method (without preconditioning), then two-sided* JD *applied to*

(a) A, with starting vectors u_1 and v_1

(b) Λ, with starting vectors $\widetilde{u}_1 := V_*^* u_1$ and $\widetilde{v}_1 := U_*^* v_1$

gives the same approximations: $\widetilde{\theta}_k = \theta_k$. Moreover, $\widetilde{u}_k = V_*^* u_k$ and $\widetilde{v}_k = U_*^* v_k$. In particular, if A is normal, then $\|\widetilde{u}_k\| = \|u_k\|$ and $\|\widetilde{v}_k\| = \|v_k\|$.

Proof. The first approximate eigenvalues are the same in both cases:

$$\widetilde{\theta}_1 := \frac{\widetilde{v}_1^* \Lambda \widetilde{u}_1}{\widetilde{v}_1^* \widetilde{u}_1} = \frac{v_1^* U_* \Lambda V_*^* u_1}{v_1^* (U_* V_*^*) u_1} = \frac{v_1^* A u_1}{v_1^* u_1} =: \theta_1.$$

For the right residuals in the first step of the method we have $\widetilde{r}_u^{(1)} = (\Lambda - \widetilde{\theta}_1 I)\widetilde{u}_1 = (\Lambda - \theta_1 I) V_*^* u_1$, so $U_* \widetilde{r}_u^{(1)} = r_u^{(1)}$. In the same way we find a similar relation for the left residuals: $V_* \widetilde{r}_v^{(1)} = r_v^{(1)}$. So $\widetilde{r}_u^{(1)} = V_*^* r_u^{(1)}$ and $\widetilde{r}_v^{(1)} = U_*^* r_v^{(1)}$. Denote $\mathcal{K}_m(A, r)$ for the Krylov subspace of dimension m, generated by A and r. For the Krylov subspaces we have (generalization of [6, p. 264]):

$$\mathcal{K}_m(A, r_u^{(1)}) = \mathcal{K}_m(U_*\Lambda V_*^*, U_*\widetilde{r}_u^{(1)}) = U_*\mathcal{K}_m(\Lambda, \widetilde{r}_u^{(1)}),$$

and likewise $\mathcal{K}_m(A^*, r_v^{(1)}) = \mathcal{K}_m(V_*\Lambda^* U_*^*, V_*\widetilde{r}_v^{(1)}) = V_*\mathcal{K}_m(\Lambda^*, \widetilde{r}_v^{(1)})$. With little extra work one can check that same relations hold for the shifted and projected matrices that are present in the correction equations, for instance

$$\left(I - \frac{uv^*}{v^*u}\right)(A - \theta I)\left(I - \frac{uv^*}{v^*u}\right) = U_*\left(I - \frac{\widetilde{u}\widetilde{v}^*}{\widetilde{v}^*\widetilde{u}}\right)(\Lambda - \theta I)\left(I - \frac{\widetilde{u}\widetilde{v}^*}{\widetilde{v}^*\widetilde{u}}\right)V_*^*.$$

So, using the notation $\widetilde{P} = I - \frac{\widetilde{u}\widetilde{v}^*}{\widetilde{v}^*\widetilde{u}}$,

$$\mathcal{K}_m(P(A - \theta I)P, r_u^{(1)}) = U_*\mathcal{K}_m(\widetilde{P}(\Lambda - \theta I)\widetilde{P}, \widetilde{r}_u^{(1)}).$$

We conclude that the approximate solutions from the first correction equations satisfy $\widetilde{s}^{(1)} = V_*^* s^{(1)}$ and $\widetilde{t}^{(1)} = U_*^* t^{(1)}$. By induction we can prove that $\widetilde{U}_k = V_*^* U_k$ and $\widetilde{V}_k = U_*^* V_k$, so the projected matrices are the same in both cases: $\widetilde{H}_k := \widetilde{V}_k^* \Lambda \widetilde{U}_k = V_k^* A U_k = H_k$. In particular, the approximations to the eigenvalues are the same, and the approximate eigenvectors (u_k, v_k) and $(\widetilde{u}_k, \widetilde{v}_k)$ are transformations of each other: $\widetilde{u}_k = V_*^* u_k$ and $\widetilde{v}_k = U_*^* v_k$. In particular, if A is normal, then U_* and V_* are orthogonal, and so $\|\widetilde{r}_k\| = \|r_k\|$. □

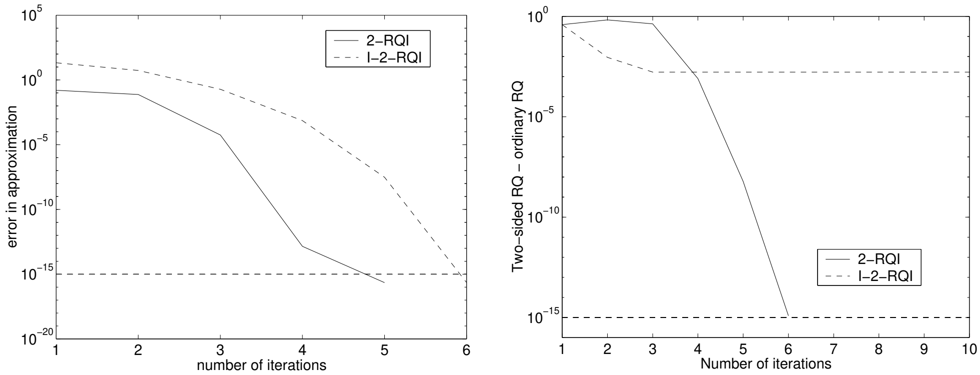
In the same way one may verify the next lemma.

LEMMA 9.2. *With the assumptions and notations of the previous lemma, if the equations of two-sided* RQI *(see* (3.2)*) in step k are solved by m_k steps of a Krylov method (without preconditioning), then two-sided* RQI *applied to*

(a) A, with starting vectors u_1 and v_1

(b) Λ, with starting vectors $\widetilde{u}_1 := V_*^* u_1$ and $\widetilde{v}_1 := U_*^* v_1$

gives 'the same' approximations: $\widetilde{\theta}_k = \theta_k$. Moreover, $\widetilde{u}_k = V_*^* u_k$ and $\widetilde{v}_k = U_*^* v_k$. In particular, if A is normal, then $\|\widetilde{u}_k\| = \|u_k\|$ and $\|\widetilde{v}_k\| = \|v_k\|$.

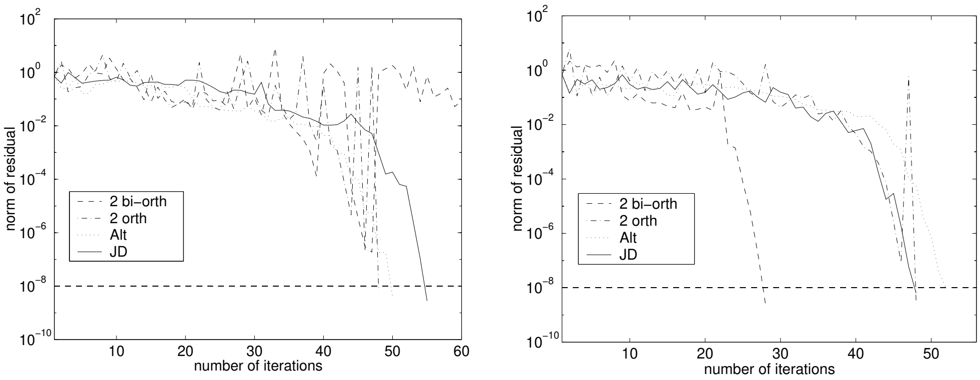


(a) (b)

FIG. 9.1: (a) The convergence of exact (solid line) and inexact (tolerance 0.5, dashed line) two-sided RQI for diag(1 : 100). (b) The difference of the two-sided Rayleigh Quotient ($\frac{v^*Au}{v^*u}$) and the right Rayleigh Quotient ($\frac{u^*Au}{u^*u}$) for exact (solid line) and inexact (tolerance 0.5, dashed line) two-sided RQI for the tridiagonal matrix with stencil (1 -2 1.2) of size 100×100.

Because of these results, our first example is $A = \text{diag}(1 : 100)$. In Figure 9.1(a), we compare exact two-sided RQI (solid line) and inexact two-sided RQI (dashed line), where the inner equations have been solved with five steps of GMRES. We take for u_1 and v_1 the 100th basisvector plus 0.2 times a random vector (MATLAB's function `rand`, 'seed' 0), and take $\xi_1 = \xi_2 = 0.5$ in (5.1). In this figure, we show the error $|\lambda - \theta|$ in the approximation to the detected eigenvalue λ. One may see the somewhat faster convergence for exact two-sided RQI. What we do not see in the figure is that inexact two-sided RQI converges to $\lambda = 100$, while exact two-sided RQI converges to $\lambda = 79$. Apparently, without subspace acceleration it is impossible to guide the process to the desired eigenvalue.

Figure 9.1(b) is an example of the fact that inexact two-sided RQI does not need to converge. Here A is the 100×100 tridiagonal matrix with stencil (1 -2 1.2), u_1 and v_1 are random vectors, and $\xi_1 = \xi_2 = 0.5$. We plot the difference between the two-sided Rayleigh Quotient ($\theta(u, v) = \frac{v^*Au}{v^*u}$) and the right Rayleigh Quotient ($\theta(u) = \frac{u^*Au}{u^*u}$). For inexact two-sided RQI, this difference (and the difference $\frac{v^*Au}{v^*u} - \frac{v^*A^*v}{v^*v}$) stabilizes. A small comfort is the fact that two-sided RQI can diagnose itself that there is a misconvergence.

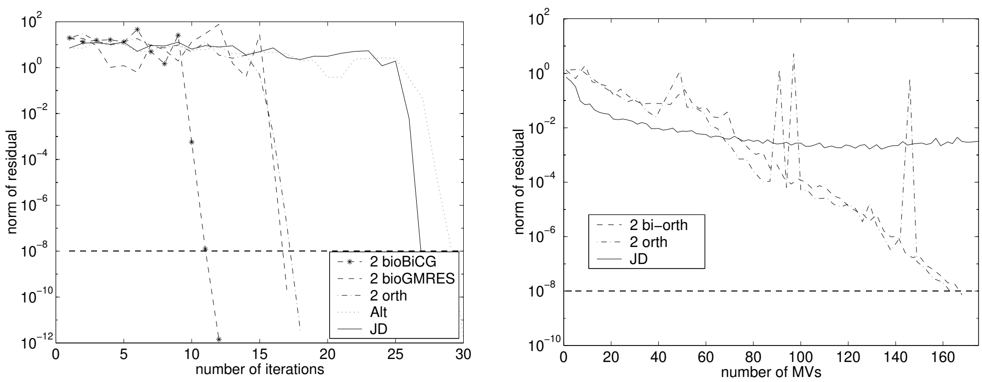


(a) (b)

FIG. 9.2: (a) The convergence histories of bi-orthogonal two-sided JD (dash), orthogonal two-sided JD (dash-dot), alternating JD (dot), and JD (solid) for the tridiagonal matrix with stencil (-1 2 1.2) of size 100×100. All correction equations are solved by 5 steps of GMRES. (b) The same as (a), but now 10 steps of GMRES.

Next, we experiment with two-sided JD variants. For Figure 9.2(a), we take the 100×100 tridiagonal matrix with stencil (-1 2 1.2). The starting vectors are random, and the correction equations are solved by five steps of unpreconditioned GMRES. All eigenvalues have real part equal to 2, and come in complex conjugate pairs. Note that for the two-sided methods, the plotted line always represents $\min\{\|r_u\|, \|r_v\|\}$. The horizontal dashed line shows the stopping tolerance. We see that alternating JD is faster (also measured in matrix-vector products (MVs)) than standard JD. For orthogonal two-sided JD we choose the variant of Equation (4.8). The method uses less iterations, but more MVs than standard JD. The convergence is very irregular; this might be improved using a target when one suspects that the process is converging. Bi-orthogonal two-sided JD almost converges, but then shows irregular behavior, and does not converge within 60 iterations. Using a target may be a good idea here as well.

For Figure 9.2(b), we change only the number of inner iteration steps to 10. Bi-orthogonal two-sided JD uses the fewest number of iterations. It uses slightly more MVs than JD. However, earlier in the process we already have more information. For instance, after 21 iterations of bi-orthogonal two-sided JD, $\kappa(\lambda) \approx 56.45$ is already approximated to a relative error of 0.5%: the condition number is well approximated before the method starts to converge. (21 iteration steps may not seem to be 'early in the process'. However, with an initial space that is 'rich' in the direction of the wanted eigenvector the initial stage of slow convergence will be absent. We will have such a situation when we continue the process for the second eigenvalue after detection of the first one.) After termination, the norms of the residuals are $\|r_u\| \approx 3.9 \cdot 10^{-8}$ and $\|r_v\| \approx 2.5 \cdot 10^{-9}$. Using only 4 extra MVs to find u more accurately (see (4.6)), we have $\|r_u\| \approx 9.3 \cdot 10^{-9}$. This is also an illustration of the situation that $\theta(u,v)$ is often more accurate than $\theta(u)$ and $\overline{\theta(v)}$: we have $|\lambda - \theta(u,v)| \approx 3.6 \cdot 10^{-15}$, while $|\lambda - \theta(u)| \approx 3.4 \cdot 10^{-10}$ and $|\lambda - \overline{\theta(v)}| \approx 1.6 \cdot 10^{-11}$. Alternating JD uses slightly more MVs than standard JD, but approximates the condition of the eigenvalue after 47 iterations up to 0.1% relative accuracy. Moreover, upon termination, the norms of both residuals are small ($\|r_u\| \approx 3.8 \cdot 10^{-8}$ and $\|r_v\| \approx 8.9 \cdot 10^{-9}$). Note the irregular convergence of the orthogonal variant of two-sided JD.



(a) (b)

FIG. 9.3: (a) The convergence histories of bi-orthogonal two-sided JD (with BiCG, dash-star), bi-orthogonal two-sided JD (GMRES, dash), orthogonal two-sided JD (GMRES, dash-dot), alternating JD (GMRES, dot), and JD (GMRES, solid) for the matrix SHERMAN4. All correction equations are solved by 5 steps of GMRES. (b) The convergence histories of bi-orthogonal two-sided JD (dash), orthogonal two-sided JD (dash-dot), and JD (solid) for the matrix SHERMAN1, as a function of the MVs. All correction equations are solved to precision ($\xi_1 = \xi_2$) 0.7 by GMRES.

As the next example, we take SHERMAN4 (size 1104, available from the Matrix Market, `http://math.nist.gov/MatrixMarket`), u_1 random, and v_1 the normalized Au_1. We solve the correction equations by 25 steps. See Figure 9.3(a). Now two-sided bi-orthogonal JD with BiCG is (also measured in MVs) the fastest method. Bi-orthogonal two-sided JD with GMRES, and orthogonal two-sided JD with GMRES use less iterations, but more MVs than standard JD. Alternating JD is somewhat slower than standard JD, but finds the two eigenvectors with $\|r_u\| \approx 2.4 \cdot 10^{-8}$ and $\|r_v\| \approx 1.5 \cdot 10^{-12}$. Also in this example, the two-sided methods approximate $\kappa(\lambda)$ well already a few steps before termination.

For Figure 9.3(b), we take a symmetric matrix, the 1000×1000 matrix SHERMAN1. The starting vectors are the same as for (a). We solve the correction equations such that the relative residuals (ξ_1 and ξ_2 in (5.1), (5.3), and (5.4)) are less than 0.7.

The convergence of the two-sided methods looks roughly linear (cf. §5, the number of MVs per iteration is also almost constant), while standard JD does not converge within 175 MVs. The history of alternating JD is the same as that of JD, since the matrix is normal.

10. Discussion and Conclusion. We have discussed an alternative approach to find eigenvalues and eigenvectors of a nonnormal matrix. Two-sided JD is a natural generalization of standard JD for nonnormal matrices. Without further demonstration, we mention that most of the techniques known in JD (such as preconditioning the correction equation, using a target, restarting, and using harmonic Ritz values) easily carry over to two-sided JD.

At the introduction of two-sided JD, we have focused on the fast convergence of the method: exact two-sided JD has asymptotically cubic convergence for simple eigenvalues of nonnormal matrices. However, in practice this might not be the most important advantage of the method. Another benefit is the fact that already *during the process*, we have approximations to *both* the left *and* and the right eigenvector. We can use this information for an estimation of the condition of the eigenvalue $\kappa(\lambda)$. This, on its turn, can be used as an error estimation

$$|\lambda - \theta| \lesssim \kappa(\lambda)\|r\|,$$

which can serve as a stopping criterion. Moreover, when we spot an eigenvalue with (possibly) a high condition, we may want to try to avoid it (using a target) when we are not interested in it, or stop the method and continue with standard JD.

During or after the process we can compare the three Rayleigh Quotients $\theta(u)$, $\overline{\theta(v)}$, $\theta(u,v)$ to check for misconvergence, that is, check to see if they converge to the same value. Moreover, from (3.1) it is clear that $\theta(u,v)$ can be more accurate ($\mathcal{O}(\delta_k\varepsilon_k)$) than $\theta(u)$ and $\overline{\theta(v)}$ ($\mathcal{O}(\delta_k)$ or $\mathcal{O}(\varepsilon_k)$), and this is confirmed by numerical experiments.

Compared with two-sided Lanczos, two-sided JD is more flexible, in the sense that we can restart with any vectors we like, and add some extra vectors to the subspaces. Two-sided JD is also more stable than two-sided Lanczos, in the sense that it can easily cope with breakdown, no look-ahead versions are necessary (see §8.3).

Of course, as compared to standard JD, two-sided JD has also disadvantages. First of all, we need the action of multiplication by A^*. Two-sided JD costs approximately twice the work per iteration compared with standard JD, and also roughly twice the storage. One could argue that by two steps of ordinary RQI (or JD) one gets the fourth degree of the error, in contrary to the third degree by one step of two-sided RQI (or JD). Ostrowski states that "from this point of view, even in the case of a non-Hermitian matrix, the use of the ordinary Rayleigh quotient iteration appears to be not only permissible but even advisable" [4, p. 472]. However, Parlett [5, Rem. 3, p. 689] criticizes this statement (in the context of dense methods).

Because of the two-sided Rayleigh Quotient and the oblique projections, the method may have difficulties with eigenvalues with a large condition, affecting the stability of the method. This can result in *loss* of accuracy in determining λ; the order remarks above have little significance if $\kappa(\lambda)$ is huge.

In conclusion, two-sided JD is an elegant alternative to standard JD and two-sided Lanczos, especially in situations where the matrix is nonnormal (but not pathetically) and when it is of interest to have approximations to the left eigenvector and condition of the eigenvalue during the process. Alternating JD may also give good results, especially if the matrix is slightly nonnormal.

Acknowledgments. The authors thank Jasper van den Eshof for interesting discussions.

REFERENCES

[1] L. Elden and V. Simoncini, *Inexact Rayleigh quotient-type methods for subspace tracking*, Tech. Report 1172, IAN-CNR, December, 1999.

[2] P. Lancaster, *A generalized Rayleigh quotient iteration for lambda-matrices*, Arch. Rational Mech. Anal., 8 (1961), pp. 309–322.

[3] A. M. Ostrowski, *On the convergence of the Rayleigh quotient iteration for the computation of the characteristic roots and vectors. III. (Generalized Rayleigh quotient characteristic roots with linear elementary divisors)*, Arch. Rational Mech. Anal., 3 (1959), pp. 325–340.

[4] ———, *On the convergence of the Rayleigh quotient iteration for the computation of the characteristic roots and vectors. V. (Usual Rayleigh quotient for non-Hermitian matrices and linear elementary divisors)*, Arch. Rational Mech. Anal., 3 (1959), pp. 472–481.

[5] B. N. Parlett, *The Rayleigh quotient iteration and some generalizations for nonnormal matrices*, Math. Comp., 28 (1974), pp. 679–693.

[6] ———, *The Symmetric Eigenvalue Problem*, Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 1998. Corrected reprint of the 1980 original.

[7] 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. International Linear Algebra Year (Toulouse, 1995).

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

[9] ———, *The Jacobi-Davidson method for eigenvalue problems and its relation with accelerated inexact Newton schemes*, in Iterative Methods in Linear Algebra, II., S. D. Margenov and P. S. Vassilevski, eds., vol. 3 of IMACS Series in Computational and Applied Mathematics, New Brunswick, NJ, U.S.A., 1996, IMACS, pp. 377–389. Proceedings of the Second IMACS International Symposium on Iterative Methods in Linear Algebra, June 17-20, 1995, Blagoevgrad.

[10] P. Smit and M. H. C. Paardekooper, *The effects of inexact solvers in algorithms for symmetric eigenvalue problems*, Linear Algebra Appl., 287 (1999), pp. 337–357. Special issue celebrating the 60th birthday of Ludwig Elsner.

[11] A. Stathopoulos and Y. Saad, *Restarting techniques for the (Jacobi-)Davidson symmetric eigenvalue methods*, Electron. Trans. Numer. Anal., 7 (1998), pp. 163–181 (electronic). Large scale eigenvalue problems (Argonne, IL, 1997).

[12] G. W. Stewart and J.-g. Sun, *Matrix Perturbation Theory*, Academic Press, San Diego, California, 1990.

[13] J. van den Eshof, *The convergence of Jacobi-Davidson for Hermitian eigenproblems*, Preprint 1165, Dep. Math., University Utrecht, Utrecht, the Netherlands, November 2000.

[14] A. van der Sluis and H. A. van der Vorst, *The convergence behavior of Ritz values in the presence of close eigenvalues*, Linear Algebra Appl., 88/89 (1987), pp. 651–694.