Universiteit Utrecht

Department of Mathematics

The convergence of Jacobi-Davidson for Hermitian eigenproblems

by

Jasper van den Eshof

Preprint

nr. 1165

November, 2000

THE CONVERGENCE OF JACOBI-DAVIDSON FOR HERMITIAN EIGENPROBLEMS

JASPER VAN DEN ESHOF*

Abstract. Rayleigh Quotient iteration is an iterative method with some attractive convergence properties for finding (interior) eigenvalues of large sparse Hermitian matrices. However, the method requires the accurate (and, hence, often expensive) solution of a linear system in every iteration step. Unfortunately, replacing the exact solution with a cheaper approximation may destroy the convergence. The (Jacobi-)Davidson correction equation can be seen as a solution for this problem. In this paper we deduce quantitative results to support this viewpoint and we relate it to other methods. This should make some of the experimental observations in practice more quantitative in the Hermitian case. Asymptotic convergence bounds are given for fixed preconditioners and for the special case if the correction equation is solved with some fixed relative residual precision. A new dynamic tolerance is proposed and some numerical illustration is presented.

Key words: Hermitian matrices; Eigenvalue problem; Jacobi-Davidson; Davidson's method; Inexact Inverse iteration; Convergence rate

AMS subject classification: 65F15.

1 Introduction

We are interested in methods to calculate an eigenvalue λ and its associated eigenvector x, possibly in the interior of the spectrum, of a large sparse Hermitian matrix A:

$$Ax = \lambda x \ .$$

Iterative methods that employ a *shift-and-invert* approach, often are successful methods for finding such eigenpairs (λ, x), examples of these methods include the *shift-and-invert power method* (Inverse iteration) and *Rayleigh quotient iteration* (RQI) (see, for example [9, 3]). This latter method has very nice local convergence properties, but practical implementations are often expensive due to the required accurate solution of a system involving a shifted matrix of A in every iteration. It is a tempting idea to replace this exact solution by a cheaper approximate solution, for example by using a preconditioner or an iterative method. However, straightforwardly replacing the exact solution in RQI may destroy the local convergence properties and may even result in the inability to find a good approximation to the eigenvector, see also Experiment 3.1 in Section 3.

In literature, many methods have been proposed to overcome this problem. Most notably are the *Davidson* method in [2], the *Jacobi-Davidson* method of Sleijpen and Van der Vorst [11] and *Inexact Rayleigh quotient iteration* as described by, for instance, Smit and Paardekooper [12].

Convergence results for the Davidson method can be found in [1, 7]. They show that if the preconditioner is kept fixed and positive definite, the Davidson method converges to exterior

*Department of Mathematics, Utrecht University, P.O. Box 80.010, NL-3508 TA Utrecht, The Netherlands. E-mail: eshof@math.uu.nl.

eigenvalues in the spectrum. In [5], Lai, Lin, and Lin analyze Inverse Iteration when the necessary inversion is done such that in every iteration the linear system is solved with a fixed relative residual precision (*Inexact Inverse Iteration*). They argue that an increasingly tighter precision is sufficient to assure convergence. Similar work has been done for linear problems (cf. [4]). Lehoucq and Meerbergen consider Inexact Cayley Transforms with fixed parameters in the framework of Rational Krylov sequence methods for generalized eigenproblems [6]. It is observed that the Cayley Transforms can be an improvement over the inexact shift-and-invert approach, in the sense that they allow for solutions with a fixed modest tolerance. In [12], Smit and Paardekooper also analyze Inverse Iteration when the linear system is solved with a fixed residual precision. They include the situation when the fixed shift is replaced with the Rayleigh quotient (Inexact RQI). They argue that, with this shift, a fixed residual precision is sufficient to assure convergence.

In this paper we show that the (Jacobi-)Davidson correction equation can be seen as an improvement over the obvious strategy of replacing the exact solution of the linear system in RQI by an approximation, in a similar fashion as the extra requirement of Inexact RQI on the solution of the linear system guarantees local convergence. This offers insight into the local behavior of the Jacobi-Davidson correction equation.

This insight can be important for a number of reasons. It can, in the first place, help to devise effective preconditioners for eigenproblems. Secondly, it gives an idea how accurately the systems should be solved to assure local convergence, especially when restarts in the subspace acceleration process are used. Furthermore, the interpretation and techniques used, can in most cases be easily extended to non-normal problems.

This paper is organized as follows. In Section 2 a short introduction is given to RQI. Working with a fixed preconditioner usually hampers the local convergence properties. We will explain this by substituting $\widehat{A} \equiv A + E$ for A in the linear system. We refer to the resulting approach as *Preconditioned Rayleigh quotient iteration.* Inexact Rayleigh quotient from [12] is reviewed as Preconditioned Rayleigh quotient iteration with an extra requirement on the solution of the linear system that restores local convergence. The influence of replacing A with $\widehat{A}$ in the Jacobi-Davidson correction equation is investigated in Section 3. We show that this correction equation can be interpreted as an improvement over Preconditioned RQI. This results in convergence bounds for the correction equation of the Jacobi-Davidson method. These results can also be interpreted for the Davidson method, including the situation when searching for interior eigenvalues. In Section 4 the influence on the convergence is considered in case the Jacobi-Davidson correction equation is solved using an iterative solver, this is a special case of Section 3. The iterations on the linear system can in practice be stopped if the residual of the linear system satisfies a relative precision. We show that this is sufficient to assure local convergence and propose a dynamic tolerance. We conclude with a few remarks about the global convergence of Jacobi-Davidson by applying the dynamic tolerance from the previous section in some numerical experiments.

2 Rayleigh quotient iteration

For simplicity we will assume that the matrices A and $\widehat{A} \equiv A + E$ are Hermitian. The matrix A has eigenvalues λ_i and corresponding eigenvectors x_i with $\|x_i\| = 1$; $\sigma(A)$ is the collection of all eigenvalues of A. We are interested in an eigenpair $(\lambda, x) = (\lambda_j, x_j)$ for a specific j, possibly with λ in the interior of the spectrum. This eigenvalue λ is assumed to be simple.

Algorithm 2.1 Preconditioned Rayleigh quotient iteration

Input: A Hermitian matrix A and a normalized starting vector u_0
Output: A new approximation u_m

for $k = 1, \ldots, m$

$\theta = u_{k-1}^* A u_{k-1}$
$z = (A + E_k - \theta I)^{-1} u_{k-1}$
$u_k = z/\|z\|$

end

We define also the constants $\gamma_{\min} \equiv \min_{\mu \in \sigma(A)\setminus\{\lambda\}} |\lambda - \mu|$ and $\gamma_{\max} \equiv \max_{\mu \in \sigma(A)} |\lambda - \mu|$. The corresponding quantities for $\widehat{A}$ are denoted with a hat.

Rayleigh quotient iteration is a well-known method to find an approximation of the eigenpair (λ, x). In this method a new approximation u'_{rq} for x is constructed based on the best approximation (u with $\|u\| = 1$) currently known:

$$u'_{\mathrm{rq}} = (A - \theta I)^{-1} w, \quad \text{where } \theta \equiv u^* A u \tag{1}$$

and w is an appropriate vector. For future convenience we will not assume a particular choice for w. In the standard Rayleigh quotient method $w = u$. It seems reasonable to assume that $\angle(u'_{\mathrm{rq}}, x) = 0$ if $\theta = \lambda$, see Section 4.3 of [9] for a discussion. The following proposition illustrates the attractive local convergence behavior of this method.

Proposition 2.1. *If u'_{rq} is defined by (1), λ is simple ($\gamma_{\min} > 0$) and $w \not\perp x$ then we have asymptotically for u'_{rq}:*

$$|\tan\angle(u'_{\mathrm{rq}}, x)| \leq \frac{\gamma_{\max}}{\gamma_{\min}} \sin^2\angle(u, x)\, |\tan\angle(w, x)| + \mathcal{O}(\sin^4\angle(u, x))$$

Proof. The proof is a straightforward generalization of the proof of Theorem 4.7.1 in [9]. $\square$

Proposition 2.1 shows that RQI converges asymptotically quadratic and for $w = u$ even cubic. This is an attractive property of this method and one might hope that this property remains if we, in order to reduce the cost of the expensive inversion, replace the required inversion by a cheaper alternative.

If (1) is solved approximately, the solution u'_{rq} can be represented as

$$u'_{\mathrm{rq}} = (\widehat{A} - \theta I)^{-1} w, \quad \text{where } \theta \equiv u^* A u \tag{2}$$

and $\widehat{A} \equiv A + E$ for some perturbation E. This equation forms the heart of the Preconditioned RQI method in Algorithm 2.1 where it is allowed that E changes in every step.

Suppose (θ, u) is close to the eigenpair (λ, x). Multiplying u with $(\widehat{A} - \theta I)^{-1}$ gives a large component in the direction of $\widehat{x}$, an eigenvector corresponding to an eigenvalue $\widehat{\lambda}$ (closest to λ) of the nearby linear problem $A + E$. The vector x is not a stationary point for (2) for general E. In general we expect for a fixed preconditioner convergence to $\widehat{x}$ if (θ, u) is very close to

(λ, x). The matrix $\widehat{A}$ needs to have an eigenvector (with eigenvalue close to θ) that makes a smaller angle with x then u. In other words, the matrix $\widehat{A}$ must contain "new information". Experiment 3.1 in Section 3 illustrates that local convergence cannot be expected for general perturbations E.

In [12], Smit and Paardekooper discuss Inexact Rayleigh quotient iteration. This approach can be seen as Preconditioned RQI with the extra requirement that the residual of the linear system satisfies the following (relative) precision (u is still normalized and $\theta \equiv u^* A u$):

$$\|(A - \theta I) u'_{\mathrm{rq}} - u\| \leq \eta < 1 \ . \tag{3}$$

This can be interpreted as implicitly working with an $\widehat{A}$ which has an increasingly better approximate eigenvector for x when u gets closer to x.

Like in [12], where the convergence of Inexact RQI is studied for symmetric A, we can easily show that Inexact RQI converges locally quadratic.

Proposition 2.2. *If u'_{rq} satisfies (3) then we have asymptotically:*

$$|\tan \angle(u'_{\mathrm{rq}}, x)| \leq \frac{\gamma_{\max}}{\gamma_{\min}} \frac{\eta}{\sqrt{1 - \eta^2}} \sin^2 \angle(u, x) + \mathcal{O}(|\sin^3 \angle(u, x)|) \ .$$

Proof. Equation (3) guarantees the existence of a d of unit-length and $\widetilde{\eta}$ such that:

$$(A - \theta I) u'_{\mathrm{rq}} = w, \quad w = u + \widetilde{\eta} d, \quad 0 \leq \widetilde{\eta} \leq \eta \ .$$

A geometric argument shows that

$$|\tan \angle(w, x)| \leq \frac{|\sin \angle(u, x)| + \widetilde{\eta}}{\sqrt{1 - (|\sin \angle(u, x)| + \widetilde{\eta})^2}} \ ,$$

which is bounded if $\widetilde{\eta} \leq \eta < 1$ and $\angle(u, x)$ is small enough with respect to $1 - \eta$. Now apply Proposition 2.1 with this w. □

The result in Proposition 2.2 is a factor two sharper than Corollary 4.3 in [12].

Suppose that a preconditioned iterative solver is used for (1) with a preconditioner $\widehat{A}$ that satisfies

$$\|(A - \theta I)(\widehat{A} - \theta I)^{-1} - I\| \leq \eta < 1, \tag{4}$$

then only one iteration is necessary for this θ to find an u'_{rq} that satisfies (3). However, for fixed $\widehat{A}$, (4) is in general not satisfied anymore when θ gets very close to λ. Proposition 2.2 only gives a condition for local convergence; not a very constructive approach. Solving a (nearly) singular system to a prescribed residual accuracy can still be very expensive.

In the next section we will see that the correction equation of the Jacobi-Davidson method can be seen as an improvement over Preconditioned RQI.

3 The Jacobi-Davidson correction equation with fixed preconditioner

We consider the Jacobi-Davidson correction equation with slightly more general projections (cf. [10]):

$$(I - P^*)(A - \theta I)(I - P) t = r \equiv (A - \theta I) u, \quad t = (I - P) t \tag{5}$$

Algorithm 3.1 Preconditioned Jacobi-Davidson without subspace acceleration

Input: A Hermitian matrix A and a normalized starting vector u_0
Output: A new approximation u_m

for $k = 1, \ldots, m$

$\theta = u_{k-1}^* A u_{k-1}$
$r_k = (A - \theta I) u_{k-1}$
Solve $(I - P_k)(A + E_k - \theta I)(I - P_k)t = r_k$ for $P_k t = 0$
$z = u_{k-1} - t$
$u_k = z / \|z\|$

end

---

with

$$P \equiv \frac{uw^*}{w^*u} \text{ and } \theta \equiv u^*Au \ .$$

We assume that $w \not\perp x$. A new approximation of the eigenvector is formed by $u'_{\text{jd}} \equiv u - t$. If $w = u$ we have the correction equation of the original Jacobi-Davidson method from [11]. We consider more general projections to facilitate a discussion on the convergence of Davidson's method.

We investigate the effect on the convergence when a preconditioner is used to approximately solve this correction equation by replacing A on the left of (5) with the preconditioner $\widehat{A} \equiv A + E$.

(6) $$(I - P^*)(\widehat{A} - \theta I)(I - P)t = r \equiv (A - \theta I)u, \quad t = (I - P)t \ .$$

This correction equation forms the heart of Algorithm 3.1. In analogy with the previous section, we will call this algorithm Preconditioned Jacobi-Davidson (without subspace acceleration).

In [10] it is noted that if $r \perp w$ and w is an eigenvector of $\widehat{A}$ then the Preconditioned JD correction equation is equivalent to the Davidson correction equation, given by:

(7) $$t = (\widehat{A} - \theta I)^{-1} r \ .$$

Therefore, when $r \perp w$, the local convergence behavior of the Davidson method can be assessed by (6). The requirement $r \perp w$ is, for example, fulfilled if w is contained in the test-space of the subspace acceleration.

The following lemma relates (6) to Preconditioned RQI with preconditioner $\widetilde{A} \equiv A + (I - P^*)E(I - P)$.

Lemma 3.1. *Suppose (6) has an unique solution t, $u'_{\text{jd}} \equiv u - t$, $\theta \equiv u^*Au$, and u' satisfies*

$$(\widetilde{A} - \theta I)u' = w, \quad \textit{with } \widetilde{A} \equiv A + (I - P^*)E(I - P),$$

then

$$u' = \alpha u'_{\text{jd}}, \quad \textit{with } \alpha = \frac{w^*u'}{w^*u},$$

and

$$u^* \widetilde{A} u = u^* A u \ .$$

Proof. Decompose $u' = Pu' + (I-P)u' = \alpha u - \widehat{t}$, with $\widehat{t} = -(I-P)u'$. We will show that this $\widehat{t}$ is a multiple of t in (6). We have that

$$(A + (I-P^*)E(I-P) - \theta I)(\alpha u - \widehat{t}) = w \ .$$

Multiplying from the left with $(I-P^*)$ and using that $(I-P^*)w = 0$ gives:

$$(I-P^*)(A + (I-P^*)E(I-P) - \theta I)(\alpha u - \widehat{t}) = 0$$

Now use that $(I-P)u = 0$, $(I-P^*)r = r$ and $\widehat{t} = (I-P)\widehat{t}$:

$$(I-P^*)(A + E - \theta I)\widehat{t} = \alpha r \ .$$

We see that $\alpha^{-1}\widehat{t}$ satisfies (6). The uniqueness of the solution of this equation completes the first part of the proof. The last statement follows, again, from the fact that $(I-P)u = 0$. □

The key observation of this lemma is that one step Preconditioned JD with preconditioner $\widehat{A}$ is equivalent with one step Preconditioned RQI with preconditioner $\widetilde{A}$. This allows us to give a nice (and well-known) consequence for the asymptotic convergence of the exactly solved JD correction equation (cf. Theorem 3.2 in [10] and Section 4.1, Observation (c) in [11]).

Corollary 3.1. *Under the assumptions of Proposition 2.1, correction equation (5) leads, for general w, to asymptotic quadratic convergence and if $w = u$ to asymptotic cubic convergence.*

The Preconditioned JD correction equation can possibly be a good alternative for Preconditioned RQI, because we are only implicitly confronted with $(I-P^*)E(I-P)$ instead of E. The question is whether this change is sufficient to expect convergence for fixed preconditioners under less strict conditions on the eigenvectors of $\widehat{A}$. The following lemma expresses that $\widetilde{A}$ may have a good approximate eigenvector for x when u is close to x.

Lemma 3.2. *If $\delta \equiv \gamma_{\min} - \|(I-P^*)E(I-P)\| > 0$ and $(\widetilde{\lambda}, \widetilde{x})$ is the eigenpair of $\widetilde{A}$ with $\widetilde{\lambda}$ closest to λ, then*

$$|\sin\angle(\widetilde{x}, x)| \leq \delta^{-1}\|(I-P^*)E(I-P)x\| \leq \delta^{-1}\|(I-P^*)E(I-P)\| \ |\sin\angle(u,x)|$$

and

$$|\widetilde{\lambda} - \lambda| \leq \|(I-P^*)E(I-P)\| \ |\sin\angle(u,x)| \ .$$

Proof. The first inequality follows from

$$\|(I-P^*)E(I-P)x\| = \|(\widetilde{A} - \lambda I)x\| \geq$$
$$\|(I - \widetilde{x}\widetilde{x}^*)(\widetilde{A} - \lambda I)x\| = \|(I - \widetilde{x}\widetilde{x}^*)(\widetilde{A} - \lambda I)(I - \widetilde{x}\widetilde{x}^*)x\| \geq$$
$$\min_{\mu \in \sigma(\widetilde{A})\backslash\{\widetilde{\lambda}\}} |\mu - \lambda| \|(I - \widetilde{x}\widetilde{x}^*)x\| = \min_{\mu \in \sigma(\widetilde{A})\backslash\{\widetilde{\lambda}\}} |\mu - \lambda| |\sin\angle(\widetilde{x}, x)| \ .$$

Using the Bauer-Fike Theorem (Theorem 7.2.2 in [3]) it follows that:

$$\min_{\mu \in \sigma(\widetilde{A})\backslash\{\widetilde{\lambda}\}} |\mu - \lambda| \geq \gamma_{\min} - \|(I-P^*)E(I-P)\| = \delta \ .$$

Noting that $(I-P)x = (I-P)(I-uu^*)x$ and $\|(I-uu^*)x\| = |\sin\angle(u,x)|$ completes the first part of the proof. The second statement follows, for example, from an application of the Bauer-Fike Theorem and remembering that $\widetilde{\lambda}$ is the closest eigenvalue to λ. □

The important consequence of Lemma 3.2 is that if $\|(I-P^*)E(I-P)\|/\delta < 1$ the matrix $\widetilde{A}$ has an eigenvector $\widetilde{x}$ that makes a smaller angle with x than the vector u. We can say that, in this case, $\widetilde{A}$ contains "new information". This gives us already an asymptotic convergence bound for the Preconditioned JD correction equation if $\theta = \widetilde{\lambda}$. In the following theorem the last missing details are filled in for more general θ in (6).

Theorem 3.1. *If E satisfies the condition*

$$\gamma_{\min} - \|(I - \frac{wx^*}{x^*w})E(I - \frac{xw^*}{w^*x})\| > 0, \tag{8}$$

with $w \not\perp x$ and t is a solution of (6), then asymptotically we have for $u'_{\mathrm{jd}} \equiv u - t$:

$$|\sin \angle(u'_{\mathrm{jd}}, x)| \leq \frac{\|(I - \frac{wx^*}{x^*w})E(I - \frac{xw^*}{w^*x}))\|}{\gamma_{\min} - \|(I - \frac{wx^*}{x^*w})E(I - \frac{xw^*}{w^*x})\|}|\sin \angle(u, x)| + \mathcal{O}(\sin^2 \angle(u, x)) \ .$$

Proof. If $u \to x$ then the δ from Lemma 3.2 goes to the expression on the left in (8). So δ is asymptotically bounded from below. This and Lemma 3.2 guarantee that for $\angle(u, x)$ small enough $\widetilde{\gamma}_{\min}$ is nonzero and we can apply Proposition 2.1.

$$|\sin \angle(u'_{\mathrm{jd}}, x)| \leq |\sin \angle(\widetilde{x}, x)| + |\sin \angle(u'_{\mathrm{jd}}, \widetilde{x})|$$

$$\begin{aligned} &\leq |\sin \angle(\widetilde{x}, x)| + \frac{\widetilde{\gamma}_{\max}}{\widetilde{\gamma}_{\min}}|\tan \angle(w, \widetilde{x})| \sin^2 \angle(u, \widetilde{x}) + \mathcal{O}(\sin^4 \angle(u, \widetilde{x})) \\ &\leq |\sin \angle(\widetilde{x}, x)| + \mathcal{O}\left((|\sin \angle(u, x)| + |\sin \angle(\widetilde{x}, x)|)^2\right) = \\ &|\sin \angle(\widetilde{x}, x)| + \mathcal{O}(\sin^2 \angle(u, x)) \ . \end{aligned}$$

In the second line we have applied Proposition 2.1, in the third line Lemma 3.2. The proof is concluded by using Lemma 3.2 and:

$$\|(I - P^*)E(I - P)\| = \|(I - \frac{wx^*}{x^*w})E(I - \frac{xw^*}{w^*x}))\| + \mathcal{O}(|\sin \angle(u, x)|) \ .$$

□

A consequence of this theorem is that correction equation (7) converges linearly if $\|E\|$ is small enough. The obliqueness of the projections may play a role in the constant. Note however that $|w^*x|^{-1}$ is only 2, even for $\angle(w, x) = \pi/3$. This effect might be small in practice.

Theorem 3.1 gives us an asymptotic convergence bound for the Davidson method in case the residual, r, is orthogonal to w and w is an eigenvector of $\widehat{A}$. The requirement $r \perp w$ can be forced by inserting w in the test-space of the subspace acceleration. This possibly explains the good convergence behavior of the Davidson method. The analysis in [1, 7] does not rely on the requirement that $r \perp w$ and can only be applied for exterior eigenpairs. The vector w is in practice often not available. So, local convergence can not be guaranteed by inserting w in the test-space. The Jacobi-Davidson correction equation solves this problem by explicitly using the projections, as the following corollary shows.

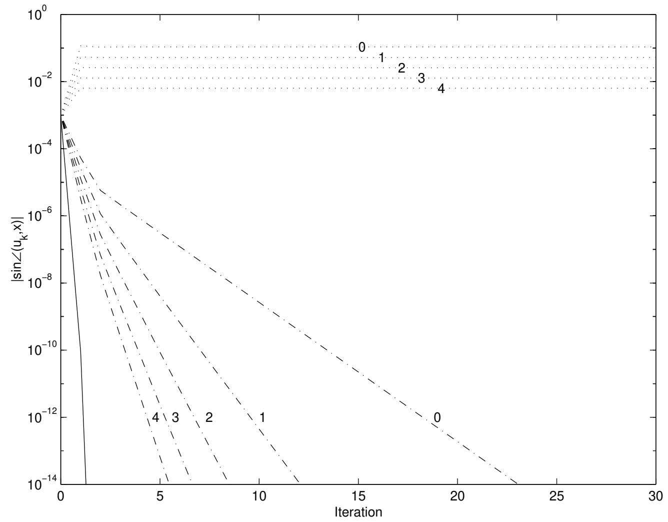

Figure 1: Convergence history Algorithm 2.1 (..) and Algorithm 3.1 (-.) with preconditioner $\widehat{A}_\epsilon$, with $\epsilon = 2^{-j}$ and $j = 0, 1, 2, 3, 4$ and exact RQI (-)

Corollary 3.2. *If E satisfies condition (8) with $w = x$ and t is a solution of (6) with $w = u$ then asymptotically we have for $u'_{\text{jd}} \equiv u - t$:*

$$|\sin\angle(u'_{\text{jd}}, x)| \leq \frac{\|(I - xx^*)E(I - xx^*)\|}{\gamma_{\min} - \|(I - xx^*)E(I - xx^*)\|} |\sin\angle(u, x)| + \mathcal{O}(\sin^2\angle(u, x))$$

We illustrate Corollary 3.2 for a simple test problem.

Experiment 3.1. The matrix $A \in \mathbb{R}^{100\times 100}$ is diagonal with $A_{ii} = i$. We can choose this matrix diagonal without loss of generality. We also generated a random symmetric matrix, E, with eigenvalues uniform in the interval $[-1, 1]$ and model a preconditioner with $\widehat{A}_\epsilon \equiv A + \epsilon E$. With the starting-vector

$$\sqrt{\frac{100}{199}}\,(1/10, \ldots, 1/10, 1, 1/10, \ldots, 1/10)^T$$

we searched for the eigenvalue $\lambda = 50$.

Figure 1 shows the convergence history of the values of $|\sin\angle(u_k, x)|$, for Algorithm 2.1, Algorithm 3.1, and Rayleigh quotient iteration with exact inversions (there is no subspace acceleration). In this picture we see that exact RQI only needs 2 iterations to find a very good approximate eigenvector. This picture also illustrates the stagnation of preconditioned RQI when an eigenvector of $\widehat{A}_\epsilon$ has been found, which was detected by monitoring $\|\widehat{A}_\epsilon u_k - (u_k^*\widehat{A}_\epsilon u_k)u_k\|$. In practical situations this could mean that a very small (and impractical) value of ϵ is necessary for preconditioned RQI to find an accurate approximation. The Jacobi-Davidson correction equation seems, due to the projections, more able to handle this $\widehat{A}_\epsilon$. Making ϵ a factor 2 smaller appears to speed up the local convergence with a factor 2, even for these relatively large values of ϵ.

This experiment and Corollary 3.2 clearly illustrate the idea behind using projections in the JD correction equation. Linear convergence is expected if $\|E\|$ is small enough with respect to the gap ratio but without making other requirements on the eigenvectors of $\widetilde{A}$. The projections in the Preconditioned Jacobi-Davidson method have the geometric interpretation that we implicitly apply a shift-and-invert step on a system that has an increasingly better approximate eigenvector for x. Condition (3) can also be interpreted this way. This condition can only be fulfilled when u'_{rq} is constructed with a preconditioner that has an increasingly better approximate eigenvector when θ becomes closer to λ.

4 Solving the Jacobi-Davidson correction equation with an iterative solver

In this section we look at a special case of the preconditioner discussed in the previous section and consider the situation where the Jacobi-Davidson correction equation is solved with a fixed residual precision. In practice, this can be accomplished by a suitable Krylov subspace method for linear systems , for example MINRES [8].

So, let us consider the solution of (5) with $w = u$. We denote the $n - 1$ eigenpairs of A , restricted to $u^\perp$, by $(\widetilde{\lambda}_i, \widetilde{x}_i)$. We have the following relation between the eigenvalues of A and those of the restricted operator.

Lemma 4.1. *For every eigenvalue $\lambda_i \in \sigma(A)$ there exists an eigenvalue $\widetilde{\lambda} \in \sigma((I-uu^*)A(I-uu^*)) \setminus \{0\}$ such that:*

$$|\lambda_i - \widetilde{\lambda}| \leq \|A - \lambda_i I\| \; |\cos \angle(u, x_i)| \ .$$

Proof. With $\widetilde{u}_i = (I - uu^*)x_i / \|(I - uu^*)x_i\|$ we get that there exists a $\widetilde{\lambda} \in \sigma((I - uu^*)A(I - uu^*))$ such that:

$$|\lambda_i - \widetilde{\lambda}_i| \leq \|(I - uu^*)A(I - uu^*)\widetilde{u}_i - \lambda_i \widetilde{u}_i\| =$$

$$\|(I - uu^*)(A - \lambda_i I)\widetilde{u}_i\| \leq \|(A - \lambda_i I)\| \; \|(I - x_i x_i^*)\widetilde{u}_i\|$$

The first inequality follows from the Bauer-Fike Theorem (Theorem 7.2.2 in [3]). The last step can be proved by noting that:

$$\|(I - x_i x_i^*)\widetilde{u}_i\| = |\sin \angle(u^\perp, x_i)| = |\cos \angle(u, x_i)| \ .$$

□

Lemma 4.1 says that if u has more or less equal components in all eigenvector directions, then the spectrum of the restricted operator is more or less distributed as the spectrum of A. On the other hand, if u makes a very small angle with the particular eigenvector x, then the spectrum of the restricted operator is a first order perturbation of the set $\sigma(A) \setminus \{\lambda\}$. We expect in this case that JD behaves similar as the Davidson method if only a few steps of an iterative solver are used in the asymptotic case.

The inner-iterations can in practice be terminated if, for some fixed $\eta < 1$, a $\widehat{t}$ is found that satisfies:

$$\|(I - uu^*)(A - \theta I)(I - uu^*)\widehat{t} - r\| \leq \eta \|r\| \ . \tag{9}$$

Because the residual reduction, for methods like MINRES, is most effective for components in directions with eigenvalues away from the shift θ, it is conceivable that only a relatively modest number of MINRES iterations is sufficient to achieve (9) for the asymptotic situation.

Theorem 4.1. *If $\widehat{t}$ satisfies (9) then we have asymptotically for $u'_{\rm jd} \equiv u - \widehat{t}$:*

$$|\tan \angle(u'_{\rm jd}, x)| \leq \eta \frac{\gamma_{\max}}{\gamma_{\min}} |\sin \angle(u, x)| + \mathcal{O}(\sin^2 \angle(u, x)) \ .$$

Proof. Equation (9) guarantees the existence of a d with $\|d\|_2 = 2$ and an $\widetilde{\eta} \leq \eta$, such that

$$(I - uu^*)(A - \theta I)(I - uu^*)\widehat{t} = r + \widetilde{\eta}\|r\|d \ .$$

We obtain

$$u'_{\rm jd} \equiv u - \widehat{t} = u - t + e,$$

where e satisfies

$$(I - uu^*)(A - \theta I)(I - uu^*)e = -\widetilde{\eta}\|r\|d \ .$$

The smallest singular value of the operator at the left can, with Lemma 4.1, shown to be $\gamma_{\min} + \mathcal{O}(|\sin \angle(u, x)|)$. The asymptotic bound follows by bounding the right-hand side with

$$\|r\| \leq \|(A - \lambda I)u\| \leq \gamma_{\max} |\sin \angle(u, x)| \tag{10}$$

and using Proposition 2.1 and that $\|u - t\| = \|u\| + \|t\| \geq 1$. □

The local convergence can be accelerated by selecting a smaller value for η. For example, one could consider the dynamic tolerance:

$$\|(I - uu^*)(A - \theta I)(I - uu^*)\widehat{t} - r\| \leq \eta \|r\|^2 \ . \tag{11}$$

Note that this expression is not invariant for scalings of A. So, we can try to divide η by an estimate of the spread of the spectrum (in our experiments we take η a multiple of $\|r_0\|_2^{-1}$). However, condition (11) still guarantees quadratic local convergence, as the following theorem shows.

Theorem 4.2. *If $\widehat{t}$ satisfies (11) then we have asymptotically for $u'_{\rm jd} \equiv u - \widehat{t}$:*

$$|\tan \angle(u'_{\rm jd}, x)| \leq \eta \gamma_{\max} \frac{\gamma_{\max}}{\gamma_{\min}} \sin^2 \angle(u, x) + \mathcal{O}(|\sin^3 \angle(u, x)|) \ .$$

Proof. It follows from (10) that

$$\|r\|^2 \leq \gamma_{\max}^2 \sin^2 \angle(u, x).$$

The proof is completed by following the same strategy as in the proof of Theorem 4.1. □

In the second experiment Theorem 4.1 and Theorem 4.2 are demonstrated.

Experiment 4.1. We take the same test problem as in Experiment 3.1 and to simulate the solution of the correction equation with a prescribed residual precision, Algorithm 3.1 was executed and in step k the terms $d_k \eta \|r_k\|$ and $d_k \eta \|r_0\|^{-1} \|r_k\|^2$ were added to the residual to simulate (9) and (11) respectively, with d_k random, $d_k \perp u_{k-1}$, and $\|d_k\| = 1$.

Figure 2 illustrates the result. The convergence when condition (9) is used demonstrates an almost linear dependence on η (although η is chosen relatively large). For condition (11) the local convergence seems about as fast as for exact Rayleigh quotient iteration as stated in Theorem 4.2 and not sensitively depending on η.

Condition (11) is based on results for the asymptotic situation and it is not clear whether this condition is effective for more realistic problems. In the next section some numerical experiments are described that show that this tolerance still can be useful.

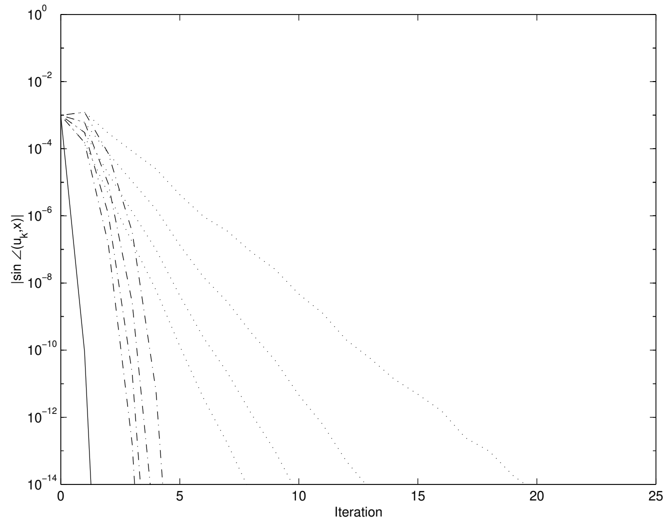

Figure 2: Convergence history Algorithm 3.1 when the correction equation is solved with a relative residual precision given by (9) (..) with $\eta = \frac{1}{4}, \frac{1}{8}, \frac{1}{16}$ and $\frac{1}{32}$ and (11) (-.) with $\eta = \frac{1}{4\|r_0\|}, \frac{1}{8\|r_0\|}, \frac{1}{16\|r_0\|}$ and $\frac{1}{32\|r_0\|}$ and exact RQI (-).

5 Numerical experiments

In the previous sections we focused on the local convergence. The reason for this was that problems can occur for Inexact Rayleigh quotient iteration in this stage of the convergence. Studying the global convergence of Jacobi-Davidson is a much more difficult subject because the different components of a Jacobi-Davidson implementation can have complicated interaction. For example, the subspace acceleration, that we ignored in the preceding sections, plays an important role in the global convergence. Using a few numerical experiments and the dynamic tolerance (11) from the previous section, we address some practical problems.

We apply the Jacobi-Davidson method for finding the largest eigenvalue (in absolute value) of a number of Hermitian matrices from the Matrix Market. We compare two different conditions for terminating GMRES. In the first approach we apply a fixed number, m, iterations GMRES in every iteration step of JD. For convenience, we refer to this as *Method 1*. In the second approach (*Method 2*) the number of GMRES iterations is also bounded by m but we terminate early if the solution of the linear system satisfies the relative precision (9) with

$$\eta = \frac{\|r_k\|}{\|r_0\|}, \tag{12}$$

where r_k is the (eigenvector) residual in step k (this gives us the situation of (11))

The implementation of Jacobi-Davidson consists of an implementation of GMRES without preconditioning for solving (5) and subspace acceleration. The experiments are conducted in Matlab. The starting vector is a normalized vector with all ones.

In the previous section we considered using a dynamic tolerance (only) , cf. (11). Experiments (not shown here) learn that, if we solve (5) with an iterative solver, using this tolerance is only a good idea in some situations. One cause of problems are misselections in the subspace acceleration. In this case a lot of MVs are spent on an irrelevant expansion. Another problem is given by stagnation at small residues for similar reasons. Therefore,we

Matrix	Exact	Method 1			Method 2		
		$m=5$	$m=10$	$m=15$	$m=5$	$m=10$	$m=15$
SHERMAN1	-/29	121/20	287/26	465/29	98/21	149/23	141/18
BCSPWR05	-/17	121/20	133/12	177/11	77/14	90/10	99/9
BCSPWR06	-/35	169/28	254/23	305/19	192/34	225/24	272/24
GR3030	-/59	163/27	320/29	529/33*	97/22	101/16	124/15
BFW782B	-/15	73/12	100/9	145/9	64/12	67/8	60/6
CAN1054	-/11	79/13	100/9	145/9	74/13	73/8	98/8
JAGMESH8	-/8	199/33	188/17	193/12	203/35	192/19	181/13
LSHP1009	-/7	115/19	122/11	145/9	119/21	113/12	126/10
NOS3	-/>200	205/34	584/53	625/39	122/44	128/39	130/37
PLAT1919	-/>200	211/35	353/32*	657/41	40/16	44/15	49/15
ZENIOS	-/29	55/9	100/9	161/10	34/10	34/9	38/9

Table 1: Number of MVs and number of iterations (MVs/Iterations) to reduce the residual to 10^{-10} when searching for the largest (in absolute value) eigenvalue. The results are given for exact inversions and Method 1 and Method 2 with m iterations GMRES. The star denoted entries indicate that a different eigenvalue was found than the largest (misconvergence).

here added a bound on m. This gives us Method 2. Which we will compare to Method 1 for the same m. Table 1 gives the total number of matrix vector multiplications (MVs) and JD iterations necessary to reduce the residual of eigenvector approximation to 10^{-10} for some matrices from the Matrix Market.

We see from Table 1 that using the dynamic tolerance can in some cases improve the required number of MVs significantly. For the *PLAT1919* matrix and $m = 5$ even with a factor of more than 5. In four instances, using an additional tolerance increased the number of MVs, but not very dramatic. Furthermore, note that the columns for the dynamic tolerance show no misconvergence.

We now give a possible explanation for the observed improvements. Figure 3 gives the convergence history for the *BCSPWR05* matrix. From the stagnation in this figure, we suspect that an exact inversion of the correction equation does not introduce the interesting eigenvector component in the first few iterations. Convergence is slow until the wanted eigenvector is represented well enough. For $m = 0$ (equivalent with Arnoldi) the convergence is relatively fast for this matrix. For Method 1, 5 iterations GMRES are used to solve the correction equation. In the first iteration it is likely that the approximate solution of GMRES is similar to the exact solution in the direction of the interesting eigenvector. This is suggested by the quick convergence of Arnoldi and the fact that this eigenvector is not (yet) deflated from the system by the projections. This results in a behavior similar to that for exact inversions and give stagnation, see Figure 3. In other words, Method 1 does not profit from the good global convergence of Arnoldi but, even more, can suffer from it with a bad starting vector.

Method 2 forces a number of "Arnoldi" iterations in the beginning and switches to more exact inversions automatically. The fact that Method 2 better exploits the good global convergence properties of Arnoldi can also help reduce the chance on misconvergence.

Now, consider the most extreme example, the *PLAT1919* matrix. For this matrix and

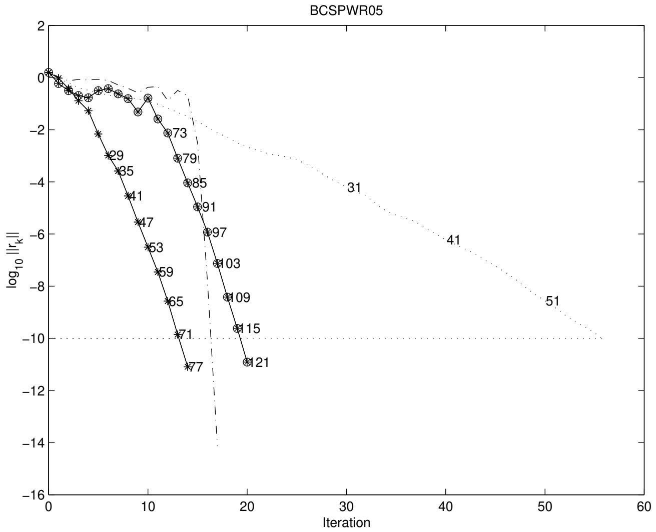

Figure 3: Number of JD iteration steps (x-axis) and MVs (numbers in plot) when searching for largest eigenvalue with exact inversions (-.), $m = 0$ (...) and $m = 5$ (-) for Method 2 (*) and Method 1 (o)

starting vector the convergence with exact inversions (RQI) is very slow due the bad starting vector. With $m = 0$ we need, however, only about 30 MVs. And, Method 2 indeed prevents misconvergence and gives good global convergence.

Two final notes. If no projections are used in the correction equation (Davidson method) and too many GMRES iterations are done this can even completely remove the interesting information from t (i.e. $u - t \perp x$). In that sense the projections play also an important role. This is also observed in [11]. As a final note, we remark that if we are working with a fixed number of MVs per iteration (Method 1) and there is stagnation due to the reason described above, then there is interesting information in the subspace constructed by the linear solver. This raises the question how this information can be detected and exploited automatically. This is future work.

Acknowledgments The author thanks Henk van der Vorst and Gerard Sleijpen for many helpful comments and suggestions. The Matlab code used in Section 5 was written by Gerard Sleijpen (*http://www.math.uu.nl/people/sleijpen/*).

References

[1] M. Crouzeix, B. Philippe, and M. Sadkane. The Davidson method. *SIAM J. Sci. Comput.*, 15(1):62–76, 1994.

[2] Ernest R. Davidson. The iterative calculation of a few of the lowest eigenvalues and corresponding eigenvectors of large real-symmetric matrices. *J. Comput. Phys.*, 17:87–94, 1975.

[3] G. H. Golub and C. F. Van Loan. *Matrix Computations*. The John Hopkins University Press, Baltimore, London, 3rd edition, 1996.

[4] G. H. Golub and Q. Ye. Inexact preconditioned conjugate gradient method with inner-outer iteration. *SIAM J. Sci. Comput.*, 21(4):1305–1320, 1999.

[5] Yu-Ling Lai, Kun-Yi Lin, and Wen-Wei Lin. An inexact inverse iteration for large sparse eigenvalue problems. *Num. Lin. Alg. Appl.*, 4:425–437, 1997.

[6] R. B. Lehoucq and Karl Meerbergen. Using generalized Cayley transformations within an inexact rational Krylov sequence method. *SIAM J. Matrix Anal. Appl.*, 20(1):131–148 (electronic), 1999.

[7] S. Oliveira. On the convergence rate of a preconditioned subspace eigensolver. *Computing*, 63(3):219–231, 1999.

[8] Christopher C. Paige and Michael A. Saunders. Solution of sparse indefinite systems of linear equations. *SIAM J. Numer. Anal.*, 12(4):617–629, 1975.

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

[10] Gerard L. G. Sleijpen, Albert G. L. Booten, Diederik R. Fokkema, and Henk A. van der Vorst. Jacobi-Davidson type methods for generalized eigenproblems and polynomial eigenproblems. *BIT*, 36(3):595–633, 1996. International Linear Algebra Year (Toulouse, 1995).

[11] Gerard L. G. Sleijpen and Henk A. Van der Vorst. A Jacobi-Davidson iteration method for linear eigenvalue problems. *SIAM Review*, 42(2):267–293, 2000.

[12] P. Smit and M. H. C. Paardekooper. The effects of inexact linear solvers in algorithms for symmetric eigenvalue problems. *Linear Alg. Appl.*, 287:337–357, 1998.