A NEW JUSTIFICATION OF THE JACOBI–DAVIDSON METHOD FOR LARGE EIGENPROBLEMS

HEINRICH VOSS*

Abstract. The Jacobi–Davidson method is known to converge at least quadratically if the correction equation is solved exactly, and it is common experience that the fast convergence is maintained if the correction equation is solved only approximately. In this note we derive the Jacobi–Davidson method in a way that explains this robust behavior.

1. Introduction. In this paper we consider the large and sparse eigenvalue problem

$$Ax = \lambda x \tag{1.1}$$

or more generally the nonlinear eigenproblem

$$T(\lambda)x = 0 \tag{1.2}$$

where $A \in \mathbb{C}^{n\times n}$ and $T : D \to \mathbb{C}^{n\times n}$, $D \subset \mathbb{C}$ is a family of sparse matrices.

For the linear problem (1.1) iterative projection methods have proven to be very efficient if a small number of eigenvalues and eigenvectors are desired. Here the eigenproblem is projected to a subspace of small dimension which yields approximate eigenpairs. If an error tolerance is not met then the search space is expanded in an iterative way with the aim that some of the eigenvalues of the reduced matrix become good approximations of some of the wanted eigenvalues of the given large matrix.

Particularly efficient are Krylov subspace methods like the Lanczos and the Arnoldi algorithm which provide rapid convergence to well separated and extreme eigenvalues and corresponding eigenvectors. For interior eigenvalues these methods tend to exhibit difficulties which can be remedied by shift-and-invert techniques, i.e. by applying the method to the matrix $(A - \sigma I)^{-1}$ where σ denotes a shift which is close to the wanted eigenvalues.

However, for truly large eigenproblems it is very costly or even infeasible to solve the shift-and-invert equation $(A - \sigma I)x = y$ by a direct method as LU factorization, and an iterative method has to be employed to solve it approximately.

Unfortunately, methods like the Lanczos algorithm and the Arnoldi algorithm are very sensitive to inexact solutions of $(A - \sigma I)x = y$, and therefore the combination of these methods with iterative solvers of the shift-and-invert equation usually is inefficient (cf. [3, 5, 12, 13, 14, 15]).

An iterative projection method which is more robust to inexact expansions of search spaces than Krylov subspace methods is the Jacobi–Davidson method which was introduced approximately 10 years ago by Sleijpen and van der Vorst [17] for the linear eigenproblem (1.1), and which was extended to matrix pencils in [4], to polynomial eigenproblems in [16], and to the general nonlinear eigenvalue problem (1.2) in [2] and [19]. A survey has recently been given in [6], pseudo codes are contained in [1].

Usually the Jacobi–Davidson expansion of a search space $\mathcal{V}$ is derived as orthogonal correction t of a current Ritz pair (θ, x) which is the solution of the so called

*Institute of Numerical Simulation, Hamburg University of Technology, D-21071 Hamburg, Germany, voss@tu-harburg.de

correction equation

$$(I - xx^H)(A - \theta I)(I - xx^H)t = -(A - \theta I)x, \quad t \perp x. \tag{1.3}$$

It has been shown in [17] that the expanded space $\text{span}\{\mathcal{V}, t\}$ contains the direction $(A - \theta I)^{-1}x$ which is obtained by one step of the Rayleigh quotient iteration. Hence, one can expect quadratic convergence, which is even cubic in the Hermitian case.

It is common experience that fast convergence is maintained if the correction equation (1.3) is solved only approximately. But the way the expansion of the search space was derived by Sleijpen and van der Vorst does not indicate why the Jacobi–Davidson method is more robust to inaccurate solutions of the correction equation than Krylov type methods to inexact solutions of the shift-and-invert system.

In this note we rederive the Jacobi–Davidson method in a way that explains its robustness.

2. A geometric derivation of a robust search space expansion. Consider the linear eigenvalue problem (1.1). Let $\mathcal{V}$ be the current search space of an iterative projection method. Assume that $x \in \mathcal{V}$ with $\|x\| = 1$ is the current approximation to the eigenvector we are aiming at, and let $\theta = x^H A x$ be the corresponding Rayleigh quotient. Because of its good approximation property we want to expand the search space by the direction of inverse iteration $v = (A - \theta I)^{-1}x / \|(A - \theta I)^{-1}x\|$.

However, in a truly large problem the vector v will not be accessible but only an inexact solution $\tilde{v} := v + e$ of $(A - \theta I)v = x$, and the next iterate will be a solution of the projection of (1.1) onto the space $\tilde{\mathcal{V}} := \text{span}\{\mathcal{V}, \tilde{v}\}$.

We assume that x is already a good approximation to an eigenvector of A. Then v will be an even better approximation, and therefore the eigenvector we are looking for will be very close to the plane $E := \text{span}\{x, v\}$. We therefore neglect the influence of the orthogonal complement of x in $\mathcal{V}$ on the next iterate and discuss the nearness of the planes E and $\tilde{E} := \text{span}\{x, \tilde{v}\}$. If the angle between these two planes is small, then the projection of (1.1) onto $\tilde{\mathcal{V}}$ should be similar to the one onto $\text{span}\{\mathcal{V}, v\}$, and the approximation properties of inverse iteration should be maintained. If this angle can become large, then it is not surprising that the convergence properties of inverse iteration are not reflected by the projection method.

We denote by $\phi_0 = \arccos(x^H v)$ the angle between x and v, and the relative error of $\tilde{v}$ by $\varepsilon := \|e\|$.

THEOREM 2.1. *The maximal possible acute angle between the planes E and $\tilde{E}$ is*

$$\beta(\varepsilon) = \begin{cases} \arccos\sqrt{1 - \varepsilon^2/\sin^2\phi_0} & \text{if} \quad \varepsilon \le |\sin\phi_0| \\ \frac{\pi}{2} & \text{if} \quad \varepsilon \ge |\sin\phi_0| \end{cases} \tag{2.1}$$

Proof. For $\varepsilon > |\sin\phi_0|$ the vector x is contained in the ball with center v and radius ε, and therefore the maximum acute angle between E and $\tilde{E}$ is $\frac{\pi}{2}$.

For $\varepsilon \le |\sin\phi_0|$ we assume without loss of generality that $v = (1, 0, 0)^T$, $\tilde{v} = (1 + e_1, e_2, e_3)^H$, and $x = (\cos\phi_0, \sin\phi_0, 0)^T$. Obviously the angle between E and $\tilde{E}$ is maximal, if the plane $\tilde{E}$ is tangential to the ball B with center v and radius ε. Then $\tilde{v}$ is the common point of ∂B and the plane $\tilde{E}$, i.e. the normal vector $\tilde{n}$ of $\tilde{E}$ has the same direction as the perturbation vector e:

$$e = \begin{pmatrix} e_1 \\ e_2 \\ e_3 \end{pmatrix} = \gamma \tilde{n} = \gamma \begin{pmatrix} \cos\phi_0 \\ \sin\phi_0 \\ 0 \end{pmatrix} \times \begin{pmatrix} 1 + e_1 \\ e_2 \\ e_3 \end{pmatrix} = \gamma \begin{pmatrix} e_3 \sin\phi_0 \\ -e_3 \cos\phi_0 \\ e_2 \cos\phi_0 - (1 + e_1)\sin\phi_0 \end{pmatrix}. \tag{2.2}$$

Hence, we have $e_1 = \gamma \sin\phi_0 e_3$, $e_2 = -\gamma\cos\phi_0 e_3$, and the third component yields

$$e_3 = \gamma(-\gamma\cos^2\phi_0 e_3 - (1+\gamma\sin\phi_0 e_3)\sin\phi_0) = -\gamma^2 e_3 - \gamma\sin\phi_0,$$

i.e.

$$e_3 = -\frac{\gamma}{1+\gamma^2}\sin\phi_0. \tag{2.3}$$

Moreover, from

$$\varepsilon^2 = e_1^2 + e_2^2 + e_3^2 = \gamma^2\sin^2\phi_0 e_3^2 + \gamma^2\cos^2\phi_0 e_3^2 + e_3^2 = (1+\gamma^2)e_3^2,$$

we obtain

$$\varepsilon^2 = \frac{\gamma^2}{1+\gamma^2}\sin^2\phi_0, \quad \text{i.e. } \gamma^2 = \frac{\varepsilon^2}{\sin^2\phi_0 - \varepsilon^2}.$$

Inserting into (2.3) yields

$$e_3^2 = \frac{1}{1+\gamma^2}\varepsilon^2 = \Big(1 - \frac{\varepsilon^2}{\sin^2\phi_0}\Big)\varepsilon^2,$$

and since the normal vector of E is $n = (0,0,1)^T$, we finally get

$$\cos\beta(\varepsilon) = \frac{n\times e}{\|n\|\cdot\|e\|} = \frac{e_3}{\varepsilon} = \sqrt{1 - \frac{\varepsilon^2}{\sin^2\phi_0}}.$$

□

Obviously for every $\alpha\in\mathbb{R}$, $\alpha\neq 0$ the plane E is also spanned by x and $x+\alpha v$. If $\tilde{E}(\alpha)$ is the plane which is spanned by x and a perturbed realization $x+\alpha v+e$ of $x+\alpha v$ then by the same arguments as in the proof of Theorem 2.1 the maximum angle between E and $\tilde{E}(\alpha)$ is

$$\gamma(\alpha,\varepsilon) = \begin{cases} \arccos\sqrt{1-\varepsilon^2/\sin^2\phi(\alpha)} & \text{if } \varepsilon \le |\sin\phi(\alpha)| \\ \frac{\pi}{2} & \text{if } \varepsilon \ge |\sin\phi(\alpha)| \end{cases} \tag{2.4}$$

where $\phi(\alpha)$ denotes the angle between x and $x+\alpha v$. Since the mapping

$$\phi \mapsto \arccos\sqrt{1-\varepsilon^2/\sin^2\phi}$$

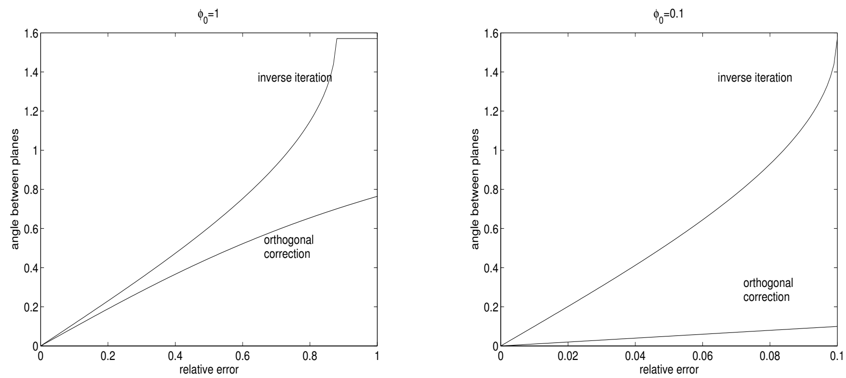
decreases monotonically the expansion of the search space by an inexact realization of $t := x+\alpha v$ is most robust with respect to small perturbations, if α is chosen such that x and $x+\alpha v$ are orthogonal, i.e. by

$$t = x - \frac{x^H x}{x^H(A-\theta I)^{-1}x}(A-\theta I)^{-1}x. \tag{2.5}$$

Then the maximum acute angle between E and $\tilde{E}(\alpha)$ satisfies

$$\gamma(\alpha,\varepsilon) = \begin{cases} \arccos\sqrt{1-\varepsilon^2} & \text{if } \varepsilon\le 1 \\ \frac{\pi}{2} & \text{if } \varepsilon \ge 1 \end{cases}. \tag{2.6}$$

Figure 1 shows the maximum angles between the planes $E = \text{span}\{x,v\}$ and $\tilde{E} = \text{span}\{x,\tilde{v}\}$ if $\tilde{v}$ is obtained by inexact evaluation of the direction of inverse iteration v and of the orthogonal correction t, respectively, for two angles $\phi_0 = 1$ and $\phi_0 = 0.1$ between x and v. It demonstrates that for a large angle ϕ_0 the robustness does not increase very much, but for small angles, i.e. in case where x is already quite accurate, the gain of robustness is essential.



3. Jacobi–Davidson method. Obviously, the expansion t in (2.5) of the current search space $\mathcal{V}$ is the solution of the equation

$$(I - xx^H)(A - \theta I)(I - xx^H)t = (A - \theta I)x, \quad t \perp x. \tag{3.1}$$

This is the so called correction equation of the Jacobi–Davidson method which was derived by Sleijpen and van der Vorst in [17] as a generalization of an approach of Jacobi [7] for improving the quality of an eigenpair of a symmetric matrix. Hence, the Jacobi–Davidson method is the most robust realization of an expansion of a search space such that the direction of inverse iteration is contained in the expanded space in the sense that it is least sensitive to inexact solves of linear systems $(A-\theta I)v = x$.

Similarly, we obtain the Jacobi–Davidson expansions for more general eigenvalue problems. Consider the generalized eigenvalue problem

$$Ax = \lambda Bx \tag{3.2}$$

where B is nonsingular. Then given an approximation (θ, x) to an eigenpair the inverse iteration is defined by $v := (A - \theta B)^{-1}Bx$. Again, we expand the current search space by $t := x + \alpha v$, where α is chosen such that x and $x + \alpha v$ are orthogonal, i.e. by

$$t = x - \frac{x^H x}{x^H (A - \theta B)^{-1} Bx}(A - \theta B)^{-1} Bx,$$

and this is the solution of the well known correction equation

$$\Big(I - \frac{Bxx^H}{x^H Bx}\Big)\Big(A - \theta B\Big)\Big(I - \frac{xx^H}{x^H x}\Big)t = (A - \theta B)x, \quad t \perp x \tag{3.3}$$

of the Jacobi–Davidson method [4].

If B is Hermitian and positive definite, and angles are measured with respect to the scalar product $\langle x, y\rangle_B := x^H By$, then the robustness requirement $\langle x, x + \alpha v\rangle_B = 0$ yields the expansion

$$t = x - \frac{x^H Bx}{x^H B(A - \theta B)^{-1} Bx}(A - \theta B)^{-1} Bx,$$

which is the solution of the symmetric correction equation (cf. [16])

$$\Big(I - \frac{Bxx^H}{x^H Bx}\Big)\Big(A - \theta B\Big)\Big(I - \frac{xx^H B}{x^H Bx}\Big)t = (A - \theta B)x, \quad t \perp_B x. \tag{3.4}$$

Finally, we consider the nonlinear eigenproblem (1.2) where the elements of T are assumed to be differentiable with respect to λ. Then given an eigenpair approximation (θ, x) the direction of inverse iteration is $v = T(\theta)^{-1}T'(\theta)x$. $t := x + \alpha v$ is orthogonal to x if

$$t = x - \frac{x^H x}{x^H T(\theta)^{-1} T'(\theta) x} T(\theta)^{-1} T'(\theta) x,$$

and this is the solution of the correction equation

$$\Big(I - \frac{T'(\theta) x x^H}{x^H T'(\theta) x}\Big) T(\theta) \Big(I - \frac{x x^H}{x^H x}\Big) t = T(\theta) x, \quad t \perp x. \tag{3.5}$$

which was discussed in [2, 19], and for polynomial eigenvalue problems in [16].

4. Inexact Krylov subspace methods. In [10] Meerbergen and Rose investigate an inexact shift-and-invert Arnoldi method for the generalized eigenvalue problem $Ax = \lambda Bx$. They demonstrate the superior numerical performance of a Cayley transformation over that of a shift–invert transformation within an Arnoldi method when using an iterative linear solver. Similarly Lehoucq and Meerbergen [8] showed that the Cayley transformation leads to a more robust eigensolver than the usual shift-and-invert transformation when the linear systems are solved inexactly within the rational Krylov method.

Aiming at the eigenvalue $\tilde\lambda$ that is closest to some shift σ in both methods the current search space $\mathcal{V}$ is expanded by

$$t_{SI} = (A - \sigma B)^{-1} B x \tag{4.1}$$

where x is a Ritz vector with respect to $\mathcal{V}$ corresponding to the Ritz value θ closest to σ.

Since

$$(A - \sigma B)^{-1}(A - \theta B)x = x + (\sigma - \theta)(A - \sigma B)^{-1} B x$$

and $x \in \mathcal{V}$, this expansion is equivalent to the one given by the Cayley transformation

$$t_C = (A - \sigma B)^{-1}(A - \theta B)x \tag{4.2}$$

if (4.1) and (4.2) are evaluated in exact arithmetic.

However, since $|x^H t_{SI}|/\|t_{SI}\| \to 1$ as $\theta \to \tilde\lambda$ and x approaches an eigenvector corresponding to $\tilde\lambda$ whereas $x^H t_C/\|t_C\| \to 0$, the considerations in Section 2 indicate that we may expect a more robust behavior of Arnoldi's method and the rational Krylov method, if the search space is expanded by an inexact realization of t_C than by an approximation to t_{SI}.

Similar considerations hold for the nonlinear Arnoldi method [9, 18] for problem (1.2). There the expansion of the search space is motivated by the residual inverse iteration $t_{RI} = x - T(\sigma)^{-1}T(\theta)x$ (cf. [11]) which converges quickly if σ is close to the wanted eigenvalue. Since in iterative projection methods the new search direction is orthogonalized against the basis of the current search space for stability reasons and since x is already contained in $\mathcal{V}$, the expansion was chosen to be $t_A := T(\sigma)^{-1}T(\theta)x$. In this case we have $|x^H t_{RI}|/\|t_{RI}\| \to 1$ and $x^H t_A/\|t_A\| \to 0$ such that the expansion by t_A turns out to be more robust than the one by t_{RI}.

REFERENCES

[1] Z. Bai, J. Demmel, J. Dongarra, A. Ruhe, and H.A. van der Vorst, editors. *Templates for the Solution of Algebraic Eigenvalue Problems: A Practical Guide*. SIAM, Philadelphia, 2000.

[2] T. Betcke and H. Voss. A Jacobi–Davidson–type projection method for nonlinear eigenvalue problems. *Future Generation Computer Systems*, 20(3):363 – 372, 2004.

[3] A. Bouras and V. Frayssé. A relaxation strategy for the Arnoldi method in eigenproblems. Technical Report TR/PA/00/16, CERFACS, Toulouse, 2000.

[4] D.R. Fokkema, G.L.G. Sleijpen, and H.A. Van der Vorst. Jacobi-Davidson style QR and QZ algorithms for the partial reduction of matrix pencils. *SIAM J.Sci.Comput.*, 20:94 – 125, 1998.

[5] G.H. Golub and Q. Ye. An inverse free preconditioned Krylov subspace method for symmetric generalized eigenvalue problems. *SIAM J. Sci. Comput.*, 24(1):312 – 334, 2002.

[6] M.E. Hochstenbach and Y. Notay. The Jacobi–Davidson method, 2006. To appear in GAMM Mitteilungen.

[7] C.G.J. Jacobi. Über ein leichtes Verfahren, die in der Theorie der Säkularstörungen vorkommenden Gleichungen numerisch aufzulösen. *Crelle J. Reine u. Angew. Math.*, 30:51 – 94, 1846.

[8] R. B. Lehoucq and K. Meerbergen. Using generalized Cayley transformation within an inexact rational Krylov sequence method. *SIAM J. Matrix Anal. Appl.*, 20:131–148, 1998.

[9] K. Meerbergen. Locking and restarting quadratic eigenvalue solvers. *SIAM Journal on Scientific Computing*, 22(5):1814 – 1839, 2001.

[10] K. Meerbergen and D. Roose. The restarted Arnoldi method applied to iterative linear system solvers for the computation of rightmost eigenvalues. *SIAM J. Matrix Anal. Appl.*, 18:1–20, 1997.

[11] A. Neumaier. Residual inverse iteration for the nonlinear eigenvalue problem. *SIAM J. Numer. Anal.*, 22:914–923, 1985.

[12] Y. Notay. Robust parameter free algebraic multilevel preconditioning. *Numer. Lin. Alg. Appl.*, 9:409 – 428, 2002.

[13] Y. Notay. Inner iterations in eigenvalue solvers. Technical Report GANMN 05-01, Université Libre de Bruxelles, 2005.

[14] V. Simoncini. Variable accuracy of matrix–vector products in projection methods for eigencomputation. *SIAM J. Numer. Anal.*, 43:1155 – 1174, 2005.

[15] V. Simoncini and L. Eldén. Inexact Rayleigh quotient-type methods for eigenvalue computations. *BIT*, 42:159 – 182, 2002.

[16] G.L. Sleijpen, G.L. Booten, D.R. Fokkema, and H.A. van der Vorst. Jacobi-Davidson type methods for generalized eigenproblems and polynomial eigenproblems. *BIT*, 36:595 – 633, 1996.

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

[18] H. Voss. An Arnoldi method for nonlinear eigenvalue problems. *BIT Numerical Mathematics*, 44:387 – 401, 2004.

[19] H. Voss. A Jacobi–Davidson method for nonlinear eigenproblems. In M. Buback, G.D. van Albada, P.M.A. Sloot, and J.J. Dongarra, editors, *Computational Science – ICCS 2004, 4th International Conference, Kraków, Poland, June 6–9,2004,Proceedings, Part II*, volume 3037 of *Lecture Notes in Computer Science*, pages 34–41, Berlin, 2004. Springer Verlag.