THE INEXACT RATIONAL KRYLOV SEQUENCE METHOD*

R. B. LEHOUCQ† AND KARL MEERBERGEN‡

Abstract. The rational Krylov sequence (RKS) method is a generalization of Arnoldi's method. It constructs an orthogonal reduction of a matrix pencil into an upper Hessenberg pencil. The RKS method is useful when the matrix pencil may be efficiently factored. However, it requires the solution of a linear system at every step. This article considers solving the resulting linear systems in an inexact manner by using an iterative method. We show that a Cayley transformation used within the RKS method is more efficient and robust than the usual shift-and-invert transformation. A relationship with the recently introduced Jacobi–Davidson method of Sleijpen and van der Vorst is also established.

1. Introduction. Suppose that a few eigenvalues near a complex number μ and possibly corresponding eigenvectors of the generalized matrix eigenvalue problem

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

are needed. Assume that both A and B are large complex matrices of order n. Also suppose that at least one of A or B is nonsingular so that equation (1.1) has n eigenvalues. Without loss of generality, assume that B is invertible. Following standard convention, we refer to (A, B) as a matrix pencil. For us, n is considered large when it is prohibitive to compute all the eigenvalues as a dense algorithm in LAPACK [1] would attempt to do.

A standard approach is to perform inverse iteration [18, page 405] with the matrix $A - \mu B$. The sequence of iterates

$$v,\ (A-\mu B)^{-1}Bv,\ [(A-\mu B)^{-1}B]^2 v,\ \ldots \tag{1.2}$$

is produced. Under some mild assumptions, the sequence converges toward the desired eigenvector, and a Rayleigh quotient calculation gives an estimate of the eigenvalue. Another approach is to extract the approximate eigenpair by using the information from the subspace defined by joining together m iterates of the sequence (1.2). This leads to a straightforward extension [20] of the ideas introduced by Ericsson and Ruhe [14] for the spectral (shift-and-invert) transformation Lanczos method. Starting with the vector v, Arnoldi's method [2] builds, step by step, an orthogonal basis for the *Krylov* subspace

$$\mathcal{K}_m(T^{SI}, v) \equiv \mathrm{Span}\{v,\ T^{SI}v,\ \ldots,\ (T^{SI})^{m-1}v\} \quad \text{where} \quad T^{SI} = (A-\mu B)^{-1}B.$$

One improvement to the inverse iteration scheme given is to vary the shift $\mu \equiv \mu_j$ at every step. For example, set μ_j equal to the Rayleigh quotient $z^H Az/z^H Bz$, where z is an unit vector in the direction of $(T^{SI})^j v$. Ruhe [29, 31] elegantly shows how to build an orthogonal basis for the *rational Krylov* subspace

$$\mathrm{Span}\{v,\ T_1^{SI}v,\ \cdots,\ (T_{m-1}^{SI}\cdots T_1^{SI})v\}, \quad \text{where} \quad T_j^{SI} = (A-\mu_j B)^{-1}B.$$

* The work of R. B. Lehoucq was supported by the Mathematical, Information, and Computational Sciences Division subprogram of the Office of Computational and Technology Research, U.S. Department of Energy, under Contract W-31-109-Eng-38. The work by Karl Meerbergen was supported by the project *Iterative Methods in Scientific Computing*, contract number HCM network CHRC-CT93-0420, coordinated by CERFACS, Toulouse, France.

† Mathematics and Computer Science Division, Argonne National Laboratory, Argonne, IL 60439 USA, lehoucq@mcs.anl.gov, http://www.mcs.anl.gov/home/lehoucq/index.html.

‡ Department of Mathematics, Utrecht University, 3584 CD Utrecht, The Netherlands, meerbergen@math.ruu.nl, http://www.math.ruu.nl/people/meerbergen.

The resulting algorithm is called a rational Krylov sequence (RKS) method and is a generalization of the shift-and-invert Arnoldi method where the shift is varied during each step.

All the methods considered require the solution of $(A - \mu B)x = By$ for x. This is typically accomplished by factoring $A - \mu B$. For example, when $A - \mu B$ is sparse, a direct method [6, 7, 9, 10, 12, 11] may be employed. If the shifts μ_j are not varied, then use of one of these direct methods in conjunction with ARPACK [19] is a powerful combination for computing a few solutions of the generalized eigenvalue problem (1.1).

However, for large eigenvalue problems ($n > 10,000$), direct methods using the RKS method may not provide an efficient solution because of the potentially prohibitive storage requirements. The motivation for the current study is to investigate the use of iterative methods for the linear systems of equations arising in the RKS method. We shall call these methods *inexact* RKS ones. One benefit is that for the many eigenvalue problems arising from a discretization of partial differential equations, an intelligent preconditioner may often be constructed. In particular, we shall demonstrate that a Cayley transformation $T_j^C \equiv (A - \mu_j B)^{-1}(A - \nu_j B)$ performs more robustly than a shift-and-invert transformation $T_j^{SI} \equiv (A - \mu_j B)^{-1}B$ when using iterative methods for the linear solves.

Fittingly, the literature on approaches for finding a few solutions to the generalized eigenvalue problem (1.1), where only approximate solutions to the linear systems are available, is sparse. Szyld [38] considers the situation where the matrix pencil is symmetric positive definite. Algorithms based on Jacobi-Davidson methods are discussed in [15, 34]. The recent report by Meerbergen and Roose [21] provided motivation for the current article.

Our article is organized as follows. We introduce the RKS method in §2. The inexact RKS method is introduced in §3 along with a connection with inverse iteration and some examples illustrating our ideas. In §4, we illustrate our method for a generalized eigenvalue problem. We compare the inexact RKS and Jacobi–Davidson methods in §5. We conclude the paper in §6 with a summary of the main ideas and some remaining questions.

In this article, matrices are denoted by upper-case Roman characters. Vectors are denoted by lower-case Roman characters. The range of the matrix V is denoted by $\mathcal{R}(V)$. The Hermitian transpose of the vector x is denoted by x^H.

2. The Rational Krylov Sequence Method. The method is outlined by the algorithm listed in Figure 2. For the practical RKS algorithm given in [31], Ruhe considers the shift-and-invert transformation $T_j^{SI} = (A - \mu_j B)^{-1}B$ rather than $T_j^C = (A - \mu_j B)^{-1}(A - \nu_j B)$. In exact arithmetic, both transformations lead to the same rational Krylov space, since

$$T_j^C = I + (\mu_j - \nu_j)T_j^{SI}. \tag{2.1}$$

However, in finite-precision arithmetic and/or in conjunction with iterative methods for linear systems, substantial differences may exist. We call the μ_j's the poles, the ν_j's the zeros, and the $V_j t_j$'s the continuation vectors. The selection of these quantities is postponed until §3. This section will discuss some relationships in Steps 1–7, the form of Gram–Schmidt orthogonalization we employ, and finally the computation of approximate eigenpairs and their convergence.

By eliminating w from Steps 2–5, we obtain the relationship

$$(A - \mu_j B)^{-1}(A - \nu_j B)V_j t_j \equiv V_{j+1}\tilde{h}_j, \tag{2.2}$$

- Choose a starting vector v_1 with $\|v_1\| = 1$.
- Set $V_0 \leftarrow [\ v_1\ ]$, and $\tilde{L}_0, \tilde{K}_0 \leftarrow [\ ]$.
- For $j = 1, 2, \ldots, m$
  1. Select a pole μ_j, a zero $\nu_j \neq \mu_j$ and a vector $t_j \in R^j$ with $\|t_j\| = 1$.
  2. Form $w = V_j t_j$ (*continuation vector*).
  3. Form $w \leftarrow (A - \mu_j B)^{-1}(A - \nu_j B)w$.
  4. Orthogonalize $w := w - V_j h_j$ with $h_j = V_j^H w$ (*Gram-Schmidt*).
  5. Set $V_{j+1} = \left[\begin{array}{cc} V_j & w/h_{j+1,j} \end{array}\right]$ where $h_{j+1,j} = \|w\|$.
  6. Set $\tilde{l}_j = \left[\begin{array}{c} h_j \\ h_{j+1,j} \end{array}\right] - \left[\begin{array}{c} t_j \\ 0 \end{array}\right]$ and $\tilde{k}_j = \left[\begin{array}{c} h_j \\ h_{j+1,j} \end{array}\right] \mu_j - \left[\begin{array}{c} t_j \\ 0 \end{array}\right] \nu_j$.
  7. Set $\tilde{L}_j = \left[\begin{array}{cc} \tilde{L}_{j-1} & \tilde{l}_j \\ 0 & \end{array}\right]$ and $\tilde{K}_j = \left[\begin{array}{cc} \tilde{K}_{j-1} & \tilde{k}_j \\ 0 & \end{array}\right]$.
  8. Compute approximate eigenpairs of interest.
  9. Check whether the approximate eigenpairs satisfy the convergence criterion.

FIG. 2.1. *Computing the Rational Krylov Sequence (RKS) for the matrix pencil (A,B)*

where $\tilde{h}_j = \left[\begin{array}{cccc} h_{1,j} & h_{2,j} & \cdots & h_{j+1,j} \end{array}\right]^T$. Let $\tilde{t}_j = \left[\begin{array}{cc} t_j^T & 0 \end{array}\right]^T$. Rearranging Equation (2.2) results in

$$\begin{aligned} (A - \mu_j B)V_{j+1}\tilde{h}_j &= (A - \nu_j B)V_{j+1}\tilde{t}_j\,, \\ AV_{j+1}(\tilde{h}_j - \tilde{t}_j) &= BV_{j+1}(\tilde{h}_j\mu_j - \tilde{t}_j\nu_j). \end{aligned}$$

By putting together the relations for $j = 1, \ldots, m$, we have that

$$AV_{m+1}(\tilde{H}_m - \tilde{T}_m) = BV_{m+1}(\tilde{H}_m M_m - \tilde{T}_m N_m), \tag{2.3}$$

where $\tilde{h}_j$ and $\tilde{t}_j$ are associated with the jth columns of $\tilde{H}_m$ and $\tilde{T}_m$, respectively, and $M_m = \mathrm{diag}(\mu_1, \ldots, \mu_m)$, $N_m = \mathrm{diag}(\nu_1, \ldots, \nu_m)$.

A final simplification is to rewrite Equation (2.3) as

$$AV_{m+1}\tilde{L}_m = BV_{m+1}\tilde{K}_m\,, \tag{2.4}$$

where $\tilde{L}_m \equiv \tilde{H}_m - \tilde{T}_m$ and $\tilde{K}_m \equiv \tilde{H}_m M_m - \tilde{T}_m N_m$. We remark that as long as $h_{j+1,j}$ is nonzero, both $\tilde{H}_m$ and $\tilde{L}_m$ are unreduced upper Hessenberg (rectangular) matrices and thus of full rank.

In Appendix A, we show that the use of T_j^{SI} leads to the RKS relation

$$AV_{m+1}\tilde{L}_m(M_m - N_m)^{-1} = BV_{m+1}\tilde{K}_m(M_m - N_m)^{-1}\ . \tag{2.5}$$

2.1. Orthogonalization. The orthogonalization of Step 3 of Algorithm 2 uses an iterative classical Gram-Schmidt algorithm. This is the same approach used by Sorensen [37] based on the analysis [5] of reorthogonalization in the Gram–Schmidt algorithm.

2.2. Computing Eigenvalue Estimates. We now consider the calculation of approximate eigenpairs for the RKS method. We discuss how to compute Ritz pairs. Harmonic Ritz pairs [23, 27, 24, 36] may also be computed, as was shown by Ruhe [31], but these are not considered here. The main purpose of this article is to study the use of iterative linear system solvers in RKS and not the various ways to extract

eigenvalues. Therefore, we use standard Ritz values throughout, though the theory can easily be extended to harmonic Ritz values.

Consider a matrix C and a subspace $\mathcal{R}(X)$, where $X \in \boldsymbol{C}^{n\times k}$ is of full rank. The pair $(\theta, y \equiv Xz)$ is called a Ritz pair of C with respect to the subspace $\mathcal{R}(X)$ if and only if

$$Cy - \theta y \perp \mathcal{R}(X) \ .$$

This is referred to as a Galerkin projection. Two important properties of a Galerkin projection are the following. First, if $\mathcal{R}(X) \equiv \boldsymbol{C}^n$, the Ritz pairs are exact eigenpairs of C. Second, if C is normal, the Ritz values lie in the convex hull of the eigenvalues of C. For example, if C is Hermitian, the Ritz values lie between the smallest and largest eigenvalue of C.

The following theorem shows how Ritz pairs may be computed from the RKS method outlined by the algorithm listed in Figure 2.

THEOREM 2.1. *$(\theta, y \equiv V_{m+1}\tilde{L}_m z)$ is called a Ritz pair for $B^{-1}A$ in $\mathcal{R}(V_{m+1}\tilde{L}_m)$ if and only if*

$$\tilde{L}_m^H \tilde{K}_m z = \theta \tilde{L}_m^H \tilde{L}_m z. \tag{2.6}$$

Proof. Following the definition and Equation (2.4), (θ, y) is a Ritz pair when

$$B^{-1}AV_{m+1}\tilde{L}_m z - \theta V_{m+1}\tilde{L}_m z = V_{m+1}(\tilde{K}_m - \theta\tilde{L}_m)z \perp \mathcal{R}(V_{m+1}\tilde{L}_m).$$

Thus, $(V_{m+1}\tilde{L}_m)^H V_{m+1}(\tilde{K}_m - \theta\tilde{L}_m)z = 0$, and the desired relation (2.6) is established. □

We denote by $\theta_i^{(m)}$ the approximate eigenvalues computed available after m steps of the RKS algorithm of Figure 2. Unless otherwise stated, we assume that the Ritz values are in increasing distance from μ_m, that is, $|\theta_1^{(m)} - \mu_m| \le |\theta_2^{(m)} - \mu_m| \le \cdots \le |\theta_m^{(m)} - \mu_m|$. The associated Ritz vector is denoted by $y_i^{(m)}$. The sub- and superscripts are omitted whenever the context is clear.

2.2.1. Computing Ritz Pairs. The generalized eigenvalue problem (2.6) may be solved as a standard one. Since $\tilde{L}_m$ is an unreduced upper Hessenberg matrix, $\tilde{L}_m$ is of full rank, and hence $\tilde{L}_m^H\tilde{L}_m$ is invertible. Thus, the standard eigenvalue problem

$$\tilde{L}_m^\dagger \tilde{K}_m z = z\theta, \quad \text{where} \quad \tilde{L}_m^\dagger = (\tilde{L}_m^H\tilde{L}_m)^{-1}\tilde{L}_m,$$

is solved. We remark that $\tilde{L}_m^\dagger$ is the Moore-Penrose generalized inverse of $\tilde{L}_m$. The explicit formation of the inverse of $\tilde{L}_m^H\tilde{L}_m$ is not required. Instead, $\tilde{L}_m^\dagger\tilde{K}_m$ may be computed by least squares methods, for example with the LAPACK [1] software. The Ritz vector is $y = V_{m+1}\tilde{L}_m z$.

2.3. Stopping Criterion. The accuracy of a Ritz pair $(\theta, y = V_{m+1}L_m z)$ is typically estimated by the residual norm $\|Ay - By\theta\|$. From Equation (2.4), it follows that

$$f \equiv Ay - \theta By = BV_{m+1}(\tilde{K}_m - \theta\tilde{L}_m)z \equiv BV_{m+1}g. \tag{2.7}$$

Thus, a simple check for convergence of a Ritz pair in Algorithm 2 is when

$$\|g\| \le \text{TOL}$$

is satisfied for a user-defined error tolerance TOL. Since $(A+E)y = By\theta$, where $E = -fy^H/(y^Hy)$, it follows that if $\|B^{-1}E\| = \|-V_{m+1}g\| = \|g\|$ is small relative to $\|B^{-1}A\|$, then (θ, y) is an eigenpair for a nearby problem. If θ is not a poorly conditioned eigenvalue of the matrix pencil and $\|B^{-1}\|$ is not large, then the size of $\|g\|$ indicates the accuracy of the computed Ritz value.

This conclusion motivates us to say that the sequence of Ritz pairs $(\theta_i^{(m)}, y_i^{(m)})$ *converges* toward an eigenpair of Equation (1.1) if and only if $\|g_i^{(m)}\|$ tends to zero as m increases toward n. Although this convergence is not rigorously defined (we necessarily have $\|g_i^{(n)}\| = 0$), it does allow us to track the progress of a Ritz pair after step m of Algorithm 2.

3. The Inexact RKS Method. At Steps 3–5 of the RKS algorithm in Figure 2 the Cayley transformation

$$V_{j+1}\tilde{h}_j = (A-\mu_j B)^{-1}(A-\nu_j B)V_j t_j$$

is computed by a two step process. First, the linear system

$$(A-\mu_j B)w = (A-\nu_j B)V_j t_j \tag{3.1}$$

is solved for w. Next, w is orthogonalized against V_j, and the solution $V_{j+1}\tilde{h}_j$ results. These two steps account for the largest source of errors arising when computing in floating-point arithmetic. Since our interest is in using a (preconditioned) iterative method for the solution of Equation (3.1), we neglect the errors in the Gram–Schmidt orthogonalization phase, as explained in §2.1.

Let us formally analyze the errors arising from the solution of Equation (3.1). With a robust implementation of a direct method, a *backward* stable solution is computed. Let $x_j = V_{j+1}\tilde{h}_j$ denote the computed solution and $s_j \equiv (A-\nu_j B)V_j t_j - (A-\mu_j B)x_j$ the associated residual. Thus,

$$(A-\mu_j B + s_j x_j^H/\|x_j\|^2)x_j = (A-\nu_j B)V_j t_j\,.$$

Here, $\|s_j x_j^H\|/\|x_j\|^2 = \|s_j\|$. If $\|s_j\|/\|A-\mu_j B\|$ is a modest multiple of machine precision, we say that the direct method is backward stable. Note that even if a backward stable solution x_j is in hand, however, it may share few, if any, digits of accuracy with w. Moreover, achieving such a backward stable solution with an iterative method may be prohibitively expensive. Therefore, we shall study the situation for which a *large* backward error is allowed for the solution of the linear system.

In order to give an indication of what we mean by large, a few (brief) words about iterative linear system solvers are needed. A linear system $Cx = b$ is said to be solved with a *relative residual tolerance* τ when the solution, x, satisfies $\|b - Cx\| \le \tau\|b\|$. Basic iterative solvers include the Jacobi and Gauss-Seidel relaxation methods. These are called stationary solvers, since the solution can be written as $x = M^{-1}b + Gx^0$, where x^0 denotes the initial solution. In all our experiments, $x^0 = 0$ so that $\|b-Cx\| = \|b - CM^{-1}b\| \le \tau\|b\|$ with $\tau = \|I - CM^{-1}\|$. Thus, we obtain, roughly speaking, the same relative residual norm for any b. In general, Krylov methods [4, 16] are much more powerful. GMRES [33], BiCGSTAB(ℓ) [35], and QMR [17] are among those most widely used. The performance of these solvers often substantially improves when a suitable preconditioner is employed. See [3] for templates for all these solvers. To summarize, then, what we mean by a large error is that τ lies in the interval $[\,10^{-8}, 10^{-2}\,]$.

By putting all the s_j for $j = 1, \ldots, m$ together in $S_m \equiv \left[\begin{array}{ccc} s_1 & \cdots & s_m \end{array} \right]$, we have

$$AV_{m+1}\tilde{L}_m = BV_{m+1}\tilde{K}_m - S_m, \tag{3.2}$$

which we call an inexact rational Krylov sequence (I-RKS) relation. This relation may be rewritten as

$$(A + S_m\tilde{L}_m^{\dagger}V_{m+1}^H)V_{m+1}\tilde{L}_m = BV_{m+1}\tilde{K}_m \ , \tag{3.3}$$

where $\tilde{L}_m^{\dagger} = (\tilde{L}_m^H\tilde{L}_m)^{-1}\tilde{L}_m^H$ is the generalized Moore-Penrose inverse. In other words, we have computed an exact RKS for the pencil $(A+E_m, B)$, where $E_m = S_m\tilde{L}_m^{\dagger}V_{m+1}^H$. Denote by $\sigma_{\min}^{-1}(\tilde{L}_m)$ the reciprocal of the minimum singular value of $\tilde{L}_m$. Thus, if

$$\|E_m\| \leq \|S_m\| \, \|\tilde{L}_m^{\dagger}\| = \|S_m\|\sigma_{\min}^{-1}(\tilde{L}_m)$$

is large, the Ritz pairs from §2.2.1 may not be those of a pencil near (A, B). This situation implies that even if we use a direct method for the linear systems, a nearly rank deficient $\tilde{L}_m$ might lead to inaccurate Ritz pairs. We call the Ritz pairs for $(A+E_m, B)$ (§2.2.1) *inexact* Ritz pairs for (A, B). We define and discuss a few quantities that will prove helpful in the discussion that follows.

- *Cayley residual* s_j : this is the residual of the linear system (3.1).
- *RKS residual* $f^{(j)}$: the RKS method computes a Ritz pair $(\theta^{(j)}, y^{(j)})$ with $y^{(j)} = V_{j+1}\tilde{L}_j z^{(j)}$ and $\|y^{(j)}\| = 1$ for $(A + E_j, B)$, and so the RKS residual satisfies

$$f^{(j)} \equiv BV_{j+1}(\tilde{K}_j - \theta^{(j)}\tilde{L}_j)z^{(j)} = (A + E_j)y^{(j)} - \theta^{(j)}By^{(j)}.$$

- *True residual* $r^{(j)}$: this is the residual defined by $r^{(j)} = Ay^{(j)} - \theta^{(j)}By^{(j)}$.

These three residuals may be linked via the two relationships

$$r^{(j)} = f^{(j)} - S_j z^{(j)} = f^{(j)} - E_j y^{(j)} \quad \text{for} \quad j = 1, \ldots, m, \tag{3.4}$$

which follow from Equation (3.3) and the definition of E_j. We assume that the exact RKS method converges in m iterations. Hence, $f^{(m)} = 0$, and, in order to have the sequence $\|r^{(j)}\|$ converge to zero, $\|S_j z^{(j)}\|$ must also tend to zero. We present numerical evidence that demonstrates that this situation occurs when using an inexact Cayley transformation, whereas it does not when an inexact shift-and-invert transformation is used.

The choice of the zero of the Cayley transformation is crucial to its success, as was pointed out in [21]. Suppose that $(\theta^{(j-1)}, y^{(j-1)})$ is an (inexact) Ritz pair computed in the $(j-1)$st iteration. Then, t_j and ν_j are chosen such that $\nu_j = \theta^{(j-1)}$ and $V_j t_j = y^{(j-1)}$. We then solve the linear system

$$(A - \mu_j B)x_j = r^{(j-1)} \ . \tag{3.5}$$

With an iterative system solver with relative residual tolerance τ, we obtain

$$\|s_j\| \leq \tau\|r^{(j-1)}\| \ . \tag{3.6}$$

If the shift-and-invert transformation is used, the system

$$(A - \mu_j B)x_j = BV_j t_j$$

- Given v_1, $\|v_1\| = 1$. Let $t_1 = [1]$ and $\theta^{(0)} = 0$.
- **For** $j = 1, 2, \ldots, m$.
  1. Compute residual $r^{(j-1)} = AV_j t_j - \theta^{(j-1)} BV_j t_j$.
  2. **If** $\|r^{(j-1)}\| <$ TOL **then** exit.
  3. Let the zero $\nu_j = \theta^{(j-1)}$ and select a pole $\mu_j \neq \nu_j$.
  4. Solve (approximately) the linear system $(A - \mu_j B)x = r^{(j-1)}$ for x.
  5. Orthonormalize x against V_j.
  6. Update $\tilde{L}_j$ and $\tilde{K}_j$.
  7. Solve the eigenvalue problem $\tilde{L}_j^{\dagger} \tilde{K}_j z = \theta z$ (see § 2.2.1).
  8. Let $\theta^{(j)}$ be Ritz value of interest and the continuation vector be $t_{j+1} = \tilde{L}_j z^{(j)} / \|\tilde{L}_j z^{(j)}\|$ associated with $\theta^{(j)}$.

FIG. 3.1. *Computing eigenvalues of the pencil* (A, B) *by the inexact rational Krylov sequence (I-RKS) method*

is solved. Solving this linear system with the same relative residual tolerance as for the Cayley transform, we obtain

$$\|s_j\| \leq \tau \|BV_j t_j\| \leq \tau \|B\| \,. \tag{3.7}$$

Figure 3 shows an algorithm that implements I-RKS using the Cayley transformation. We now illustrate a few properties of this algorithm by means of an example.

Consider the matrices $A = \mathrm{diag}(1, \cdots, 5)$ and $B = I$. The pencil (A, I) has eigenpairs (j, e_j), $j = 1, \ldots, 5$. The goal is to compute the eigenpair $(1, e_1)$ with I-RKS using a fixed pole $\mu_j = 0.7$ and starting with $v_1 = [\ 1, \ldots, 1\ ]^T/\sqrt{5}$. The Cayley system

$$(A - \mu_j I)x_j = r^{(j-1)}$$

is solved as $x_j = M^{-1} r^{(j-1)}$, where

$$M^{-1} = \begin{bmatrix} (1-\mu_j)^{-1} & 10^{-2} & & \\ 10^{-2} & \ddots & \ddots & \\ & \ddots & \ddots & 10^{-2} \\ & & 10^{-2} & (5-\mu_j)^{-1} \end{bmatrix} .$$

The residual tolerance is $\tau = \|I - (A - \mu_j I)M^{-1}\| \approx 5 \cdot 10^{-2}$. We performed $m = n = 5$ iterations, so $\mathcal{R}(V_5 L_5) \equiv \boldsymbol{C}^n$, which implies that $f_i^{(5)} = 0$ for $i = 1, \ldots, 5$. Thus, the computed eigenpairs are exact eigenpairs of $A + E_5$. We found that

$$A + E_5 = \begin{bmatrix} 1.0000 & 0.0120 & -0.0697 & 0.3708 & -0.4728 \\ -0.0000 & 1.9987 & -0.5981 & 4.4591 & -5.6013 \\ -0.0001 & 0.1003 & 0.4666 & 17.1897 & -21.1757 \\ -0.0002 & 0.0340 & -4.4220 & 36.8172 & -40.7251 \\ -0.0002 & 0.0151 & -3.7375 & 26.8228 & -27.8127 \end{bmatrix}$$

- Given v_1, $\theta^{(0)}$ and $\tilde{l}_0 = [1]$.
- **For** $j = 1, 2, \ldots m$.
  1. Select a pole μ_j.
  2. Let the zero be $\nu_j = \theta^{(j-1)}$ and the continuation vector be $t_j = \tilde{l}_{j-1}/\|\tilde{l}_{j-1}\|$
  3. Compute the residual $r^{(j-1)} = AV_j t_j - \theta^{(j-1)} V_j t_j$
  4. **If** $\|r^{(j-1)}\| <$ TOL then exit
  5. Solve $\left\{ \begin{array}{ll} \text{(Cayley:)} & (A-\mu_j B)V_{j+1}x_j = r^{(j-1)} \\ \text{(Shift-and-invert:)} & (A-\mu_j B)V_{j+1}x_j = BV_j t_j \end{array} \right\}$ for x_j
  6. Compute $\tilde{l}_j$ and $\tilde{k}_j$, the jth column of $\tilde{L}_j$ and $\tilde{K}_j$, respectively.
  7. Compute the Rayleigh quotient $\theta^{(j)} = \tilde{l}_j^H \tilde{k}_j / \tilde{l}_j^H \tilde{l}_j$.

FIG. 3.2. *Inverse iteration computed using inexact transformations*

and has eigenpairs

$$\begin{array}{rcccc} i = & 1 & 2 & 3 & 4 \ \& \ 5 \\ \theta_i^{(5)} = & 1.0000 & 2.0123 & 2.5340 & 3.4618 \pm 6.3095i \\ y_i^{(5)} = & \left[\begin{array}{r} 1.0000 \\ 0.0000 \\ 0.0000 \\ 0.0000 \\ 0.0000 \end{array} \right. & \begin{array}{r} 0.0165 \\ 0.9812 \\ -0.1801 \\ -0.0602 \\ -0.0310 \end{array} & \begin{array}{r} -0.0177 \\ -0.1340 \\ 0.9229 \\ 0.3188 \\ 0.1681 \end{array} & \left. \begin{array}{r} 0.0045 \pm 0.0068i \\ 0.0249 \pm 0.0951i \\ 0.0149 \pm 0.3758i \\ -0.0826 \pm 0.7214i \\ -0.1810 \pm 0.5374i \end{array} \right] \end{array} .$$

Since $f_i^{(5)} = 0$, the true residual has the form $r_i^{(5)} = -E_5 y_i^{(5)}$. For example $\|r_1^{(5)}\| = 6 \cdot 10^{-5}$ but $1 \cdot 10^{-1} < \|r_i^{(5)}\| < 1 \cdot 10^{1}$ for $i > 1$.

This example shows that E_5 is nearly rank deficient and that the desired eigenvector of (A, I) is nearly its nullvector. Therefore, the desired eigenvalue, in this case, $\lambda_1 = 1$, can be computed with a small true residual. It should be noted that the perturbation E_5 is small in the direction of only one eigenspace, which implies that I-RKS is not able to compute several eigenvalues at the same time. This is not the situation when the linear systems are solved with a direct method.

Because of the Galerkin projection, I-RKS computes the eigenpairs of $A + E_5$ exactly after $m = 5$ iterations. In general, however, $r_i^{(5)} \neq 0$, since the inexact Ritz pair is not computed from a Galerkin projection with A. We also remark that $\theta_4^{(5)}$ and $\theta_5^{(5)}$ are complex, which would not be the case with Galerkin projection, since A is a real symmetric matrix. This is in contrast with other iterative eigenvalue solvers, such as the Arnoldi method and the Jacobi-Davidson method.

In the remainder of this section, we discuss why I-RKS works. A link with inverse iteration is established in §3.1, and a formal justification is given in §3.2.

3.1. Inverse Iteration. From Equation (2.2) and the matrix identity (2.1), it follows that

$$V_{j+1}\tilde{h}_j = V_j t_j + (\mu_j - \nu_j)(A - \mu_j B)^{-1} B V_j t_j .$$

Thus, with $\tilde{l}_j = \tilde{L}_j e_j$, we have

$$V_{j+1}\tilde{l}_j = (\mu_j - \nu_j)(A - \mu_j B)^{-1} B V_j t_j , \tag{3.8}$$

and hence $V_{j+1}\tilde{l}_j$ is the linear combination of the columns of V_{j+1} obtained by performing one step of inverse iteration on the vector $V_j t_j$. An inductive argument easily establishes the following property.

LEMMA 3.1. *If* $t_1 = 1$ *and* $t_j = \tilde{l}_{j-1}/\|\tilde{l}_{j-1}\|$ *for* $j > 1$, *then*

$$V_{j+1}\tilde{l}_j = \zeta_j \prod_{i=1}^{j} \left((A - \mu_i B)^{-1} B\right) v_1,$$

where $\zeta_j \equiv \|\tilde{l}_j\| = |\nu_j - \mu_j|\,\|(A-\mu_i B)^{-1} B V_j t_j\|$ *and* v_1 *is the starting vector of RKS.*

Lemma 3.1 indicates how to compute an approximate eigenvalue. If we denote $\tilde{k}_j \equiv \tilde{K}_j e_j$, Equation (2.4) gives the Rayleigh quotient

$$\theta^{(j)} = \frac{(V_{j+1}\tilde{l}_j)^H B^{-1} A (V_{j+1}\tilde{l}_j)}{(V_{j+1}\tilde{l}_j)^H (V_{j+1}\tilde{l}_j)} = \frac{\tilde{l}_j^H \tilde{k}_j}{\tilde{l}_j^H \tilde{l}_j} \tag{3.9}$$

as an estimate of an eigenvalue.

An algorithm for inverse iteration is given in Figure 3.1. The approximate eigenpair on iteration j is $(\theta^{(j)}, y^{(j)} = V_{j+1}\tilde{l}_j/\|\tilde{l}_j\|)$, so we can use the relationships (3.4) with $z^{(j)} = e_j/\|\tilde{l}_j\|$. Recall that we use $\nu_j = \theta^{(j-1)}$ and $V_j t_j = y^{(j-1)}$. The entries $\theta^{(0)}$ and v_1 determine the initial guesses for the eigenpair. We now compare inexact inverse iteration computed via the RKS method using the shift-and-invert and Cayley transformations with an example.

EXAMPLE 3.1. The Olmstead model [26] represents the flow of a layer of viscoelastic fluid heated from below. The equations are

$$\begin{cases} \dfrac{\partial u}{\partial t} &= (1-C)\dfrac{\partial^2 v}{\partial X^2} + C\dfrac{\partial^2 u}{\partial X^2} + Ru - u^3 \\ B\dfrac{\partial v}{\partial t} &= u - v \end{cases}$$

with boundary conditions $u(0) = u(1) = 0$ and $v(0) = v(1) = 0$. Here u represents the speed of the fluid and v is related to viscoelastic forces. The equation was discretized with central differences with gridsize $h = 1/(n/2)$. After the discretization, the equation may be written as $\dot{x} = f(x)$ with $x^T = [u_1, v_1, u_2, v_2, \ldots, u_{n/2}, v_{n/2}]$. The size of the Jacobian matrix $A = \partial f/\partial x$ is $n = 100$. We consider the Jacobian for the parameter values $B = 2$, $C = 0.1$ and $R = 4.7$ for the trivial steady state $[u, v] = 0$. Thus, the interest is in the eigenvalue of largest real part.

We ran the algorithm in Figure 3.1. The linear systems were solved by 20 iterations of Gauss-Seidel starting with a zero initial vector. Since this solver is stationary, the relative residual norm, τ, is almost constant. The initial guess for the eigenvalue was $\theta^{(0)} = 0$. The initial vector for RKS was $v_1 = [\,1, \ldots, 1\,]^T/\sqrt{n}$. The poles μ_j were set equal to 5 for all j. The residuals $r^{(j)}$, $f^{(j)}$ and $S_j z^{(j)}$ are shown in Table 3.1. All three sequences decrease when the Cayley transform is used.

We redid the experiments using the shift-and-invert transformation. The results are also shown in Table 3.1. Both $\|S_j z^{(j)}\|$ and $\|r^{(j)}\|$ stagnate near the same value. Note, however, that $\|f^{(j)}\|$ tends to zero.

Table 3.1 shows that the true residual decreases when the Cayley transformation is used, but stagnates for the shift-and-invert transformation. The following result indicates what occurs under some mild conditions when performing inexact inverse iteration with either the shift-and-invert or the Cayley transformation.

TABLE 3.1
Numerical results for inverse iteration on Example 3.1 using inexact Cayley and shift-and-invert transformations. The table shows the norms of true residual $r^{(j)}$, $S_j z^{(j)}$, and the RKS residual $f^{(j)}$. The norm of $\tilde{l}_j$ is also displayed for the Cayley transformation.

	Cayley				Shift-and-invert		
j	$\|r^{(j)}\|$	$\|S_j z^{(j)}\|$	$\|f^{(j)}\|$	$\|\tilde{l}_j\|$	$\|r^{(j)}\|$	$\|S_j z^{(j)}\|$	$\|f^{(j)}\|$
1	$1\cdot10^{0}$	$8\cdot10^{-1}$	$7\cdot10^{-1}$	5.2	$4\cdot10^{0}$	$5\cdot10^{0}$	$4\cdot10^{-1}$
2	$1\cdot10^{1}$	$1\cdot10^{1}$	$5\cdot10^{-1}$	0.6	$7\cdot10^{-1}$	$7\cdot10^{-1}$	$7\cdot10^{-2}$
3	$1\cdot10^{0}$	$1\cdot10^{0}$	$2\cdot10^{-1}$	1.6	$4\cdot10^{-1}$	$4\cdot10^{-1}$	$3\cdot10^{-2}$
4	$2\cdot10^{-1}$	$2\cdot10^{-1}$	$8\cdot10^{-2}$	1.2	$5\cdot10^{-1}$	$5\cdot10^{-1}$	$1\cdot10^{-2}$
5	$1\cdot10^{-1}$	$9\cdot10^{-2}$	$4\cdot10^{-2}$	1.0	$5\cdot10^{-1}$	$5\cdot10^{-1}$	$7\cdot10^{-3}$
6	$5\cdot10^{-2}$	$5\cdot10^{-2}$	$2\cdot10^{-2}$	1.0	$5\cdot10^{-1}$	$5\cdot10^{-1}$	$4\cdot10^{-3}$
7	$2\cdot10^{-2}$	$2\cdot10^{-2}$	$8\cdot10^{-3}$	1.0	$5\cdot10^{-1}$	$5\cdot10^{-1}$	$2\cdot10^{-3}$
8	$9\cdot10^{-3}$	$9\cdot10^{-3}$	$3\cdot10^{-3}$	1.0	$5\cdot10^{-1}$	$5\cdot10^{-1}$	$1\cdot10^{-3}$
9	$4\cdot10^{-3}$	$4\cdot10^{-3}$	$1\cdot10^{-3}$	1.0	$5\cdot10^{-1}$	$5\cdot10^{-1}$	$5\cdot10^{-4}$
10	$2\cdot10^{-3}$	$1\cdot10^{-3}$	$4\cdot10^{-4}$	1.0	$5\cdot10^{-1}$	$5\cdot10^{-1}$	$3\cdot10^{-4}$
11	$6\cdot10^{-4}$	$6\cdot10^{-4}$	$2\cdot10^{-4}$	1.0	$5\cdot10^{-1}$	$5\cdot10^{-1}$	$1\cdot10^{-4}$
12	$2\cdot10^{-4}$	$2\cdot10^{-4}$	$6\cdot10^{-5}$	1.0	$5\cdot10^{-1}$	$5\cdot10^{-1}$	$7\cdot10^{-5}$
13	$8\cdot10^{-5}$	$8\cdot10^{-5}$	$2\cdot10^{-5}$	1.0	$5\cdot10^{-1}$	$5\cdot10^{-1}$	$4\cdot10^{-5}$
14	$3\cdot10^{-5}$	$3\cdot10^{-5}$	$8\cdot10^{-6}$	1.0	$5\cdot10^{-1}$	$5\cdot10^{-1}$	$2\cdot10^{-5}$
15	$1\cdot10^{-5}$	$1\cdot10^{-5}$	$3\cdot10^{-6}$	1.0	$5\cdot10^{-1}$	$5\cdot10^{-1}$	$1\cdot10^{-5}$

THEOREM 3.2. *Assume that there is an integer $k \leq m$ and value $\gamma > 0$ such that $\|\tilde{l}_j\| \geq \gamma$ for $j > k$. Assume that $\|f^{(j)}\| \leq \rho\|f^{(j-1)}\|$ for $j \geq k$ and ρ a positive number.*

If a Cayley transform is used, then for $j \geq k+1$,

$$\|r^{(j)}\| \leq \left(\rho + \frac{\tau}{\gamma}\right)^{j-k} \|f^{(k)}\| + \left(\frac{\tau}{\gamma}\right)^{j-k} \|r^{(k)}\| \tag{3.10}$$

and when a shift-and-invert transformation is used,

$$\|r^{(j)}\| \leq \rho^{j-k}\|f^{(k)}\| + \frac{\tau}{\gamma}\|B\| \ . \tag{3.11}$$

Here, τ is the relative residual tolerance used for the linear solves (see equations (3.6) and (3.7)).

Proof. With $z^{(j)} = e_j/\|\tilde{l}_j\|$, (3.4) becomes $r^{(j)} = f^{(j)} - s_j/\|\tilde{l}_j\|$. With $\|\tilde{l}_j\| \geq \gamma$, it follows that

$$\begin{aligned} \|r^{(j)}\| &\leq \|f^{(j)}\| + \|s_j\|/\|\tilde{l}_j\| \\ &\leq \|f^{(j)}\| + \|s_j\|/\gamma \ . \end{aligned} \tag{3.12}$$

For the Cayley transform, we prove (3.10) by induction on j. We clearly have that

$$\|r^{(k)}\| \leq \|f^{(k)}\| + \|r^{(k)}\| \ ,$$

which satisfies (3.10) for $j = k$. Suppose that (3.10) holds for some integer $j - 1 \geq k$. From the hypothesis of the theorem, we have that $\|f^{(j)}\| \leq \rho\|f^{(j-1)}\| \leq \cdots \leq \rho^{j-k}\|f^{(k)}\|$. Combining this with equations (3.12) and (3.6) results in

$$\|r^{(j)}\| \leq \|f^{(j)}\| + \|s_j\|/\gamma \leq \rho^{j-k}\|f^{(k)}\| + \tau/\gamma\|r^{(j-1)}\|.$$

Using our inductive hypothesis on $\|r^{(j-1)}\|$ gives

$$\begin{aligned}\|r^{(j)}\| &\leq (\rho^{j-k} + \tau/\gamma(\rho+\tau/\gamma)^{j-k-1})\|f^{(k)}\| + (\tau/\gamma)^{j-k}\|r^{(k)}\| \\ &\leq (\rho+\tau/\gamma)^{j-k}\|f^{(k)}\| + (\tau/\gamma)^{j-k}\|r^{(k)}\| \ ,\end{aligned}$$

from which (3.10) follows. For shift-and-invert, (3.11) follows from (3.12) and (3.7), which completes the proof. □

The value ρ denotes the convergence ratio of the exact RKS process on $(A + E_m, B)$. This process converges when $\rho < 1$. The theorem shows that if $\rho + \tau/\gamma < 1$, inexact inverse iteration with the Cayley transformation converges. Since $\rho+\tau/\gamma \geq \rho$, inexact inverse iteration, in general, converges slower than the exact one. With the shift-and-invert transformation, although the term $\|f^{(j)}\|$ may converge to zero, $\|r^{(j)}\|$ may stagnate, since the contribution to the true residual, coming from s_j, is constant. We also remark that when a direct method is used for the linear system of equations, τ is a multiple of machine precision. Thus, whether a shift-and-invert or Cayley transformation is used, the true residual $\|r^{(j)}\|$ decreases at a rate proportional to ρ.

For the exact Cayley transformation, we have

$$V_{j+1}\tilde{h}_j = (A - \mu_j B)^{-1} f^{(j-1)}$$

and $\tilde{l}_j = \tilde{h}_j - \tilde{t}_j$ and $\|t_j\| = 1$. Hence, we have

$$1 - \delta_j \leq \|\tilde{l}_j\| \leq 1 + \delta_j \quad \text{where} \quad \delta_j = \|(A - \mu_j B)^{-1}\| \ \|f^{(j-1)}\| \ .$$

Thus, $\gamma = \max(0, \min_{j \geq k} 1 - \delta_j)$. If $f^{(j-1)}$ converges to zero and $\|(A - \mu_j B)^{-1}\|$ remains modest, $1 \pm \delta_j$ tends to one for increasing j. Computation reveals that quite often $\|\tilde{l}_j\| \approx 1$ after a very small number of iterations. This also holds for inexact inverse iteration, since it can be seen as exact inverse iteration applied to $(A+E_m, B)$, as is the case in Table 3.1. Hence, for large enough k, $\gamma \approx 1$, such that the convergence rate of inverse iteration for large k using the Cayley transform can be estimated by $\rho + \tau$.

3.2. Inexact Rational Krylov. We now formally discuss the algorithm listed in Figure 3. The Ritz vector $y^{(j)} = V_{j+1}\tilde{L}_j z^{(j)}$, computed as explained in § 2.2.1, is a linear combination of

$$V_{i+1}\tilde{l}_i = \zeta_i (A - \mu_i B)^{-1} B y^{(i-1)} \quad \text{with} \quad y^{(i-1)} = V_i \tilde{L}_{i-1} z^{(i-1)} \quad \text{for} \quad i = 1, \ldots, j \ .$$

We observed from the numerical experiments that the last component of $z^{(j)}$ is large compared with the initial components. The explanation rests with the fact that $V_{j+1}\tilde{l}_j$ is the improvement of the previous Ritz vector by inverse iteration, thus giving the best approximation of the desired eigenvector among the $V_{i+1}\tilde{l}_i$'s.

The inexact Ritz pairs $(\theta^{(i)}, y^{(i)})$ lead to true residuals $r^{(i)}$. If the Cayley transform is used, the Cayley residual on iteration i satisfies $\|s_i\| \leq \tau \|r^{(i-1)}\|$. The true residual on the jth iteration is decomposed as $r^{(j)} = f^{(j)} - S_j z^{(j)}$, where

$$\|S_j z^{(j)}\| \leq \sum_{i=1}^{j} \|s_i\| \ |e_i^T z^{(j)}|$$

gives an upper bound to $\|S_j z^{(j)}\|$. In the right-hand side, $\|s_i\|$ is independent of j and can be quite large for small i. Since $|e_i^T z^{(j)}|$ forms a decreasing sequence for increasing j, we have a decreasing sequence $\|S_j z^{(j)}\|$.

TABLE 3.2
Numerical results for the Olmstead model of Example 3.2. The table shows the order of accuracy for the residual norm of the rightmost Ritz pair, the norm of $S_j z^{(j)}$, and the first four components of $z^{(j)}$.

j	$\|r_j\|$	$\|S_j z^{(j)}\|$	$\|e_1^T z^{(j)}\|$	$\|e_2^T z^{(j)}\|$	$\|e_3^T z^{(j)}\|$	$\|e_4^T z^{(j)}\|$
1	$6\cdot10^{-1}$	$1\cdot10^{-1}$	$2\cdot10^{-1}$			
2	$2\cdot10^{0}$	$3\cdot10^{-1}$	$5\cdot10^{-1}$	$3\cdot10^{0}$		
3	$2\cdot10^{-2}$	$1\cdot10^{-2}$	$5\cdot10^{-3}$	$6\cdot10^{-1}$	1	
4	$2\cdot10^{-2}$	$1\cdot10^{-2}$	$3\cdot10^{-3}$	$6\cdot10^{-1}$	1	2
5	$8\cdot10^{-4}$	$7\cdot10^{-4}$	$1\cdot10^{-4}$	$4\cdot10^{-2}$	$6\cdot10^{-2}$	$3\cdot10^{-1}$
6	$3\cdot10^{-4}$	$3\cdot10^{-4}$	$2\cdot10^{-5}$	$2\cdot10^{-2}$	$3\cdot10^{-2}$	$3\cdot10^{-1}$
7	$2\cdot10^{-5}$	$2\cdot10^{-5}$	$2\cdot10^{-6}$	$1\cdot10^{-3}$	$2\cdot10^{-3}$	$2\cdot10^{-2}$
8	$5\cdot10^{-7}$	$5\cdot10^{-7}$	$7\cdot10^{-8}$	$2\cdot10^{-5}$	$5\cdot10^{-5}$	$1\cdot10^{-3}$
9	$3\cdot10^{-8}$	$3\cdot10^{-8}$	$4\cdot10^{-9}$	$2\cdot10^{-6}$	$2\cdot10^{-6}$	$6\cdot10^{-5}$

EXAMPLE 3.2. We now discuss an example for which $e_i^T z^{(j)}$ and $S_j z^{(j)}$ tend to zero in the I-RKS method. The matrix arises from the same problem as in Example 3.1, but now $n = 200$. We ran Algorithm I-RKS from Figure 3 with fixed $\mu_j = 5$, starting with vector $v_1 = [1, \cdots, 1]^T/\sqrt{n}$. The linear systems were solved by GMRES preconditioned by ILU. The number of iterations of GMRES was determined by the relative error tolerance, which was selected as $\tau = 10^{-4}$. Table 3.2 shows the residual norm and the norm of the error term $S_j z^{(j)}$. Both $\|S_j z^{(j)}\|$ and $\|r^{(j)}\|$ tend to zero. For large j, $\|S_j z^{(j)}\| \approx \|r^{(j)}\|$. This is the case because $f^{(j)}$ converges more rapidly to zero than $S_j z^{(j)}$. Table 3.2 also illustrates the fact that $e_i^T z^{(j)}$ decreases for a fixed i and increasing j.

4. A Numerical Example. This example illustrates the use of inexact rational Krylov methods for the solution of a generalized eigenvalue problem. We also make a comparison between inexact inverse iteration with the Cayley transform and I-RKS.

The simulation of flow of a viscous fluid with a free surface on a tilted plane, leads, with a finite element approach, to an eigenvalue problem $Ax = Bx\lambda$ with $A, B \in \boldsymbol{R}^{536\times536}$ and B a singular matrix. The computation of the eigenvalue nearest -10 is the objective. Since our theory is valid only for nonsingular B, we interchange the role of A and B by computing the eigenvalue $\gamma = \lambda^{-1}$ of $Bx = Ax\gamma$ nearest $\mu = -10^{-1}$.

The fact that B is singular implies that $\gamma = 0$ is an eigenvalue. It has been shown that the presence of this eigenvalue can disturb the calculation of a nonzero eigenvalue when the spectral transformation Lanczos method [13, 25], the shift-invert Arnoldi method [28, 22], or the rational Krylov method [8] are used. One way to reduce the impact of $\gamma = 0$ is to start the I-RKS method with an initial vector v_1 that is poor in the eigenspace corresponding to $\gamma = 0$ [25]. This can be achieved by selecting $v_1 = (B - \mu A)^{-1}Bv$ with v arbitrary.

The eigenvalue γ nearest -0.1 was computed by use of I-RKS (Fig. 3) with fixed pole $\mu_j = -0.1$. The linear systems were solved by GMRES preconditioned with `ILUT(lfil=40,tol=1.e-3)` [32] with $\tau = 10^{-4}$. The initial vector v_1 was computed from the system $(B - \mu A)v_1 = Bv$ with $v = [\ 1, \cdots, 1\ ]^T$ using the GMES-ILUT solver. The algorithm was stopped when $\|r_1^{(j)}\| \le \text{TOL} = 10^{-13}$.

The numerical results are shown in Table 4.1 for inexact rational Krylov (I-RKS)

TABLE 4.1

Numerical results for the tilted plane problem from § 4. The methods used are inexact rational Krylov (I-RKS) and inverse iteration with the Cayley transform. On iteration j, $\theta^{(j)}$ is the inexact Ritz value, s_j the Cayley residual, and $g^{(j)} = (\tilde{K}_j - \theta^{(j)}\tilde{L}_j)z^{(j)}$.

	I-RKS (Fig. 3)				Inverse Iteration (Fig. 3.1)			
j	$(\theta_1^{(j)})^{-1}$	$\|s_j\|$	$\|r_1^{(j)}\|$	$\|g_1^{(j)}\|$	$(\theta_1^{(j)})^{-1}$	$\|s_j\|$	$\|r^{(j)}\|$	$\|g^{(j)}\|$
1	-9.40554	$5\cdot10^{-9}$	$1\cdot10^{-6}$	$3\cdot10^{-3}$	-9.40554	$5\cdot10^{-9}$	$1\cdot10^{-6}$	$2\cdot10^{-3}$
2	-9.48481	$1\cdot10^{-10}$	$3\cdot10^{-8}$	$5\cdot10^{-6}$	-9.49928	$1\cdot10^{-10}$	$2\cdot10^{-7}$	$4\cdot10^{-4}$
3	-9.48825	$1\cdot10^{-12}$	$1\cdot10^{-9}$	$8\cdot10^{-8}$	-9.48705	$2\cdot10^{-11}$	$4\cdot10^{-9}$	$6\cdot10^{-6}$
4	-9.48831	$6\cdot10^{-13}$	$2\cdot10^{-11}$	$1\cdot10^{-9}$	-9.48845	$3\cdot10^{-12}$	$4\cdot10^{-9}$	$8\cdot10^{-7}$
5	-9.48832	$2\cdot10^{-15}$	$6\cdot10^{-13}$	$6\cdot10^{-11}$	-9.48831	$3\cdot10^{-13}$	$5\cdot10^{-10}$	$8\cdot10^{-8}$
6	-9.48832	$5\cdot10^{-17}$	$2\cdot10^{-14}$	$1\cdot10^{-12}$	-9.48832	$5\cdot10^{-14}$	$5\cdot10^{-11}$	$8\cdot10^{-9}$
7					-9.48832	$4\cdot10^{-14}$	$5\cdot10^{-12}$	$6\cdot10^{-10}$
8					-9.48832	$5\cdot10^{-15}$	$4\cdot10^{-13}$	$3\cdot10^{-11}$
9					-9.48832	$3\cdot10^{-17}$	$3\cdot10^{-14}$	$6\cdot10^{-12}$

and inexact inverse iteration using the Cayley transform. First, note that $\|f^{(j)}\| \leq \|A\|\|g^{(j)}\|$, so $\|g^{(j)}\|$ does not measure the RKS residual (see also (2.7)). Also note that for both I-RKS and inverse iteration, the sequences $\|r^{(j)}\|$, $\|s_j\|$ and $\|g^{(j)}\|$ decrease. Both methods converge to $\lambda = \gamma^{-1} \approx -9.486$. Finally, note that I-RKS is faster than inverse iteration.

5. A Connection with the Jacobi–Davidson Method. Based upon the recent work of Sleijpen and van der Vorst [36] which investigated a new method for standard eigenvalue problems ($B = I$) that only uses approximate linear systems solutions. The two manuscripts [15, 34] extend the method for generalized eigenvalue problems. We now proceed to show a connection between RKS and Jacobi–Davidson when the linear systems are solved exactly.

Consider the linear system

$$(A - \mu_j B)w = (A - \tilde{\nu}_j B)p_j\,, \tag{5.1}$$

where $p_j = V_j\tilde{L}_{j-1}z^{(j-1)}$ is a Ritz vector of interest. This amounts to selecting the jth continuation vector $t_j = \tilde{L}_{j-1}z^{(j-1)}$ as in the Algorithm I-RKS in Figure 3 with associated Ritz value

$$\tilde{\nu}_j = \frac{p_j^H A p_j}{p_j^H B p_j}.$$

The right-hand side in (5.1) is then the residual of the eigenpair $(\tilde{\nu}_j, p_j)$ and is orthogonal to p_j. Since we are interested in expanding our search space (the span of the columns of V_j), multiply both sides of Equation (5.1) by the projector $I - Bp_jp_j^H/(p_j^H Bp_j)$. Using the fact that $(A - \tilde{\nu}_j B)p_j \perp p_j$, results in

$$(I - \frac{Bp_jp_j^H}{p_j^H Bp_j})(A - \mu_j B)w = (A - \tilde{\nu}_j B)p_j\,.$$

Since $p_j \in \mathcal{R}(V_j)$, the component of w in the direction of p_j does not play a role when w is added to the subspace $\mathcal{R}(V_j)$. Thus, we are interested in finding only the

component of w orthogonal to p_j and so the linear system

$$(I - \frac{Bp_jp_j^H}{p_j^HBp_j})(A-\mu_jB)(I-\frac{p_jp_j^H}{p_j^Hp_j})w = (A-\tilde{\nu}_jB)p_j \tag{5.2}$$

is solved instead. The Jacobi–Davidson method calls Equation (5.2) the *correction equation*. We may rewrite Equation (5.2) with $d \equiv (I - p_jp_j^H/(p_j^Hp_j))w \perp p_j$ as

$$(A-\mu_jB)d = \varepsilon_jBp_j + (A-\tilde{\nu}_jB)p_j = (A-(\tilde{\nu}_j+\varepsilon_j)B)p_j\,. \tag{5.3}$$

The orthogonality of p_j and d leads to

$$\varepsilon_j = \frac{p_j^H(A-\mu_jB)^{-1}(A-\tilde{\nu}_jB)p_j}{p_j^H(A-\mu_jB)^{-1}Bp_j}.$$

Choosing the pole $\nu_j \equiv \tilde{\nu}_j + \varepsilon_j$ gives a relationship between the Jacobi–Davidson and RKS methods when Cayley transformations are used. In words, Jacobi–Davidson is a RKS method with a specific Cayley transformation. The difference is that in the correction equation (5.2) of the Jacobi–Davidson method, p_j is orthogonal to the right-hand side. This is not the case for the RKS method with Cayley transformation defined by Equation (5.3).

The solution of the linear system (5.2) leads to quadratic convergence when $\mu_j = \tilde{\nu}_j$. Theorem 3.2 in [34] establishes this result under some mild conditions while Appendix A in [34] demonstrates a connection with Newton's method.

6. Conclusions. This paper demonstrated the use of iterative linear system solvers in Ruhe's rational Krylov sequence method. The analysis of the convergence of inexact inverse iteration showed the importance of the use of the Cayley transformation instead of the usual shift-and-invert transformation.

A theoretical link between the inexact rational Krylov method and the Jacobi-Davidson method was made by observing a connection between the correction equation and the Cayley transformation.

We called the eigenpairs computed by I-RKS inexact Ritz pairs, since they are Ritz pairs for a perturbed RKS method. The classical properties of Galerkin projection are lost due to this inexactness. The fact that I-RKS solves a perturbed problem with small perturbations in the desired eigendirections motivates the application of the RKS techniques developed for the process using exact linear solves. These techniques include use of complex poles and zeros for real A and B [30], harmonic Ritz pairs, deflation and purging [31], and the implicit application of a rational filter [8]. The practical advantage of the inexact rational Krylov method is that the cheaply computed matrices $\tilde{L}_m$ and $\tilde{K}_m$ are used to compute the Ritz pairs. The Jacobi-Davidson method requires the explicit formation of $V_m^HAV_m$ and V_mBV_m.

A. Relation between the Shift-and-Invert and Cayley Transformations. In this appendix, we prove that the use of the shift-and-invert or the Cayley transformations lead to similar recurrence relations.

LEMMA A.1. *If in Step 3 of Algorithm 2, we use* $T_j^C = (A-\mu_jB)^{-1}(A-\nu_jB)$ *on the jth iteration, and we obtain the relation*

$$AV_{m+1}\tilde{L}_m = BV_{m+1}\tilde{K}_m\,,$$

then the use of $T_j^{SI} = (A-\mu_jB)^{-1}B$ *leads to*

$$AV_{m+1}\tilde{L}_m(M_m-N_m)^{-1} = BV_{m+1}\tilde{K}_m(M_m-N_m)^{-1}$$

with $N_m = \mathrm{diag}(\nu_1, ..., \nu_m)$ *and* $M_m = \mathrm{diag}(\mu_1, ..., \mu_m)$.

Proof. From $T_j^C V_j t = V_j t + (\mu_j - \nu_j) T_j^{SI} V_j t$, we have

$$T_j^C V_j t_j = V_{j+1} h_j \tag{A.1}$$

$$V_j t_j + (\mu_j - \nu_j) T_j^{SI} V_j t_j = V_{j+1} \tilde{h}_j$$

$$T_j^{SI} V_j t_j = V_{j+1} (\tilde{h}_j - \tilde{t}_j)(\mu_j - \nu_j)^{-1} \ . \tag{A.2}$$

Multiplying (A.1) and (A.2) on the left by $(A - \mu_j B)$ and putting together the equations for $j = 1, ..., m$ produce the relations

$$A V_{m+1} (\tilde{H}_m - \tilde{T}_m) = B V_{m+1} (\tilde{H}_m M_m - \tilde{T}_m N_m)$$

and

$$\begin{aligned} A V_{m+1} (\tilde{H}_m - \tilde{T}_m)(M_m - N_m)^{-1} &= B V_{m+1} ((\tilde{H}_m - \tilde{T}_m)(M_m - N_m)^{-1} M_m - \tilde{T}_m) \\ &= B V_{m+1} (\tilde{H}_m M_m - \tilde{T}_m N_m)(M_m - N_m)^{-1}, \end{aligned}$$

respectively. Noting that $\tilde{L}_m = \tilde{H}_m - \tilde{T}_m$ and $\tilde{K}_m = \tilde{H}_m M_m - \tilde{T}_m N_m$ completes the proof. □

REFERENCES

[1] E. Anderson, Z. Bai, C. Bischof, J. Demmel, J. Dongarra, J. Du Croz, A. Greenbaum, S. Hammarling, A. McKenney, S. Ostrouchov, and D. Sorensen. *LAPACK users' guide*. SIAM, Philadelphia, PA, 1995.

[2] W. E. Arnoldi. The principle of minimized iterations in the solution of the matrix eigenvalue problem. *Quart. Appl. Math.*, 9:17–29, 1951.

[3] R. Barrett, M. Berry, T. Chan, J. Demmel, J. Donato, J. Dongarra, V. Eijkhout, R. Pozo, C. Romine, and H. van der Vorst. *Templates for the solution of linear systems: Building blocks for iterative methods*. SIAM, Philadelphia, PA, 1994.

[4] A. M. Bruaset. *A survey of preconditioned iterative methods*. Pitman Research Notes in Mathematics Series. Longman Scientific & Technical, Harlow Essex, UK, 1995.

[5] J. W. Daniel, W. B. Gragg, L. Kaufman, and G. W. Stewart. Reorthogonalization and stable algorithms for updating the Gram-Schmidt QR factorization. *Math. Comp.*, 30:772–795, 1976.

[6] T. A. Davis and I. S. Duff. An unsymmetric-pattern multifrontal method for sparse LU factorization. *SIAM J. Matrix Anal. Applic.*, to appear. (also University of Florida technical report TR-94-038).

[7] Timothy A. Davis and Iain S. Duff. A combined unifrontal/multifrontal method for unsymmetric sparse matrices. Technical Report TR-95-020, Computer and Information Science and Engineering Department, University of Florida, September 1995.

[8] G. De Samblanx, K. Meerbergen, and A. Bultheel. The implicit application of a rational filter in the rks method. Technical Report TW239, Department of Computer Science, K.U.Leuven, Heverlee, Belgium, 1996.

[9] James W. Demmel, Stanley C. Eisenstat, John R. Gilbert, Xiaoye S. Li, and Joseph W. H. Liu. A supernodal approach to sparse partial pivoting. Technical Report CSD-95-883, Department of Computer Science, University of California, Berkeley, California, September 1995.

[10] I. S. Duff. ME28: A sparse unsymmetric linear equation solver for complex equations. *ACM Transactions on Mathematical Software*, 7(4):505–511, December 1981.

[11] I. S. Duff and J. K. Reid. The design of MA48, a code for direct solution of sparse unsymmetric linear systems of equations. *ACM Transactions on Mathematical Software*, 22(2):187–226, June 1996.

[12] I. S. Duff and J. A. Scott. The design of a new frontal code for solving sparse unsymmetric systems. *ACM Transactions on Mathematical Software*, 22(1):30–45, March 1996.

[13] T. Ericsson. A generalised eigenvalue problem and the Lanczos algorithm. In J. Cullum and R. A. Willoughby, editors, *Large Scale Eigenvalue Problems*, pages 95–119. Elsevier Science Publishers BV, 1986.

[14] T. Ericsson and A. Ruhe. The spectral transformation Lanczos method for the numerical solution of large sparse generalized symmetric eigenvalue problems. *Math. Comp.*, 35:1251–1268, 1980.

[15] 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. Technical Report Preprint 941, Department of Mathematics, Utrecht University, Utrecht, The Netherlands, January 1996.

[16] R. W. Freund, G. H. Golub, and N. M. Nachtigal. Iterative solution of linear systems. In A. Iserles, editor, *Acta Numerica 1992*, pages 57–100. Cambridge University Press, 1992.

[17] Roland W. Freund and Noel M. Nachtigal. QMRPACK: A package of QMR algorithms. *ACM Transactions on Mathematical Software*, 22(1):46–77, March 1996.

[18] G. Golub and C. Van Loan. *Matrix computations*. The Johns Hopkins University Press, 2nd edition, 1989.

[19] R. B. Lehoucq, D. C. Sorensen, P. Vu, and C. Yang. ARPACK: An implementation of the implicitly re-started Arnoldi iteration that computes some of the eigenvalues and eigenvectors of a large sparse matrix. Available from `netlib@ornl.gov` under the directory scalapack, 1995.

[20] K. Meerbergen and D. Roose. Matrix transformations for computing rightmost eigenvalues of real nonsymmetric matrices. *IMA J. Numer. Anal.*, 16:297–346, 1996.

[21] K. Meerbergen and D. Roose. The restarted Arnoldi method applied to iterative linear system solvers for computation of rightmost eigenvalues. *SIAM J. Matrix Anal. Applic.*, 1996. Accepted for publication.

[22] K. Meerbergen and A. Spence. Implicitly restarted Arnoldi and purification for the shift-invert transformation. *Math. Comp.*, 1997. To appear.

[23] R. B. Morgan. Computing interior eigenvalues of large matrices. *Linear Alg. Appl.*, 154–156:289–309, 1991.

[24] R. B. Morgan and Min Zeng. Estimates for interior eigenvalues of large nonsymmetric matrices. 1996. Preprint.

[25] B. Nour-Omid, B. N. Parlett, T. Ericsson, and P. S. Jensen. How to implement the spectral transformation. *Math. Comp.*, 48:663–673, 1987.

[26] W. E. Olmstead, W. E. Davis, S. H. Rosenblat, and W. L. Kath. Bifurcation with memory. *SIAM J. Appl. Math.*, 40:171–188, 1986.

[27] C. Paige, B. N. Parlett, and H. A. van der Vorst. Approximate solutions and eigenvalue bounds from Krylov subspaces. *Num. Lin. Alg. Appl.*, 2:115–133, 1995.

[28] B. Philippe and M. Sadkane. Improving the spectral transformation block Arnoldi method. In P. S. Vassilevski and S. D. Margenov, editors, *Second IMACS Symposium on Iterative Methods in Linear Algebra*, volume 3 of *IMACS Series in Computational and Applied Mathematics*, pages 57–63. IMACS Symposium on Iterative Methods in Linear Algebra, 1996.

[29] A. Ruhe. Rational Krylov sequence methods for eigenvalue computation. *Linear Alg. Appl.*, 58:391–405, 1984.

[30] A. Ruhe. The Rational Krylov algorithm for nonsymmetric eigenvalue problems, III: Complex shifts for real matrices. *BIT*, 34:165–176, 1994.

[31] A. Ruhe. Rational Krylov, a practical algorithm for large sparse nonsymmetric matrix pencils. Technical Report UCB/CSD-95-871, Computer Science Division, University of California, Berkeley, CA, 1995.

[32] Y. Saad. SPARSKIT : A basic tool kit for sparse matrix computations. Technical Report 90-20, Research Institute for Advanced Computer Science, NASA Ames Research Center, Moffet Field, CA, 1990.

[33] Y. Saad and M. H. Schultz. GMRES: A generalized minimal residual algorithm for solving nonsymmetric linear systems. *SIAM J. Sci. Statist. Comput.*, 7:856–869, 1986.

[34] G. L. G. Sleijpen, G. L. Booten, D. R. Fokkema, and H. A. van der Vorst. Jacobi Davidson type methods for generalized eigenproblems and polynomial eigenproblems: Part I. Preprint nr. 923, Department of Mathematics, Universiteit Utrecht, Utrecht, The Netherlands, 1995.

[35] G. L. G. Sleijpen, D. R. Fokkema, and H. A van der Vorst. BiCGstab(ℓ) and other hybrid Bi-CG methods. *Numer. Algor.*, 7:75–109, 1994.

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

[37] D. C. Sorensen. Implicit application of polynomial filters in a k-step Arnoldi method. *SIAM J. Matrix Analysis and Applications*, 13(1):357–385, January 1992.

[38] Daniel B. Szyld. Criteria for combining inverse and Rayleigh quotient iteration. *SIAM J. Numer. Anal.*, 25:1369–1375, 1988.