THE JACOBI-DAVIDSON METHOD FOR EIGENVALUE PROBLEMS AND ITS RELATION WITH ACCELERATED INEXACT NEWTON SCHEMES

GERARD L.G. SLEIJPEN* AND HENK A. VAN DER VORST*

Abstract. We discuss a new method for the iterative computation of a portion of the spectrum of a large sparse matrix. The matrix may be complex and non-normal. The method also delivers the Schur vectors associated with the computed eigenvalues. The eigenvectors can easily be computed from the Schur vectors, and for stability reasons we prefer the approach with Schur vectors. The method is based on the recently introduced Jacobi-Davidson algorithm [16]. This method improves the Davidson method and its generalizations. We also show how the Davidson's methods, including the new one, can be viewed as accelerated inexact Newton schemes.

Keywords: Eigenvalues and eigenvectors, Davidson's method, QR-algorithm.
AMS(MOS) subject classification: 65F15, 65N25.

1. Introduction. We consider variants of Davidson's method for the iterative computation of one or more eigenvalues and their corresponding eigenvectors of an $n \times n$ matrix $\mathbf{A}$. The original Davidson method [3], for real normal matrices A, may be viewed as an accelerated Gauss-Jacobi method, and the success of the method seems to depend quite heavily on diagonal dominance of $\mathbf{A}$ [3, 4, 19].

In the hope to enlarge the scope of the method, one has investigated the effect of replacing the diagonal preconditioner by other, more general, ones. Recent convergence results as well as numerical experiments with these generalized Davidson methods are reported in [2, 9, 10, 11]. Unfortunately, preconditioners that represent the eigenproblem well may lead to slow convergence or even to stagnation. In [16] an explanation for this unsatisfactory situation is given. Further exploiting ideas of Jacobi [8] has led to a new robust method with superior convergence properties, for non-diagonally dominant, non-normal, complex, matrices as well: this *Jacobi-Davidson* [16] method takes advantage of the quality of the preconditioner and is not sensitive to the effects of rounding errors. Moreover, the Jacobi-Davidson method usually converges faster with better preconditioners. The efficiency per step of this new method is comparable to a step of the Davidson method if the same way of preconditioning is used.

The Jacobi-Davidson method and a discussion of its relation with Davidson's method can be found in Sect. 2. A convenient way of incorporating preconditioners is given in Sect. 3. In Sect. 4, we argue that the new method can be viewed as an improvement on methods such as Davidson's, Arnoldi's, and (accelerated inexact) Shift-and-Invert. Most of the material in these sections has been taken from [16]. We have chosen to highlight part of the material in [16] here, since that paper has not appeared yet.

The basic scheme of the Jacobi-Davidson method, given in [16], needs some modification for determining a portion of the spectrum with associated eigenvectors. Our approach in Sect. 5 is based on deflation: by projecting on the orthogonal complement of the subspace spanned by detected eigenvectors we compute a partial Schur form. The convergence behaviour of this "QR-variant" of the Jacobi-Davidson method exhibits interesting parallels to the convergence behaviour of standard QR for dense

* Mathematical Institute, Utrecht University, Budapestlaan 6, Utrecht, the Netherlands, e-mail: sleijpen@math.ruu.nl, vorst@math.ruu.nl

matrices. We have also incorporated a restart strategy in our Jacobi-Davidson QR-algorithm, ALG. 2.

Davidson [5] suggests that the Davidson-Liu variant of his algorithm may be interpreted as a Newton scheme, and this has been used as an argument to explain its fast convergence. In Sect. 6 we show that this interpretation is correct and that the Jacobi-Davidson method can also be interpreted as a Newton scheme. This interpretation leads to a better understanding of the difference between the two methods.

We conclude, in Sect. 6, with a simple but illustrative numerical example. For more complicated applications in Chemistry and Plasma Physics modeling, see [1, 18].

2. Jacobi-Davidson as an improved Davidson method. We will allow the matrix $\mathbf{A}$ to be complex and non-normal.

2.1. Davidson's method. The main idea behind Davidson's method is the following one. Suppose we have some subspace $\mathcal{K}$ of dimension k, over which the projected matrix $\mathbf{A}$ has a Ritz value θ_k and a corresponding Ritz vector $\mathbf{u}_k$. Let us assume that an orthogonal basis for $\mathcal{K}$ is given by the vectors $\mathbf{v}_1$, $\mathbf{v}_2$, …, $\mathbf{v}_k$.

Quite naturally the question arises how to expand the subspace in order to find a successful update for $\mathbf{u}_k$. To that end we compute the defect $\mathbf{r} = \mathbf{A}\mathbf{u}_k - \theta_k \mathbf{u}_k$. Then Davidson, in his original paper [3], suggests to compute $\widetilde{\mathbf{t}}$ from $(\mathbf{D}_A - \theta_k \mathbf{I})\widetilde{\mathbf{t}} = \mathbf{r}$, where $\mathbf{D}_A$ is the diagonal of the matrix $\mathbf{A}$. The vector $\widetilde{\mathbf{t}}$ is made orthogonal to $\mathcal{K}$, and the resulting vector is chosen as the new $\mathbf{v}_{k+1}$ by which $\mathcal{K}$ is expanded.

It has been reported that this method can be quite successful in finding the largest eigenvalues in absolute value of strongly diagonally dominant matrices. The matrix $(\mathbf{D}_A - \theta_k \mathbf{I})^{-1}$ can be viewed as a preconditioner for the vector $\mathbf{r}$. It is tempting to see the preconditioner also as an approximation for $(\mathbf{A} - \theta_k \mathbf{I})^{-1}$, and, indeed, this approach has been followed for the construction of more complicated preconditioners (*generalized Davidson methods*, see, e.g., [2, 9, 11]). However, as is pointed out in [15], this is a wrong interpretation, since $(\mathbf{A} - \theta_k \mathbf{I})^{-1}$ maps $\mathbf{r}$ onto $\mathbf{u}_k$, and hence it would not lead to an expansion of our search space[1]. Therefore the approximation should not be too accurate [15, 2].

2.2. The Jacobi-Davidson method. For a better understanding and for finding a way to avoid stagnation, we concentrate on the kth approximation $\mathbf{u}_k$ of the eigenvector $\mathbf{x}$. We assume $\mathbf{u}_k$ to be normalized w.r.t. the Euclidean norm: $\|\mathbf{u}_k\| = 1$. Observe that the residual $\mathbf{r} = \mathbf{A}\mathbf{u}_k - \theta_k \mathbf{u}_k$ is orthogonal to $\mathbf{u}_k$ because $\theta_k = \mathbf{u}_k^* \mathbf{A}\mathbf{u}_k$ is the Ritz value associated with $\mathbf{u}_k$. Now, following a suggestion of Jacobi [8], we project the eigenproblem $\mathbf{A}\mathbf{x} = \lambda \mathbf{x}$ on $\mathrm{span}(\mathbf{u}_k)$, and on its orthogonal complement. This leads to two coupled equations for λ and the complement $\mathbf{z}$ of $\mathbf{x}$ orthogonal to $\mathbf{u}_k$:

$$\lambda = \mathbf{u}_k^* \mathbf{A}(\mathbf{u}_k + \mathbf{z}), \tag{2.1}$$

$$\mathbf{z} \perp \mathbf{u}_k \quad \text{and} \quad (\mathbf{I} - \mathbf{u}_k \mathbf{u}_k^*)(\mathbf{A} - \lambda \mathbf{I})(\mathbf{I} - \mathbf{u}_k \mathbf{u}_k^*)\,\mathbf{z} = -\mathbf{r}. \tag{2.2}$$

For this derivation, we have used the fact that $(\mathbf{I} - \mathbf{u}_k \mathbf{u}_k^*)(\mathbf{A} - \lambda \mathbf{I})\mathbf{u}_k = \mathbf{r}$. Since λ is unknown, we cannot compute the optimal update $\mathbf{z}$ for $\mathbf{u}_k$ from (2.2). However, it

[1] Any progress in this case may be attributed to the effects of rounding errors

ALGORITHM 1. *The Jacobi-Davidson Method*

1. **Start:** Choose a $k_{max} > 0$ and an initial non-trivial vector $\mathbf{v}$.
   Set $H = [\,]$, $\mathbf{V} = [\,]$, $\mathbf{W} = [\,]$.
2. **Iterate until $\dim(\mathbf{V}) = k_{max}$:**
   a. Orthogonalize $\mathbf{v}$ against $\mathbf{V}$ via modified Gram-Schmidt, normalize the resulting $\mathbf{v}$: $\mathbf{v} \leftarrow \mathbf{v}/\|\mathbf{v}\|$. Compute $\mathbf{w} \leftarrow \mathbf{A}\mathbf{v}$.
   b. $H \leftarrow \begin{bmatrix} H & \mathbf{V}^*\mathbf{w} \\ \mathbf{v}^*\mathbf{W} & \mathbf{v}^*\mathbf{w} \end{bmatrix}$, $\mathbf{V} \leftarrow [\,\mathbf{V}\,|\,\mathbf{v}\,]$, $\mathbf{W} \leftarrow [\,\mathbf{W}\,|\,\mathbf{w}\,]$.
   c. Compute the largest eigenpair (θ, y) of H (with $\|y\| = 1$).
   d. Compute the Ritz vector $\mathbf{u} \leftarrow \mathbf{V}y$, $\tilde{\mathbf{u}} \leftarrow \mathbf{A}\mathbf{u}$ $(= \mathbf{W}y)$, and the associated residual vector $\mathbf{r} \leftarrow \tilde{\mathbf{u}} - \theta\mathbf{u}$.
   e. Test for convergence. Stop if satisfied.
   f. Solve (approximately) $\mathbf{v} \perp \mathbf{u}$
   $$(\mathbf{I} - \mathbf{u}\,\mathbf{u}^*)\,(\mathbf{A} - \theta\,\mathbf{I})\,(\mathbf{I} - \mathbf{u}\,\mathbf{u}^*)\,\mathbf{v} = -\mathbf{r}.$$
3. **Restart:** Set $H \leftarrow [\,\theta\,]$, $\mathbf{V} \leftarrow [\,\mathbf{u}\,]$, $\mathbf{W} \leftarrow [\,\tilde{\mathbf{u}}\,]$, and goto 2.

seems reasonable to replace λ by the current approximation θ_k: we expect the exact solution $\mathbf{t}$ of the equation

$$\mathbf{t} \perp \mathbf{u}_k \quad \text{and} \quad (\mathbf{I} - \mathbf{u}_k\mathbf{u}_k^*)(\mathbf{A} - \theta_k\mathbf{I})(\mathbf{I} - \mathbf{u}_k\mathbf{u}_k^*)\,\mathbf{t} = -\mathbf{r} \tag{2.3}$$

to be a good correction for $\mathbf{u}_k$.

In Davidson methods, the computational effort is directed towards finding good approximate solutions of the equation

$$(\mathbf{A} - \theta_k\mathbf{I})\mathbf{t} = -\mathbf{r}, \tag{2.4}$$

whereas we now see that we should concentrate on computing an approximate solution $\tilde{\mathbf{t}}$ of the *correction equation* (2.3). Subsequently, we make $\tilde{\mathbf{t}}$ orthogonal to the *search space* $\mathcal{K}$ which defines the new orthonormal basis vector $\mathbf{v}_{k+1}$.

The algorithm for the Jacobi-Davidson method then becomes as in ALG. 1, where we have skipped indices for variables that overwrite old values (indicated by $\leftarrow$) in an iteration step, e.g., $\mathbf{u}$ instead of $\mathbf{u}_k$. We compute vectors $\mathbf{v}_1, \ldots, \mathbf{v}_k$ that form an orthonormal basis of the search space. For efficiency, we compute and store also $\mathbf{w}_j = \mathbf{A}\mathbf{v}_j$. The Ritz value θ_k is an eigenvalue of the "small" $k \times k$ matrix $H_k = \mathbf{V}_k^*\mathbf{A}\mathbf{V}_k = \mathbf{V}_k^*\mathbf{W}_k$, where $\mathbf{V}_k = [\,\mathbf{v}_1|\ldots|\,\mathbf{v}_k]$ and $\mathbf{W}_k = [\,\mathbf{w}_1|\ldots|\,\mathbf{w}_k]$. The integer $\dim(\mathbf{V}_k)$ is the number of column vectors of $\mathbf{V}_k$. The size $\|\,\mathbf{r}\,\|$ of the residual can be monitored to detect convergence, but other strategies may be considered as well.

3. Preconditioning. For general systems it will be unattractive to solve the correction equation (2.3) exactly and some iterative scheme can be considered. Formulation (2.3) is not convenient for incorporating preconditioners $\mathbf{M}$ for $\mathbf{A} - \theta_k\mathbf{I}$, $\mathbf{M} \approx \mathbf{A} - \theta_k\mathbf{I}$. A more appropriate formulation follows from the observation that the solution $\mathbf{t}$ of (2.3) satisfies

$$(\mathbf{A} - \theta_k\mathbf{I})\,\mathbf{t} = \alpha\,\mathbf{u}_k - \mathbf{r}, \quad \text{with } \alpha \text{ such that } \mathbf{t} \perp \mathbf{u}_k. \tag{3.1}$$

This suggests to compute an approximation $\widetilde{\mathbf{t}}$ for $\mathbf{t}$ as

$$\widetilde{\mathbf{t}} = \alpha\, \mathbf{M}^{-1}\mathbf{u}_k - \mathbf{M}^{-1}\mathbf{r} \tag{3.2}$$

again with α such that $\widetilde{\mathbf{t}} \perp \mathbf{u}_k$:

$$\alpha = \frac{\mathbf{u}_k^* \mathbf{M}^{-1}\mathbf{r}}{\mathbf{u}_k^* \mathbf{M}^{-1}\mathbf{u}_k}. \tag{3.3}$$

If a preconditioned iterative method is used to solve the correction equation (2.3) then, in each step of the linear solver, the solution of the "preconditioning equation" can be computed by formulas as (3.2-3). In this approach, we have, except for the first iteration step, to solve only one system involving $\mathbf{M}$ in each iteration step. The inner product $\mathbf{u}_k^*\mathbf{M}^{-1}\mathbf{u}_k$, to be computed once, can also be used in all steps of the iteration process for (2.3).

An iterative method itself can be considered to be a preconditioning; $\mathbf{M}$ is not explicitly known then. However, there is a huge difference between applying the iterative method to the projected system (2.3) directly and solving (2.3) according to (3.2-3) where the method would be applied to the matrix $\mathbf{A} - \theta_k\mathbf{I}$. If, for instance, θ_k is an accurate approximation of λ, the matrix $\mathbf{A} - \theta_k\mathbf{I}$ will be almost singular while that will not be the case for the projected operator in (2.3): faster and more stable convergence may be expected from the direct approach.

If $\mathbf{M}$ is a good approximation for $\mathbf{A} - \theta_k\mathbf{I}$ then $\mathbf{M}^{-1}\mathbf{r} \approx \mathbf{u}_k$ and, if $\mathbf{u}_k \approx \mathbf{x}$, the angle between $\mathbf{M}^{-1}\mathbf{u}_k$ and $\mathbf{u}_k$ may also be very small, so that (3.2-3) will not lead to an effective subspace expansion; the direct approach seems to be less sensitive to rounding errors.

4. Related methods. Equation (3.2) leads to several interesting approaches:

1. If we approximate $\mathbf{t}$ simply by $-\mathbf{r}$, ($\alpha = 0$ and $\mathbf{M} = \mathbf{I}$) then we obtain formally the same results as with Arnoldi's method; $\tilde{\mathbf{t}}$ will not be orthogonal to $\mathbf{u}_k$. Arnoldi can be viewed as a Jacobi-Davidson method using a 1-step Krylov method without preconditioner in the "inner loop" (the loop that solves the correction equation).
2. With $\alpha = 0$ we obtain a Davidson method with preconditioner $\mathbf{M}$. Also, in this case $\tilde{\mathbf{t}}$ will not be orthogonal to $\mathbf{u}_k$.
3. For $\mathbf{M}$ an explicit preconditioner, Olsen et al. [12] proposed to choose α as in (3.3).
4. Any approximate solution of (2.3) obtained by some preconditioned iterative method can be represented formally by (3.2-3): the choice defined in (3.3) gives a Jacobi-Davidson method.
5. If $\mathbf{M} = \mathbf{A} - \theta_k\mathbf{I}$, then (3.2) reduces to $\mathbf{t} = \alpha\,(\mathbf{A} - \theta_k\mathbf{I})^{-1}\mathbf{u}_k - \mathbf{u}_k$. Since $\mathbf{t}$ is made orthogonal to $\mathbf{u}_k$ afterwards, this choice is equivalent with solving $\mathbf{t}$ from

$$(\mathbf{A} - \theta_k\mathbf{I})\,\mathbf{t} = \mathbf{u}_k. \tag{4.1}$$

In this case the method is mathematically equivalent with accelerated Shift-and-Invert iteration with optimal shift. For symmetric $\mathbf{A}$ this is the accelerated Inverse Rayleigh Quotient method, which converges cubically [13]. In the unsymmetric case we have quadratic convergence [13].

In view of the speed of convergence of Shift-and-Invert methods it may hardly be worthwhile to accelerate them in a "Davidson" manner: the overhead is significant and the gains may only be minor.

6. If $\mathbf{M} \neq \mathbf{A} - \theta_k \mathbf{I}$, then with $\widetilde{\mathbf{t}} = \mathbf{M}^{-1}\mathbf{u}_k$ we obtain an inexact Shift-and-Invert method, with "Davidson" subspace acceleration.

The first, second and the fifth approach are well-known, and the question arises whether we may gain anything by the third method or the fourth or sixth one.

Heuristic arguments in [16], supported by experimental results, indicate that the original Davidson method works well in situations where $\widetilde{\mathbf{t}}$ does not have a strong component in the direction of $\mathbf{u}_k$. Shift-and-Invert approaches work well if the component in the direction of $\mathbf{x}$ in $\mathbf{u}_k$ is strongly increased. However, in this case this component may dominate so strongly, when we have a good preconditioner, that it prevents us to reconstruct, in finite precision arithmetic, a relevant orthogonal expansion for the search space (cf. [16]: Sect. 4). In this respect the Jacobi-Davidson is a compromise between the Davidson method and the accelerated inexact Shift-and-Invert method, since the factor α properly controls the influence of $\mathbf{u}_k$ and makes sure that we construct the orthogonal expansion of the subspace correctly: the vectors in the linear combination in (3.2) will be of comparable size and will not be almost dependent (cf. [16], Sect. 4). In this respect, Jacobi-Davidson offers the best of two worlds, and this will be illustrated by our numerical example.

5. Improvements on the basic Jacobi-Davidson algorithm. In this section we briefly discuss strategies for restart and for the computation of a portion of the spectrum.

5.1. Restart strategies. In the Jacobi-Davidson algorithm (Alg. 1) we included a familiar restart, simply by taking the Ritz vector $\mathbf{u}_k$, $k = k_{max}$, at the end of a cycle as a new initial guess. However, then the process may construct a new search space that has considerable overlap with the previous one; this phenomenon is well known for the restarted power method and restarted Arnoldi without deflation, and this may lead to a reduced speed of convergence or even to stagnation. After such a simple restart, we may expect that the process will construct also vectors with significant components in directions of eigenvectors associated with eigenvalues close to the wanted eigenvalue. And this is just the kind of information that is discarded at the restart. This suggests a strategy to retain ℓ Ritz vectors, for some $\ell > 1$, associated with the Ritz values closest to this eigenvalue as well including the Ritz vector $\mathbf{u}_k$, $k = k_{max}$, that is the approximation for the desired eigenvector. In Alg. 1, these would be the ℓ largest Ritz values and we return to step 2 (cf. Alg. 1), with the ℓ-dimensional subspace spanned by the associated Ritz vectors.

Similar restart or assembly strategies have been proposed for methods as Shift-and-Invert Arnoldi. Since, in contrast to Jacobi-Davidson, these methods find their approximate eigenvectors in Krylov subspaces, their restart strategy requires more care [17]: the new space should not only be close to the span of Ritz vectors associated with nearby Ritz values, but, for efficiency reasons, it should also be a Krylov subspace.

5.2. Computing a partial Schur form. We have formulated Jacobi-Davidson for the largest eigenvalue (cf. Alg. 1). Of course, the algorithm can be adapted for other eigenvalues as well. For instance, selecting the Ritz value nearest to some "target value" $\mu \in \mathbb{C}$ rather than the largest one, leads to the eigenvalue nearest to μ. For such "interior" eigenvalues, the use of harmonic Ritz values rather than Ritz values is recommended in [9] and [16]: Sect. 5. For ease of presentation and since, as we will see below, the advantages of harmonic Ritz are less obvious for the computation of several eigenvalues, we do not pursue this approach here.

Jacobi-Davidson is very attractive for computing a portion of the spectrum: in its search for one eigenvector, the process builds a search space with strong components also in the direction of other eigenvectors associated with nearby eigenvalues. If an eigenpair has been identified, then the eigenproblem can be deflated and the search for other eigenvalues can be continued, using the old search space minus the computed eigenvector as an initial search space for the next pair. As in the standard QR algorithm, and for similar reasons, the process will compute next eigenpairs increasingly faster. As an additional advantage, the deflation technique enables to find eigenvalues according to their multiplicity. If the preconditioner is designed to help and find eigenvalues close to some μ, then the speed of convergence may slow down if the distance between μ and the target eigenvalues grows.

For stability reasons, we compute Schur vectors rather than eigenvectors: for the matrix $\mathbf{U}$ with orthonormal column vectors (*Schur vectors*) $\mathbf{u}_1, \ldots, \mathbf{u}_m$, we have that $\mathbf{AU} = \mathbf{U}S$ for some $m \times m$ upper triangular matrix S (*a Schur matrix*) with on its diagonal the eigenvalues of $\mathbf{A}$ near(est) to some target value $\mu \in \mathbb{C}$, say. Also for stability reasons, we suggest to use orthogonal projections ($\mathbf{I} - \mathbf{U}\,\mathbf{U}^*$, etc.) rather than standard deflation ($\mathbf{A} - \lambda \mathbf{u}\,\mathbf{u}^* - \ldots$). These projections should be performed by modified Gram-Schmidt; in ALG. 2 we used the notation $\mathrm{Mod}(\mathbf{I} - \mathbf{U}\,\mathbf{U}^*)$ to indicate that we apply the projection $(\mathbf{I} - \mathbf{U}\,\mathbf{U}^*)$ with special care.

Harmonic Ritz values help to avoid "tracking" temporarily a wrong sequence of approximate eigenvalues. However, a "wrong" choice in certain steps tends to enrich the search space with components of eigenvectors that will be of interest later on, and this may speed up the convergence towards these eigenvectors in a later stage of the process (after deflation). Therefore, when computing a portion of the spectrum, it is less obvious whether harmonic Ritz values are of advantage.

ALG. 2 leads to the m eigenvalues $\lambda_1, \ldots, \lambda_m$ of the matrix $\mathbf{A}$, near(est[2]) to the target value μ, and their associated Schur vectors. Implicitly, it also computes the Schur matrix S. Replacing "$\lambda_m = \theta$" by "$S \leftarrow \begin{bmatrix} S & \mathbf{U}^*\widetilde{\mathbf{u}} \\ 0^* & \theta \end{bmatrix}$" would give S explicitly. The algorithm follows an assembly strategy as suggested in Sect. 5.1. Observe that the QR-algorithm can be used to compute at relative low costs the Schur form of the "small" $k \times k$ matrix $H = H_k$. We order the Ritz values on the diagonal of the Schur matrix R of H_k such that the distance to μ increases ($|R_{11} - \mu| \leq |R_{22} - \mu| \leq \ldots$). This facilitates the identification of the smaller search space for assembly when the search space becomes too large: the first ℓ columns R will define the ℓ Ritz vectors not converged yet with Ritz values nearest to μ. For more details, we refer to [6].

6. Jacobi-Davidson as an accelerated Newton scheme. The eigenvalue problem is nonlinear: for almost any scaling vector $\widetilde{\mathbf{u}}$ and $\mathbf{w}$, the eigenvector $\mathbf{x}$, scaled such that $\widetilde{\mathbf{u}}^*\mathbf{x} = 1$, is the solution of the equation

$$\mathbf{F}(\mathbf{x}) = 0 \quad \text{where} \quad \mathbf{F}(\mathbf{u}) = \mathbf{Au} - \theta\mathbf{u} \quad \text{with} \quad \theta = \theta(\mathbf{u}) = \frac{\mathbf{w}^*\mathbf{Au}}{\mathbf{w}^*\mathbf{u}}.$$

[2] Not all eigenvalues close to μ may be detected; occasionally, one may be missed that would have appeared later on. A similar phenomenon occurs in QR: there, although the first m detected eigenvalues in general tend to be the m smallest, they may appear in a slightly different order.

ALGORITHM 2. *The Jacobi-Davidson QR algorithm.*
The algorithm produces m Schur vectors $\mathbf{U} = [\mathbf{u}_1 \,|\, \ldots \,|\, \mathbf{u}_m]$ with eigenvalues $\lambda_1, \ldots, \lambda_m$ that are nearest to the "target value" $\mu \in \mathbb{C}$ with residual accuracy $tol > 0$.

1. **Start:** Choose ℓ and k_{max} such that $m < \ell < k_{max}$.
   Choose an initial non-trivial vector $\mathbf{v}$.
   Set $\mathbf{U} = [\,]$, $H = [\,]$, $\mathbf{V} = [\,]$, $\mathbf{W} = [\,]$.
2. **Iterate until $\mathbf{dim}(\mathbf{V}) = k_{max}$:**
   a. Orthogonalize $\mathbf{v}$ against $\mathbf{V}$ by modified Gram-Schmidt, normalize the resulting vector $\mathbf{v} \leftarrow \mathbf{v}/\|\mathbf{v}\|$. Compute $\mathbf{w} \leftarrow \mathbf{A}\mathbf{v}$.
   b. $H \leftarrow \begin{bmatrix} H & \mathbf{V}^*\mathbf{w} \\ \mathbf{v}^*\mathbf{W} & \mathbf{v}^*\mathbf{w} \end{bmatrix}$, $\mathbf{V} \leftarrow [\,\mathbf{V} \,|\, \mathbf{v}\,]$, $\mathbf{W} \leftarrow [\,\mathbf{W} \,|\, \mathbf{w}\,]$.
   c. Compute the Schur form of H with ordered diagonal:
      $H = QRQ^*$ where $Q^*Q = I$ and R upper triangular with
      $|R_{11} - \mu| \leq |R_{22} - \mu| \leq$.
   d. Set $\theta \leftarrow R_{11}$, $y \leftarrow Qe_1$.
      Compute the Ritz vector $\mathbf{u} \leftarrow \mathbf{V}y$, $\tilde{\mathbf{u}} \leftarrow \mathbf{A}\mathbf{u}$ $(= \mathbf{W}y)$,
      and the residual $\mathbf{r} \leftarrow \mathrm{Mod}(\mathbf{I} - \mathbf{U}\mathbf{U}^*)(\tilde{\mathbf{u}} - \theta\mathbf{u})$.
   e. **If** $\|\,\mathbf{r}\,\| < tol$:
      $\lambda_m = \theta$, $\mathbf{U} \leftarrow [\,\mathbf{U} \,|\, \mathbf{u}\,]$. <u>**Stop**</u> if $\dim(\mathbf{U}) = m$.
      With $J = [\,e_2 \,|\, \ldots \,|\, e_k\,]$
      set $H \leftarrow J^*RJ$, $\mathbf{V} \leftarrow \mathbf{V}(QJ)$, $\mathbf{W} \leftarrow \mathbf{W}(QJ)$,
      $R \leftarrow H$, $Q \leftarrow I$, and goto 2.d.
   f. Solve (approximately) $\mathbf{v} \perp \widetilde{\mathbf{U}} = [\,\mathbf{U} \,|\, \mathbf{u}\,]$
      $$\mathrm{Mod}(\mathbf{I} - \widetilde{\mathbf{U}}\widetilde{\mathbf{U}}^*)\,(\mathbf{A} - \theta\,\mathbf{I})\,\mathrm{Mod}(\mathbf{I} - \widetilde{\mathbf{U}}\widetilde{\mathbf{U}}^*)\,\mathbf{v} = -\mathbf{r}.$$
3. **Assembly:** With $J = [\,e_1 \,|\, \ldots \,|\, e_\ell\,]$,
   set $H \leftarrow J^*RJ$, $\mathbf{V} \leftarrow \mathbf{V}(QJ)$, $\mathbf{W} \leftarrow \mathbf{W}(QJ)$, and goto 2.

The function $\mathbf{F}$ is nonlinear and maps the hyper-plane $\{\mathbf{u} \mid \tilde{\mathbf{u}}^*\mathbf{u} = 1\}$ to the hyper-space $\mathbf{w}^\perp$. In particular, residuals $\mathbf{r} := \mathbf{F}(\mathbf{u})$ are orthogonal to $\mathbf{w}$.

If $\mathbf{u}_k$ approximates the eigenvector $\mathbf{x}$ then the next Newton approximate $\mathbf{u}_{k+1}$ is defined by

$$\mathbf{u}_{k+1} = \mathbf{u}_k + \mathbf{t}, \quad \text{where} \quad \mathbf{t} \perp \tilde{\mathbf{u}} \quad \text{satisfies} \quad \left(\frac{\partial\,\mathbf{F}}{\partial\,\mathbf{u}}\bigg|_{\mathbf{u}}\right)\mathbf{t} = -\mathbf{r} = -\mathbf{F}(\mathbf{u}_k).$$

The Jacobian of $\mathbf{F}$ acts on $\tilde{\mathbf{u}}^\perp$, and is given by

$$\left(\frac{\partial\,\mathbf{F}}{\partial\,\mathbf{u}}\bigg|_{\mathbf{u}_k}\right)\mathbf{t} = \left(\mathbf{I} - \frac{\mathbf{u}_k\mathbf{w}^*}{\mathbf{w}^*\mathbf{u}_k}\right)(\mathbf{A} - \theta_k\mathbf{I})\,\mathbf{t} \quad \text{for} \quad \mathbf{t} \perp \tilde{\mathbf{u}}.$$

Hence, the correction equation of the Newton step reads as

$$\mathbf{t} \perp \tilde{\mathbf{u}} \quad \text{and} \quad \left(\mathbf{I} - \frac{\mathbf{u}_k\mathbf{w}^*}{\mathbf{w}^*\mathbf{u}_k}\right)(\mathbf{A} - \theta_k\mathbf{I})\,\mathbf{t} = -\mathbf{r}. \tag{6.1}$$

Keeping $\tilde{\mathbf{u}}$ and $\mathbf{w}$ fixed throughout the Newton iteration produces $\mathbf{u}_k$ that converge asymptotically quadratically towards $\mathbf{x}$ if $\mathbf{u}_1$, with $\tilde{\mathbf{u}}^*\mathbf{u}_1 = 1$, is sufficiently close to $\mathbf{x}$

and $\mathbf{w}^*\mathbf{x} \neq 0$. However, since the Newton error contracts quadratically if $\mathbf{u}_k$ is close enough to $\mathbf{x}$, the non-stationary choice $\widetilde{\mathbf{u}} = \mathbf{u}_k$ and $\mathbf{w} = \mathbf{u}_k$ also gives asymptotic convergence[3]. For this choice, the Newton correction equation is precisely the one of Jacobi-Davidson, and the result on the speed of convergence is in line with our observation in Sect. 4. Apparently, Jacobi-Davidson is a Newton scheme, accelerated by "Davidson" subspaces. The acceleration is important for several reasons. It leads asymptotically to super-quadratic convergence, but, more importantly, it accelerates the "pre-quadratic" phase: the process can be steered towards a wanted eigenpair and the angle between the eigenvector and the search space will be smaller for larger search spaces.

In general it is expensive to solve the correction equation exactly. With preconditioner $\mathbf{M}$ (a nonsingular approximation of $\mathbf{A} - \theta_k\mathbf{I}$) the "inexact" equation appears by replacing $\mathbf{A} - \theta_k\mathbf{I}$ in (6.1) by $\mathbf{M}$, with solution $\widetilde{\mathbf{t}}$ given by (cf. (3.3) and (3.2)),

$$\widetilde{\mathbf{t}} = \alpha\,\mathbf{M}^{-1}\mathbf{u}_k - \mathbf{M}^{-1}\mathbf{r} \quad \text{with} \quad \alpha = \frac{\widetilde{\mathbf{u}}^*\mathbf{M}^{-1}\mathbf{r}}{\widetilde{\mathbf{u}}^*\mathbf{M}^{-1}\mathbf{u}_k}. \tag{6.2}$$

Since $\mathbf{r} \perp \mathbf{w}$, we have that $\mathbf{M}^{-1}\mathbf{r} \perp \widetilde{\mathbf{u}}$ whenever $\mathbf{M}$ maps $\widetilde{\mathbf{u}}^\perp$ onto $\mathbf{w}^\perp$. Then the correction vector $\widetilde{\mathbf{t}}$ is an expansion vector of a Davidson method: $\widetilde{\mathbf{t}} = \mathbf{M}^{-1}\mathbf{r}$ (cf. [2]). We have this bijectivity if, for instance, $\mathbf{M}$ is the diagonal of $\mathbf{A} - \theta_k\mathbf{I}$ (Davidson's original choice in [3]) and $\widetilde{\mathbf{u}} = \mathbf{w}$ is some standard basis vector e_ν.

Davidson's original method can be viewed as an accelerated inexact Newton scheme, explaining its fast convergence for strongly diagonally dominant matrices (see also [5]). Apparently, both inexact Jacobi-Davidson and Davidson's method are accelerated inexact Newton schemes. However, there is an essential difference: Davidson's method relies implicitly on skew projections: $e_\nu^*\mathbf{u}_k$ can be very small. As argued in Sect. 4, this can deteriorate the speed of convergence significantly. For any nonsingular non-diagonal $\mathbf{M}$ and any non-trivial $\widetilde{\mathbf{u}}$ there is a vector $\mathbf{w}$ such that $\mathbf{M}$ maps $\widetilde{\mathbf{u}}^\perp$ onto $\mathbf{w}^\perp$, and the scaling vectors can be chosen optimally. But again, the resulting projection may be very skew specially in the slow "non-quadratic" phase of the process. For instance, if $\widetilde{\mathbf{u}}$ is the orthogonal projection of $\mathbf{u}_k$ onto $\mathbf{M}^*(\mathbf{V}_k)$ and $\mathbf{w} = \mathbf{M}^{-*}\widetilde{\mathbf{u}}$, then $\widetilde{\mathbf{u}}^\perp$ is mapped onto $\mathbf{w}^\perp$, and the angles between $\mathbf{u}_k$ and $\mathbf{M}^*(\mathbf{V}_k)$, and between $\widetilde{\mathbf{u}}$ and $\mathbf{M}^{-1}\mathbf{u}_k$, play a role.

7. A numerical example. The simple example that we will present here should be seen only as an illustration of the new approach.

In this example, taken from [16], we compare experimentally the performance of generalized Davidson, Jacobi-Davidson, and the accelerated inexact Shift-and-Invert variant (**SI**) of approach 6 in Sect. 4, i.e., we expand our search space by the orthogonal complement of the approximate solution $\widetilde{\mathbf{t}}$ of (2.4), (2.2), and (4.1), respectively. We have solved these equations approximately by m steps of preconditioned GMRES with 0 as an initial guess. The preconditioner $\mathbf{M}$ is kept fixed throughout the iteration process. For the systems (2.4) and (4.1) $\mathbf{M}$ is used, while the projected $(\mathbf{I} - \mathbf{u}_k\mathbf{u}_k^*)\,\mathbf{M}\,(\mathbf{I} - \mathbf{u}_k\mathbf{u}_k^*)$ is used for (2.2) (cf. Sect. 3). Since we are interested in the absolute smallest eigenvalue, we take for θ_k the eigenvalue of $H_k = \mathbf{V}_k^*\mathbf{A}\mathbf{V}_k$ of smallest absolute value.

---

[3] The resulting scheme with these non-stationary scaling vectors appears also if Newton's method is applied to the map $(\theta, \mathbf{u}) \rightarrow \mathbf{A}\mathbf{u} - \theta\mathbf{u}$ from $\mathbb{C} \times \{\mathbf{u} \mid \|\mathbf{u}\| = 1\}$ to $\mathbb{C}^n$.

FIG. 7.1. *Using 5 steps of preconditioned* GMRES

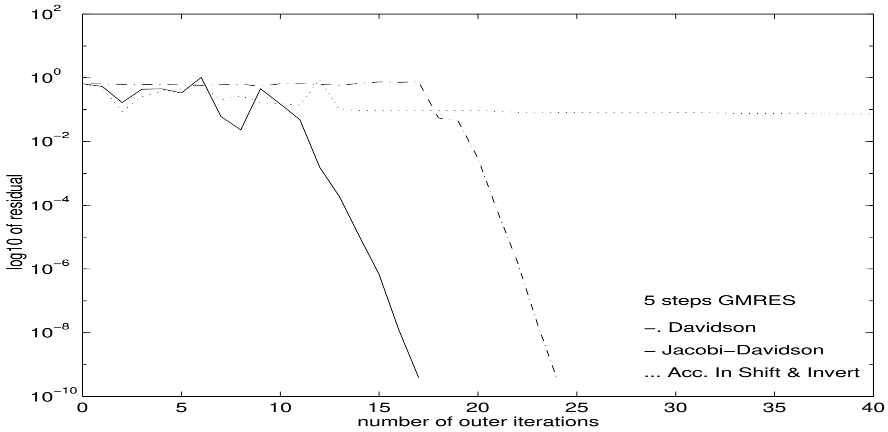


FIG. 7.2. *Using 10 steps precond.* GMRES

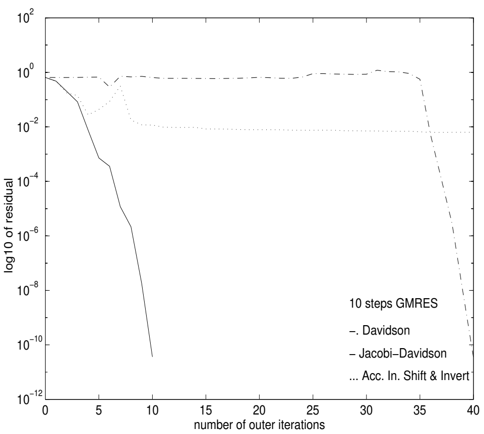
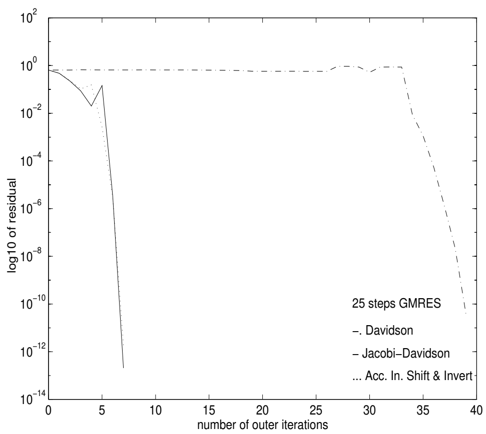


FIG. 7.3. *Using 25 steps precond.* GMRES



We have applied the three methods, Davidson, Jacobi-Davidson and **SI** to a matrix from the Harwell-Boeing set of test matrices: **A** is the SHERMAN4 matrix shifted by 0.5 (we wish to compute the eigenvalue of the SHERMAN4 matrix that is nearest to 0.5). **A** is of order 1104. All eigenvalues appear to be real. With the given shift we are aiming for the fifth eigenvalue. For **M** we have taken the ILU(2) decomposition of **A**. We show the $\log_{10}$ of the norm of the residual versus the number of outer iterations steps, which is the dimension of the search space $\mathbf{V}_k$; the FIG. 7.1, 7.2, and 7.3 show the results for, respectively, 5, 10, and 25 steps of GMRES.

Larger values of m imply more accurate approximate solutions of the "expansion equations" (2.4), (2.2), and (4.1). In line with our observations in Sect. 4 we see that improving the precision in Davidson's method slows down the speed of convergence and may even lead to stagnation (the dash-dotted curves $-\cdot-\cdot$). As might be anticipated, for **SI** we observe the opposite effect (the dotted curves ·····); the more precise we solve (4.1), the faster the method converges, while stagnation may occur if (4.1) is not solved rather precisely. The Jacobi-Davidson method does not react so drastically to the precision of the approximate solutions of (2.2) (the solid curves —): the method converges faster than both Davidson and **SI**.
As argued in Sect. 4, **SI** may be rather sensitive to rounding errors, especially if the expanding vector $\widetilde{\mathbf{t}}$ has a large component in the direction $\mathbf{u}$: for **SI**, but also

for Davidson, we had to apply modified Gram-Schmidt (ModGS) twice to maintain orthogonality of $\mathbf{V}_k$, while in Jacobi-Davidson no special precautions were required. The angle between the expanding vector $\widetilde{\mathbf{t}}$ and the available search space can become too small for an accurate computation of the orthogonal component by using ModGS only once. In such a situation, it may help to apply mod-GS more often [14]. For the present example, twice was enough, but other examples, not reported here, required more ModGS sweeps.

Acknowledgement. We gratefully acknowledge helpful discussions with Diederik Fokkema (University Utrecht) on the subject of Sect. 4.2. He also provided the numerical data for the example in Sect. 6.

REFERENCES

[1] J.G.L. Booten, P.M. Meijer, H.J.J. te Riele, and H.A. van der Vorst, *Parallel Arnoldi method for the construction of a Krylov subspace basis: an application in Magnetohydrodynamics*, Report NM-R9406, Dept. of Numerical Mathematics, CWI, Amsterdam, 1994.

[2] M. Crouzeix, B. Philippe, and M. Sadkane, *The Davidson method*, SIAM J. Sci. Comput., 15 (1994), pp. 62–76.

[3] E.R. Davidson, *The iterative calculation of a few of the lowest eigenvalues and corresponding eigenvectors of large real symmetric matrices*, J. Comput. Phys., 17 (1975), pp. 87–94.

[4] E.R. Davidson, in: *Methods in Computational Molecular Physics*, G.H.F. Diercksen and S. Wilson, Eds., Reidel, Dordrecht, 1983, pp. 95–113.

[5] E.R. Davidson, *Monster Matrices: their eigenvalues and eigenvectors*, Computers in Physics, 7 (1993), pp. 519–522.

[6] D.R. Fokkema and G.L.G. Sleijpen, *Computing a partial Schur form for large sparse matrices using Jacobi-Davidson*, Preprint, Dept. of Math., University Utrecht, 1995.

[7] G.H. Golub and C.F. Van Loan, *Matrix Computations*, Second edition, The John Hopkins University Press, Baltimore and London, 1989.

[8] C.G.J. Jacobi, *Ueber ein leichtes Verfahren, die in der Theorie der Säcularstörungen vorkommenden Gleichungen numerisch aufzulösen*, Journal für die reine und angewandte Mathematik, (1846), pp. 51-94.

[9] R.B. Morgan, *Computing interior eigenvalues of large matrices*, Lin. Alg. and its Appl., 154/156 (1991), pp. 289–309.

[10] R.B. Morgan, *Generalizations of Davidson's method for computing eigenvalues of large nonsymmetric matrices*, J. Comput. Phys., 101 (1992), pp. 287–291.

[11] 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.

[12] J. Olsen, P. Jørgensen, and J. Simons, *Passing the one-billion limit in full configuration-interaction (FCI) calculations*, Chemical Physics Letters, 169(6), (1990), pp. 463–472.

[13] A.M. Ostrowski, *On the convergence of the Rayleigh quotient iteration for the computation of characteristic roots and vectors. V*, Arch. Rational. Mech. Anal., 3 (1959), pp. 472–481.

[14] A. Ruhe, *Numerical aspects of Gram-Schmidt Orthogonalization of Vectors*, Lin. Alg. and its Appl., 52/53 (1983), pp. 591–602.

[15] Y. Saad, *Numerical methods for large eigenvalue problems*, Manchester University Press, Manchester, 1992.

[16] G.L.G. Sleijpen and H.A. van der Vorst, *A Jacobi-Davidson iteration method for linear eigenvalue problems*, To appear in SIAM J. Matrix Anal. Appl..

[17] D.C. Sorensen, *Implicitly restarted Arnoldi methods*, SIAM J. Matrix Anal. Appl., 13 (1992), pp. 357–385.

[18] H.J.J. van Dam, J.H. van Lenthe, G.L.G. Sleijpen, and H.A. van der Vorst, *An improvement of Davidson's iteration method; Applications to MRCI and MRCEPA calculations*, To appear in J. Comput. Chem..

[19] J.H. van Lenthe and P. Pulay, *A space-saving modification of Davidson's eigenvector algorithm*, J. Comput. Chem., 11 (1990), pp. 1164–1168.