NUMERICAL LINEAR ALGEBRA WITH APPLICATIONS
Numer. Linear Algebra Appl., **6,** 281–293 (1999)

Ritz and Harmonic Ritz Values and the Convergence of FOM and GMRES

Serge Goossens* and Dirk Roose

Department of Computer Science, Katholieke Universiteit Leuven, Celestijnenlaan 200A, B-3001 Heverlee, Belgium. E-mail: Serge.Goossens@cs.kuleuven.ac.be

The Ritz and harmonic Ritz values are approximate eigenvalues, which can be computed cheaply within the FOM and GMRES Krylov subspace iterative methods for solving non-symmetric linear systems. They are also the zeros of the residual polynomials of FOM and GMRES, respectively. In this paper we show that the Walker–Zhou interpretation of GMRES enables us to formulate the relation between the harmonic Ritz values and GMRES in the same way as the relation between the Ritz values and FOM. We present an upper bound for the norm of the difference between the matrices from which the Ritz and harmonic Ritz values are computed. The differences between the Ritz and harmonic Ritz values enable us to describe the breakdown of FOM and stagnation of GMRES.

KEY WORDS Krylov subspaces; FOM; GMRES; Ritz values; harmonic Ritz values

1. Introduction

Many iterative methods for the solution of linear systems and the computation of (selected) eigenvalues make use of Krylov subspaces. Given a real non-singular matrix $A \in \mathbb{R}^{n\times n}$ and a vector $r_0 \in \mathbb{R}^n$, the Krylov subspaces $\mathcal{K}_m(A, r_0) = \text{span}\{r_0, Ar_0, \ldots, A^{(m-1)}r_0\}$ for $m = 1, 2, \ldots, n$ form a nested sequence of subspaces. Krylov subspace iterative methods such as the full orthogonalization method (FOM), also known as Arnoldi's method for linear systems [1], and the generalized minimal residual method (GMRES) proposed by

* Correspondence to S. Goossens, Department of Computer Science, Katholieke Universiteit Leuven, Celestijnenlaan 200A, B-3001 Heverlee, Belgium

Contract/grant sponsor: Belgian State, Prime Minister's Office for Science, Technology and Culture; contract/grant number: IUAP P4/2. Contract/grant sponsor: Research Council K.U. Leuven; contract/grant number: OT/94/16, Contract/grant sponsor: Flemish Institute for the Promotion of Scientific and Technological Research in Industry.

CCC 1070–5325/99/040281–13 $17.50
Received 29 September 1993
Revised 7 December 1993

Saad and Schultz [22] are well-known methods for the approximate solution of large sparse non-symmetric linear systems. The classical Krylov subspace methods are described in several books, see e.g. [2,4,11,21]. An overview of the state of the art on polynomial-based iteration methods for symmetric linear systems is given in a book by Fischer [7].

We briefly describe FOM and GMRES, two well-known Krylov subspace iterative methods for solving non-symmetric linear systems. The Arnoldi process [1] computes an orthonormal basis for $\mathcal{K}_m(A, r_0)$ and forms the computational kernel of FOM and GMRES. The Arnoldi basis vectors $V_m = \begin{pmatrix} v_1 & v_2 & \dots & v_m \end{pmatrix} \in \mathbb{R}^{n\times m}$ form an orthogonal matrix. In the orthogonalization process the scalars $h_{i,j}$ are computed so that the square upper Hessenberg matrix $H_m = \left(h_{i,j}\right) \in \mathbb{R}^{m\times m}$ satisfies the fundamental relation

$$AV_m = V_m H_m + h_{m+1,m} v_{m+1} e_m^H = V_{m+1}\bar{H}_m \tag{1.1}$$

The rectangular upper Hessenberg matrix $\bar{H}_m \in \mathbb{R}^{(m+1)\times m}$ is the square upper Hessenberg matrix H_m supplemented with an extra row $\begin{pmatrix} 0 & \dots & 0 & h_{m+1,m} \end{pmatrix}$.

A theoretical comparison of FOM and GMRES is given by Brown [3] and more recently by Cullum and Greenbaum [5]. In this paper we assume that the initial guess x_0 to the solution of the linear system $Ax = b$ is zero and that the system has been scaled so that the initial residual has unit length. Further we assume that the dimension of $\mathcal{K}_m(A, r_0)$ is m. If this is not the case, then this subspace is invariant under A and the approximations discussed in this paper are exact. Both FOM and GMRES select an approximate solution from $\mathcal{K}_m(A, r_0)$ which can be written as $x = \varphi_{m-1}(A)r_0$, where $\varphi_{m-1}(\lambda) = \gamma_{m-1}\lambda^{m-1} + \dots + \gamma_1\lambda + \gamma_0 \in \mathbb{P}_{m-1}$ is a real polynomial of degree $(m-1)$. The residual corresponding to this approximate solution is

$$r = b - Ax = (I - A\varphi_{m-1}(A))\, r_0 = \tilde{\varphi}_m(A)r_0 \in \mathcal{K}_{m+1}(A, r_0) \tag{1.2}$$

FOM selects the approximate solution such that the residual $r^{\mathrm{FOM}} = \tilde{\varphi}_m^{\mathrm{FOM}}(A)r_0$ is orthogonal to $\mathcal{K}_m(A, r_0)$:

$$V_m^H\,(r_0 - AV_m y_m) = 0 \Leftrightarrow H_m y_m = e_1 \tag{1.3}$$

where e_j denotes column j of the identity matrix. GMRES selects the approximate solution which minimises the norm of the residual $\|r^{\mathrm{GMRES}}\|_2$:

$$\|r^{\mathrm{GMRES}}\|_2 = \|V_{m+1}^H\,(r_0 - AV_m y_m)\,\|_2 = \|e_1 - \bar{H}_m y_m\|_2 \tag{1.4}$$

This is equivalent to the requirement that the residual is orthogonal to the image of the Krylov subspace $A\mathcal{K}_m(A, r_0)$. The overdetermined linear system can be solved using the normal equations, where the vector $f_m = H_m^{-H} e_m$ is only defined when H_m is non-singular:

$$\bar{H}_m^H \bar{H}_m y_m = H_m^H e_1 \Leftrightarrow \left(H_m + h_{m+1,m}^2 f_m e_m^H\right) y_m = e_1 \tag{1.5}$$

The eigenvalues of H_m are called Ritz values and approximate the eigenvalues of A. We show in Section 2 that these Ritz values are the zeros of the FOM residual polynomial. In Section 3 we give an outline of the Simpler GMRES algorithm of Walker and Zhou because the upper triangular matrix which is computed in this algorithm is used in the subsequent theory. We define a transformation matrix which is also used in the proofs. In Section 4 we briefly recall the definition of harmonic Ritz values given by Sleijpen and Van der Vorst (section 5.1 in [24]) and the fact that they are eigenvalue approximations according to the minimal residual criterion. We present an upper bound for the norm of the difference between the matrices from which the Ritz and harmonic Ritz values are computed. We prove in Section 5 that the zeros of the GMRES residual polynomial are the harmonic Ritz values, based on the simpler GMRES algorithm. In Section 6 we relate the breakdown of FOM and stagnation of GMRES to the differences between the Ritz and harmonic Ritz values. Numerical results are presented for a problem which nearly causes breakdown and stagnation. In Section 7 we give numerical results for a convection–diffusion problem. Some remarks concerning related work conclude this paper.

2. Ritz values and FOM residual polynomial

The classical Galerkin approach for computing approximate eigenpairs has been discussed by several authors, see e.g. [18,20]. An approximate eigenvector $x = V_m y_m$ is sought in $\mathcal{K}_m(A, r_0)$ such that the residual of the eigenvalue equation is orthogonal to $\mathcal{K}_m(A, r_0)$

$$(Ax - \mu x) \perp \mathcal{K}_m(A, r_0) \Leftrightarrow V_m^H (AV_m y_m - \mu V_m y_m) = 0 \tag{2.1}$$

The approximate eigenvalues can thus be computed from the matrix $H_m = V_m^H A V_m$.

Definition 2.1. *The Ritz values $\vartheta_i^{(m)}$ are the eigenvalues of the Hessenberg matrix H_m.*

Hence the Ritz values are the well-known ‘Arnoldi eigenvalue estimates’.

The FOM residual polynomial $\tilde{\varphi}_m^{\mathrm{FOM}}(\lambda)$ is a multiple of the characteristic polynomial of H_m. This implies that the Ritz values are the zeros of the FOM residual polynomial. We restrict ourselves here to formulating the lemmas and the theorem. This theorem is a straightforward generalization of a result by Paige *et al*. [17] for symmetric matrices. The proofs can be found in [10].

Lemma 2.1. *Let an arbitrary polynomial $\chi_m(\lambda) = \gamma_m \lambda^m + \cdots + \gamma_1 \lambda + \gamma_0$ have strict degree m, then $\chi_m(A) r_0 \in \mathcal{K}_{m+1}(A, r_0)$ and $\chi_m(A) r_0 \notin \mathcal{K}_m(A, r_0)$ and we have that $\chi_m(A) r_0 = V_m \chi_m(H_m) e_1 + \gamma_m \zeta_m v_{m+1}$, where $\zeta_m = h_{m+1,m} \ldots h_{3,2} h_{2,1}$.*

Lemma 2.2. *All non-zero vectors in $\mathcal{K}_{m+1}(A, r_0)$ that are orthogonal to $\mathcal{K}_m(A, r_0)$ can be written as $\alpha \psi_m(A) r_0$ for some non-zero $\alpha \neq 0$, where $\psi_m(\lambda) = \det(\lambda I - H_m)$ is the characteristic polynomial of H_m.*

Theorem 2.1. *The FOM residual polynomial is a multiple of the characteristic polynomial of the Hessenberg matrix H_m.*

Hence the FOM residual polynomial $\tilde{\varphi}_m^{\mathrm{FOM}}(\lambda)$ is uniquely defined by the fact that the m Ritz values $\vartheta_i^{(m)}$ are its zeros and by the normalization $\tilde{\varphi}_m^{\mathrm{FOM}}(0) = 1$. Saad [19] has proved that the characteristic polynomial of H_m is the polynomial that minimizes $\|\psi(A) v_1\|_2$ over all monic polynomials.

3. Simpler GMRES and the transformation matrix

Walker and Zhou [26] pointed out that by starting the Arnoldi process with Ar_0 instead of r_0 a simpler GMRES is obtained which does not require the factorization of an upper Hessenberg matrix. First an orthonormal basis for the image of the Krylov subspace $A\mathcal{K}_m(A, r_0)$ is computed by starting the Arnoldi process with Ar_0. The upper triangular matrix $R_m = (\rho_{i,j}) \in \mathbb{R}^{m\times m}$ is defined by the scalars $\rho_{i,j}$ computed in the orthogonalization process. We have assumed that the dimension of $\mathcal{K}_m(A, r_0)$ is m, this implies that R_m is non-singular. The columns of the matrix $Z_m = \begin{pmatrix} z_0 & z_1 & \dots & z_{m-1} \end{pmatrix} \in \mathbb{R}^{n\times m}$ span $\mathcal{K}_m(A, r_0)$. The upper triangular matrix R_m, computed in the Arnoldi process, satisfies

$$AZ_m = W_m R_m \tag{3.1}$$

The columns of the orthogonal matrix $W_m = \begin{pmatrix} z_1 & z_2 & \dots & z_m \end{pmatrix} \in \mathbb{R}^{n\times m}$ span $A\mathcal{K}_m(A, r_0)$. The orthogonalization of the residual vector r^{GMRES} with respect to the subspace $A\mathcal{K}_m(A, r_0)$ can be done by orthogonalization against z_j for $j = 1, 2, \dots, m$. The residual vector r_m satisfies $W_m^H r_m = 0$ and

$$z_0 = r_0 = r_m + \sum_{j=1}^{m} \xi_j z_j = r_m + W_m w_m \tag{3.2}$$

where the vector $w_m = \begin{pmatrix} \xi_1 & \xi_2 & \dots & \xi_m \end{pmatrix}^H$ is defined by the scalars that have been computed during the orthogonalization. The corresponding approximate solution is then given by $Z_m R_m^{-1} w_m$.

We assume that ξ_m is non-zero. However if $\xi_m = 0$, the last vector z_m is orthogonal to the current residual. This means that GMRES stagnates and that the residual polynomial is not changed. In the remainder of this paper we make use of the transformation matrix $T_m \in \mathbb{R}^{m\times m}$, which is completely defined by describing its action on the vector w_m and on the first $(m-1)$ columns of the identity matrix I_m. This matrix transforms w_m into e_1 and shifts e_{j-1} to e_j for $j = 2, 3, \dots, m$:

$$T_m = \begin{pmatrix} 0 & & & & & 1/\xi_m \\ 1 & 0 & & & & -\xi_1/\xi_m \\ & 1 & 0 & & & -\xi_2/\xi_m \\ & & \ddots & \ddots & & \vdots \\ & & & 1 & 0 & -\xi_{m-2}/\xi_m \\ & & & & 1 & -\xi_{m-1}/\xi_m \end{pmatrix} \tag{3.3}$$

The transformation matrix T_m is only required in combination with the upper triangular matrix R_m constructed in the simpler GMRES algorithm. We define the square upper Hessenberg matrix $\tilde{R}_m \in \mathbb{R}^{m\times m}$ as the product of R_m and T_m. Its explicit representation

$$\tilde{R}_m = R_m T_m = \begin{pmatrix} \rho_{1,2} & \rho_{1,3} & \cdots & \rho_{1,m} & \tau_1 \\ \rho_{2,2} & \rho_{2,3} & \cdots & \rho_{2,m} & \tau_2 \\ & \rho_{3,3} & \cdots & \rho_{3,m} & \tau_3 \\ & & \ddots & \vdots & \vdots \\ & & & \rho_{m,m} & \tau_m \end{pmatrix} \tag{3.4}$$

only requires the computation of the m scalars τ_i $(i = 1, 2, \dots, m)$. The determinant of $\tilde{R}_m$

can easily be computed:

$$\det \tilde{R}_m = \det R_m \det T_m = (-1)^{m+1} \frac{1}{\xi_m} \prod_{j=1}^{m} \rho_{j,j} \tag{3.5}$$

The results in the next two sections will show that $\tilde{R}_m$ is related to $\bar{H}_m$ as follows:

$$\tilde{R}_m = H_m^{-H} \bar{H}_m^H \bar{H}_m = H_m + h_{m+1,m}^2 f_m e_m^H \tag{3.6}$$

4. Harmonic Ritz values

The definition of harmonic Ritz values has been given by Sleijpen and Van der Vorst (section 5.1 in [24]) and is motivated by the fact that the reciprocals of the harmonic Ritz values are in the field of values of A^{-1}, whereas the Ritz values are in the field of values of A. This observation has been used by Manteuffel and Starke [14] to construct estimates of the spectrum. Since the harmonic Ritz values arise from an implicit application of a Rayleigh–Ritz procedure to the inverted operator, it is clear that they can be used successfully for the computation of interior eigenvalues. This approach is followed in the modified Rayleigh–Ritz procedure for interior eigenvalues proposed by Morgan and Zeng [15].

Definition 4.1. *The harmonic Ritz values $\tilde{\vartheta}_i^{(m)}$ are the reciprocals of the (ordinary) Ritz values of A^{-1} computed from $A\mathcal{K}_m(A, r_0)$.*

The harmonic Ritz values are the eigenvalues of $\tilde{R}_m$ as can be seen as follows. An approximate eigenvector $x = W_m y_m$ is sought in $A\mathcal{K}_m(A, r_0)$ and since $W_m^H Z_m = T_m^{-1}$ is the inverse of the transformation matrix T_m, the projected eigenvalue problem

$$\left(A^{-1}x - \mu x\right) \perp A\mathcal{K}_m(A, r_0) \Leftrightarrow W_m^H \left(A^{-1} W_m y_m - \mu W_m y_m\right) = 0 \tag{4.1}$$

results in $\tilde{R}_m y_m = \mu^{-1} y_m$. The harmonic Ritz values provide approximations to the eigenvalues, as shown in Theorem 5.1 in [24], which is reformulated here.

Theorem 4.1. *The harmonic Ritz values are eigenvalue approximations according to the minimal residual criterion.*

Proof
We seek an approximate eigenvector $x = Z_m y_m$ in $\mathcal{K}_m(A, r_0)$.

$$(Ax - \mu x) \perp A\mathcal{K}_m(A, r_0) \Leftrightarrow T_m R_m y_m = \mu y_m \tag{4.2}$$

The eigenvalues of $T_m R_m$ are also eigenvalues of the similar matrix $R_m T_m$. ∎

The harmonic Ritz values can also be computed from (1.1) using the eigenvalue problem (see [8,24])

$$\bar{H}_m^H \bar{H}_m y_m = \mu H_m^H y_m \Leftrightarrow \left(H_m + h_{m+1,m}^2 f_m e_m^H\right) y_m = \mu y_m \tag{4.3}$$

Since $f_m = H_m^{-H} e_m$ we have that $\|f_m\|_2 \leq 1/\sigma_{\min}(H_m)$ where $\sigma_{\min}(H_m)$ is the smallest singular value of H_m. Hence we can bound the norm of the rank one update in (4.3)

$$\|h_{m+1,m}^2 f_m e_m^H\|_2 \leq \frac{h_{m+1,m}^2}{\sigma_{\min}(H_m)} \tag{4.4}$$

The harmonic Ritz values equal the Ritz values when an invariant subspace has been found, since in this case $h_{m+1,m} = 0$. Equation (4.4) shows that the differences between the Ritz and harmonic Ritz values can only be large when $h_{m+1,m}$ is large and when $\sigma_{\min}(H_m)$ is small, which is the case when GMRES stagnates. Paige *et al*. [17] showed that for a symmetric matrix the Ritz values interlace the harmonic Ritz values and since both these values converge to the eigenvalues of the matrix we expect the differences between the Ritz and harmonic Ritz values to be small in the case of a real, symmetric matrix. We show that the differences between the Ritz and harmonic Ritz values can be large when GMRES stagnates.

5. GMRES residual polynomial

In this section we show that the GMRES residual polynomial $\tilde{\varphi}_m^{\text{GMRES}}(\lambda)$ is a multiple of the characteristic polynomial of $\tilde{R}_m$. This implies that the harmonic Ritz values are the zeros of the GMRES residual polynomial.

Lemma 5.1. *If the polynomial $\chi_m(\lambda) = \gamma_m \lambda^m + \cdots + \gamma_1 \lambda + \gamma_0$ has strict degree m, then $\chi_m(A) r_0 \in \mathcal{K}_{m+1}(A, r_0)$ and $\chi_m(A) r_0 \notin \mathcal{K}_m(A, r_0)$ and we have that*

$$\chi_m(A) r_0 = W_m \chi_m(\tilde{R}_m) w_m + \gamma_0 r_m \tag{5.1}$$

Proof

We define $u_1 = R_m e_1 = R_m T_m w_m = \tilde{R}_m w_m$ and $u_j = \tilde{R}_m u_{j-1} = \tilde{R}_m^{j-1} u_1 = \tilde{R}_m^j w_m$ for $j = 2, 3, \ldots, m$. Since $\tilde{R}_m$ is upper Hessenberg u_j has non-zero entries only in its first j positions. An expression for Ar_0 is readily available as this is the first vector computed in the Arnoldi process in simpler GMRES:

$$Ar_0 = z_1 \rho_{1,1} = W_m u_1 = W_m \tilde{R}_m w_m \tag{5.2}$$

By induction we prove that

$$A^k r_0 = W_m u_k = W_m \tilde{R}_m^k w_m \tag{5.3}$$

for $k = 1, 2, \ldots, m$. Equation (5.2) shows that (5.3) is valid for $k = 1$. We assume that (5.3) is valid for $k = 1, 2, \ldots, j$ and prove that it is also valid for $k = j + 1$. Using (3.1) we have for $j < m$

$$\begin{aligned} A^{j+1} r_0 &= W_m R_m e_2 e_1^H u_j + W_m R_m e_3 e_2^H u_j + \cdots + W_m R_m e_{j+1} e_j^H u_j \\ &= W_m R_m T_m u_j = W_m \tilde{R}_m u_j = W_m u_{j+1} = W_m \tilde{R}_m^{j+1} w_m \end{aligned} \tag{5.4}$$

We use (5.3) in combination with the expansion (3.2) of the initial residual to find an expression for $\chi_m(A)r_0$ in terms of W_m and r_m:

$$\chi_m(A)r_0 = \gamma_m W_m u_m + \cdots + \gamma_1 W_m u_1 + \gamma_0(r_m + W_m w_m) \tag{5.5}$$
$$= W_m \chi_m(\tilde{R}_m) w_m + \gamma_0 r_m \tag{5.6}$$

Equation (5.6) is the desired result. ■

The last non-zero entry in u_j is in position j and equals $\tilde{\zeta}_j = \rho_{1,1}\rho_{2,2}\ldots\rho_{j,j} \neq 0$. Hence from (5.3) we can conclude that

$$A^j r_0 = W_m u_j = \tilde{\zeta}_j z_j + \hat{z}_j \tag{5.7}$$

with $\hat{z}_j \in \operatorname{span}\{z_1, z_2, \ldots, z_{j-1}\}$ for $j = 1, 2, \ldots, m$.

Lemma 5.2. *All non-zero vectors in $\mathcal{K}_{m+1}(A, r_0)$ that are orthogonal to $A\mathcal{K}_m(A, r_0)$ can be written as $\alpha\tilde{\psi}_m(A)r_0$ for some non-zero $\alpha \neq 0$, where $\tilde{\psi}_m(\lambda) = \det\left(\lambda I - \tilde{R}_m\right)$ is the characteristic polynomial of $\tilde{R}_m$.*

Proof
Let the polynomial $\chi_j(\lambda) = \tilde{\gamma}_j\lambda^j + \cdots + \tilde{\gamma}_1\lambda + \tilde{\gamma}_0$ have strict degree $j < m$. If this polynomial has a non-zero constant term $\tilde{\gamma}_0 \neq 0$ then we can see from (5.5) and (5.7) that $\chi_j(A)r_0$ has a non-zero component $\tilde{\gamma}_0\xi_m z_m$ along $z_m \in A\mathcal{K}_m(A, r_0)$ and thus is not orthogonal to $A\mathcal{K}_m(A, r_0)$. Recall that we have assumed that GMRES does not stagnate in the last step and this implies that $\xi_m \neq 0$ is non-zero. On the other hand, if this polynomial does not have a constant term $\tilde{\gamma}_0 = 0$ then we know from (5.7) that $\chi_j(A)r_0$ has a non-zero component $\tilde{\gamma}_j\tilde{\zeta}_j z_j$ along $z_j \in A\mathcal{K}_m(A, r_0)$ and thus is not orthogonal to $A\mathcal{K}_m(A, r_0)$. Since $\tilde{\psi}_m(\lambda) = \det\left(\lambda I - \tilde{R}_m\right) = \gamma_m\lambda^m + \cdots + \gamma_1\lambda + \gamma_0$ is the characteristic polynomial of $\tilde{R}_m$, we have by the Cayley–Hamilton theorem that $\tilde{\psi}_m(\tilde{R}_m) = 0$. Setting the polynomial $\chi_m(\lambda) = \alpha\tilde{\psi}_m(\lambda)$ in (5.1) we deduce that $\alpha\tilde{\psi}_m(A)r_0 = \alpha\gamma_0 r_m$ is orthogonal to $A\mathcal{K}_m(A, r_0)$. Any polynomial $\varphi_m(\lambda)$ of degree m that is not a scalar multiple of $\tilde{\psi}_m(\lambda)$ can be written as $\varphi_m(\lambda) = \alpha\tilde{\psi}_m(\lambda) + \chi_j(\lambda)$ with $\chi_j(\lambda)$ a non-zero polynomial of degree $j < m$. We have that

$$\varphi_m(A)r_0 = \alpha\tilde{\psi}_m(A)r_0 + \chi_j(A)r_0 = \alpha\gamma_0 r_m + \chi_j(A)r_0 \tag{5.8}$$

which is not orthogonal to $A\mathcal{K}_m(A, r_0)$. ■

Theorem 5.1. *The GMRES residual polynomial is a multiple of the characteristic polynomial of the Hessenberg matrix $\tilde{R}_m$.*

Proof
Since the GMRES residual polynomial $\tilde{\varphi}_m^{\mathrm{GMRES}}(\lambda) = \gamma_m\lambda^m + \cdots + \gamma_1\lambda + \gamma_0$ has degree m (5.1) yields an expression for $r^{\mathrm{GMRES}} \in \mathcal{K}_{m+1}(A, r_0)$

$$r^{\mathrm{GMRES}} = \tilde{\varphi}_m^{\mathrm{GMRES}}(A)r_0 = W_m\tilde{\varphi}_m^{\mathrm{GMRES}}(\tilde{R}_m)w_m + \gamma_0 r_m. \tag{5.9}$$

Thus $\tilde{\varphi}_m^{\mathrm{GMRES}}(\tilde{R}_m) = 0$ is necessary to have $r^{\mathrm{GMRES}} \perp A\mathcal{K}_m(A, r_0)$. By Lemma 5.2 we know that the GMRES residual polynomial $\tilde{\varphi}_m^{\mathrm{GMRES}}(\lambda) = \alpha\tilde{\psi}_m(\lambda)$ must be a scalar multiple of the characteristic polynomial of $\tilde{R}_m$ in order to eliminate all the components of the residual in $A\mathcal{K}_m(A, r_0)$. The constant α can be determined from $\tilde{\varphi}_m^{\mathrm{GMRES}}(0) = \alpha\tilde{\psi}_m(0) = 1$ since $\tilde{\psi}_m(0) \neq 0$ unless the dimension of $\mathcal{K}_m(A, r_0)$ is less than m or GMRES stagnates in the last step. The value of $\tilde{\psi}_m(0)$ can easily be computed from (3.5)

$$\tilde{\psi}_m(0) = -\det \tilde{R}_m = (-1)^m \frac{1}{\xi_m} \prod_{j=1}^{m} \rho_{j,j} \tag{5.10}$$

We obtain the following expression for the residual polynomial

$$\tilde{\varphi}_m^{\mathrm{GMRES}}(\lambda) = \frac{\tilde{\psi}_m(\lambda)}{\tilde{\psi}_m(0)} \tag{5.11}$$

This completes the proof. ∎

Hence, the GMRES residual polynomial is uniquely defined by the fact that its m zeros are the harmonic Ritz values $\tilde{\vartheta}_i^{(m)}$ and by the normalization $\tilde{\varphi}_m^{\mathrm{GMRES}}(0) = 1$. Related theoretical results have already been proved by Freund [8,9] and by Manteuffel and Otto [13]. The GMRES residual polynomial is a standard kernel polynomial and its zeros can be computed as the eigenvalues of the generalized problem (4.3) as was shown by Freund [8]. However, Theorem 5.1 shows that these zeros can be computed efficiently within the simpler GMRES algorithm. The point made in this paper is that the Walker–Zhou interpretation of GMRES leads to a simpler analysis for the involved Arnoldi matrices.

6. Example 1: stagnation of GMRES

We describe breakdown of FOM and stagnation of GMRES in terms of the Ritz and harmonic Ritz values. To illustrate the crucial points we use the following example. Consider the linear system $A_n x_n = b_n$ where

$$A_n = \begin{pmatrix} 0 & 1 & & \\ & \ddots & \ddots & \\ & & 0 & 1 \\ 1 & & & 0 \end{pmatrix} \in \mathbb{R}^{n\times n} \text{ and } b_n = \begin{pmatrix} 0 \\ \vdots \\ 0 \\ 1 \end{pmatrix} \in \mathbb{R}^n \tag{6.1}$$

The Arnoldi process generates an orthonormal basis for the Krylov subspace starting from $v_1 = \begin{pmatrix} 0 & \dots & 0 & 1 \end{pmatrix}^H$. After the first step we have $v_2 = \begin{pmatrix} 0 & \dots & 0 & 1 & 0 \end{pmatrix}^H$ and FOM breaks down. Hence the Ritz value $\vartheta^{(1)} = 0$ equals zero. The FOM residual polynomial evaluated for non-zero λ shows that the norm of the residual $\|r^{\mathrm{FOM}}\|_2$ may grow without bound when FOM encounters a breakdown. In this case GMRES stagnates. The harmonic Ritz value $\tilde{\vartheta}^{(1)} = \infty$ is set equal to infinity. Hence the GMRES residual polynomial evaluated for finite λ shows the stagnation. In simpler GMRES the stagnation can be seen from the fact that the corresponding projection $\xi_1 = 0$ is zero.

In the following numerical experiment we are concerned with the eigenvalue estimates when GMRES nearly stagnates. The matrix A_n is given in (6.1) and the right-hand side is $b_n(\varepsilon) = \begin{pmatrix} \varepsilon & \dots & \varepsilon & 1+\varepsilon \end{pmatrix}^H \in \mathbb{R}^n$. The initial guess $x_0 = 0$ is the zero vector. Our interest is in the convergence behaviour for small $\varepsilon > 0$. The eigenvalues of A_n satisfy

Table 1. Estimated and computed norm of the Ritz and harmonic Ritz values from $\mathcal{K}_m(A, b)$ for example 1

	Ritz values			Harmonic Ritz values		
m	minimum	estimated	maximum	minimum	estimated	maximum
10	0.263	0.269	0.278	3.595	3.714	3.802
19	0.491	0.501	0.521	1.919	1.995	2.037

$\lambda^n = 1$. Hence, in the complex plane, they form a regular n-polygon on the unit circle. The numerical results in this section were obtained with $\varepsilon = 10^{-6}$ and $n = 20$. The norm of the residual $\|r^{\text{GMRES}}\|_2$ decreased from $1 - 2.0 \times 10^{-12}$ to $1 - 3.8 \times 10^{-11}$ in 19 steps.

For $m < n$ the matrices $\bar{H}_m$ and V_m can be approximated successfully by $\bar{H}_m^* \in \mathbb{R}^{(m+1)\times m}$ and $V_m^* \in \mathbb{R}^{n\times m}$:

$$\bar{H}_m^* = \begin{pmatrix} 2\varepsilon & 2\varepsilon & \dots & 2\varepsilon \\ 1 & & & \\ & 1 & & \\ & & \ddots & \\ & & & 1 \end{pmatrix} \quad \text{and} \quad V_m^* = \begin{pmatrix} \varepsilon & \varepsilon & \dots & \varepsilon \\ \vdots & \vdots & & \vdots \\ \varepsilon & \varepsilon & \dots & \varepsilon \\ \varepsilon & \varepsilon & & 1+\varepsilon \\ \varepsilon & \varepsilon & \dots & -\varepsilon \\ & & & \vdots \\ \varepsilon & 1+\varepsilon & & \\ 1+\varepsilon & -\varepsilon & \dots & -\varepsilon \end{pmatrix} \tag{6.2}$$

With these approximations we see that the Ritz values are $|\vartheta^{(m)}| = \mathcal{O}\left(\sqrt[m]{2\varepsilon}\right)$. It is easy to show that the Ritz values approximately form a regular m-polygon in the complex plane. In this case the harmonic Ritz values are the reciprocals of the complex conjugates of the ordinary Ritz values $\vartheta^{(m)}$. The harmonic Ritz values are $|\tilde{\vartheta}^{(m)}| = \mathcal{O}\left(1/\sqrt[m]{2\varepsilon}\right)$, going off to infinity as ε tends to zero. Table 1 shows the quality of these estimations which increases with decreasing ε and decreasing m.

From (6.2) we estimate $\sigma_{\min}(H_m) \approx 2\varepsilon$ and $h_{m+1,m} \approx 1$. With this result we see that $\|h_{m+1,m}^2 f_m e_m^H\|_2 \approx 1/(2\varepsilon)$, showing that the bound in (4.4) can be reached. We have $\|H_m\|_2 \approx 1$ and $\|H_m + h_{m+1,m}^2 f_m e_m^H\|_2 \approx 1/(2\varepsilon)$. Hence, we cannot expect the Ritz values and the harmonic Ritz values to be equal. Since GMRES stagnates, the norm of the FOM residual $\|r^{\text{FOM}}\|_2 \approx 1/(2\varepsilon)$ is large.

These results show that the differences between the Ritz and harmonic Ritz values are significantly large when GMRES (nearly) stagnates. To illustrate this we show in Figure 1 the Ritz and harmonic Ritz values computed from $\mathcal{K}_{10}(A_{20}, b_{20}(10^{-6}))$. The harmonic Ritz values are plotted using a plus sign (+), while the Ritz values are shown with a multiplication sign (×). An asterisk (∗) is used to plot the eigenvalues. Figure 2 shows the Ritz and harmonic Ritz values computed from $\mathcal{K}_{19}(A_{20}, b_{20}(10^{-6}))$.

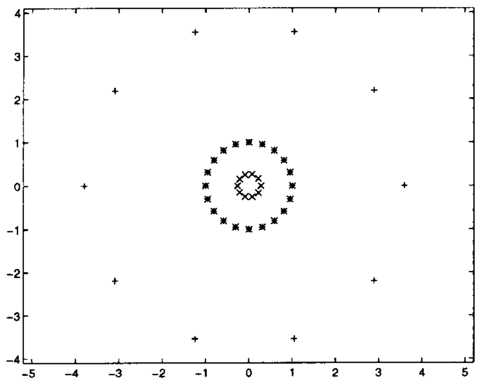

Figure 1. Ritz ($\times$) and harmonic Ritz ($+$) values from $\mathcal{K}_{10}(A, b)$ for Example 1

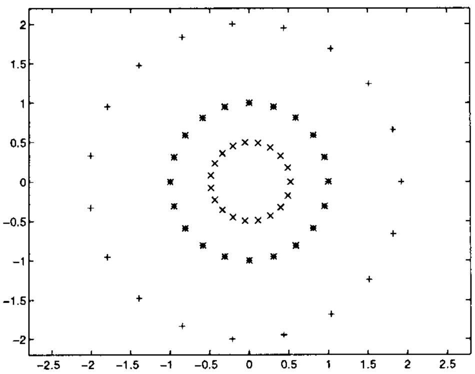

Figure 2. Ritz ($\times$) and harmonic Ritz ($+$) values from $\mathcal{K}_{19}(A, b)$ for Example 1

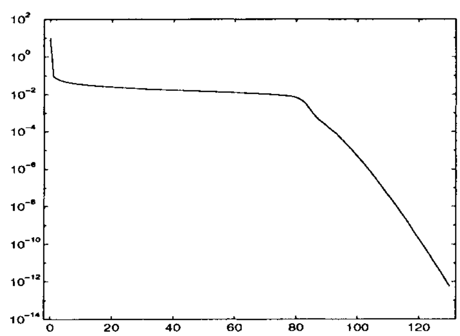

Figure 3. Norm of the residual $\|r^{\mathrm{GMRES}}\|_2$ as a function of m for Example 2

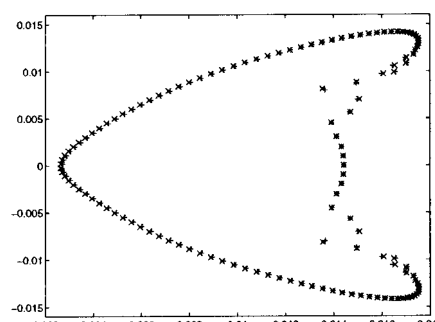

Figure 4. Ritz ($\times$) and harmonic Ritz ($+$) values from $\mathcal{K}_{130}(A, f)$ for example 2

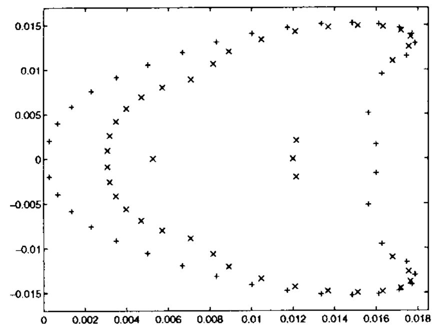

Figure 5. Ritz ($\times$) and harmonic Ritz ($+$) values from $\mathcal{K}_{41}(A, f)$ for Example 2

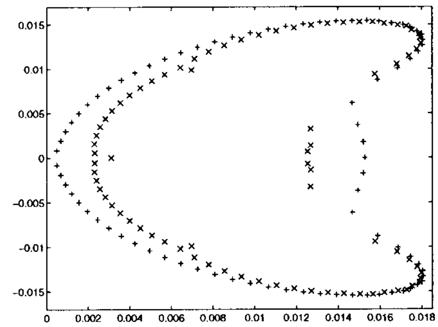

Figure 6. Ritz ($\times$) and harmonic Ritz ($+$) values from $\mathcal{K}_{80}(A, f)$ for Example 2

7. Example 2: a convection-diffusion problem

The aim is to compute the steady-state solution of a linearized convection–diffusion problem. The problem is formulated as follows: given a divergence-free ($\nabla\vec{w} = 0$) convective velocity field $\vec{w}$, find a scalar variable u satisfying

$$-\nu\nabla^2 u + \vec{w}\nabla u = f \text{ in } \Omega \tag{7.1}$$

with a Dirichlet boundary condition $u(x) = g(x)$ on $\partial\Omega$. A 2D convection dominated convection–diffusion problem is solved on a uniform square grid with mesh width $h = 1/64$. The constant velocity field $\vec{w} = \begin{pmatrix} -1/\sqrt{2} & 1/\sqrt{2} \end{pmatrix}$ has a 45° inclination with the grid lines. The diffusion parameter is set to $\nu = 10^{-6}$, resulting in a 'mesh Peclet number' of $Pe = \|w\|_2 h/(2\nu) = 7\,812.5 > 1$, which shows that the convection is dominant.

After SUPG finite element discretization a linear system $Au = f$ is obtained in which the unknown u is the discretized version of the unknown in (7.1) and the vector f corresponds to the right-hand side. The non-symmetric matrix $A = \nu L + N$ is the sum of the discretization of the diffusive term and the skew-symmetric discrete version of the convection operator N. The matrix $L = -\nabla_h^2 + S$ is the sum of the discretized Laplacian and a matrix S corresponding to stabilization terms, since streamline upwinding is used. Details about the discretization can be found in [6]. We use unpreconditioned GMRES to solve the linear system $Au = f$.

Figure 3 shows the residual norm $\|r^{\text{GMRES}}\|_2$ as a function of the dimension m of $\mathcal{K}_m(A, f)$. In Figures 4–6 the Ritz value $\vartheta_i^{(m)} = 1$ and the harmonic Ritz value $\tilde{\vartheta}_j^{(m)} = 1$ are not shown due to scaling. The harmonic Ritz values are plotted using a plus sign ($+$), while the Ritz values are shown with a multiplication sign ($\times$). The Krylov subspace $\mathcal{K}_{130}(A, f)$ of dimension $m = 130$ yields an accurate solution. Hence the differences between the Ritz and harmonic Ritz values are small. Figure 4 shows the Ritz and harmonic Ritz values computed from $\mathcal{K}_{130}(A, f)$. To illustrate that the Ritz and harmonic Ritz values differ significantly when GMRES (nearly) stagnates, we show in Figure 5 the Ritz and harmonic Ritz values computed from $\mathcal{K}_{41}(A, f)$. Figure 6 shows the Ritz and harmonic Ritz values computed from $\mathcal{K}_{80}(A, f)$.

8. Concluding remarks

Nachtigal *et al*. [16] have presented a number of arguments why the GMRES residual polynomial should be used instead of Arnoldi eigenvalue estimates. They compute the coefficients of the polynomial explicitly by transforming back to the Krylov (power) basis and incorporate a root-finding step in their hybrid algorithm. However, finding the roots of a polynomial is an ill-conditioned problem. The zeros of the GMRES residual polynomial can be computed by solving an eigenvalue problem. This is not only a cheap but also a stable procedure. Saylor and Smolarski [23] have also presented an adaptive algorithm which is a hybrid combination of GMRES and Richardson's method, but they prefer using roots that are in the field of values of A and point out that this is the crucial difference between their algorithm and the one by Nachtigal *et al*. [16].

Toh and Trefethen [25] advocate the use of $\bar{H}_m$ because it bypasses the usual consideration of Ritz values or 'Arnoldi eigenvalue estimates'. For highly non-normal matrices we cannot expect the Arnoldi iteration to be effective at determining eigenvalues. On the other hand, Greenbaum et al. [12] have shown that eigenvalues cannot be used to predict the convergence of GMRES for highly non-normal matrices. Toh and Trefethen [25] have pointed out that plotting the Ritz values in order to analyse the convergence of non-symmetric Krylov solvers is not sufficient.

We suggest that both the Ritz spectrum and the harmonic Ritz spectrum should be considered when analysing the convergence of non-symmetric Krylov solvers. The cost of computing the harmonic Ritz values only depends on m, which in practice is small, while the computation of pseudo-spectra and lemniscates is usually far too expensive.

Acknowledgements

The authors would like to thank R. W. Freund for pointing out the relation to quasi-kernel polynomials. We are also grateful to H. A. Van der Vorst for his valuable suggestions. This paper presents research results of the Belgian Programme on Interuniversity Poles of Attraction, initiated by the Belgian State, Prime Minister's Office for Science, Technology and Culture (IUAP P4/2). The scientific responsibility rests with its authors. This research is also supported by the Research Council K.U.Leuven (OT/94/16). Serge Goossens is financed by a specialization scholarship of the Flemish Institute for the Promotion of Scientific and Technological Research in Industry (IWT).

REFERENCES

1. W. E. Arnoldi. The principle of minimized iterations in the solution of the matrix eigenvalue problem. *Quart. Appl. Math.*, 9(1), 17–29, 1951.
2. 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.
3. P. N. Brown. A theoretical comparison of the Arnoldi and GMRES algorithms. *SIAM J. Sci. Stat. Comput.*, 12(1), 58–78, 1991.
4. A. M. Bruaset. *A Survey of Preconditioned Iterative Methods*. Pitman Research Notes in Mathematics Series. Longman Scientific & Technical, Essex, UK, 1995.
5. J. Cullum and A. Greenbaum. Relations between Galerkin and norm-minimising iterative methods for solving linear systems. *SIAM J. Matrix Anal. Appl.*, 17(2), 223–246, 1996.
6. H. C. Elman, D. J. Silvester and A. J. Wathen. Iterative methods for problems in computational fluid dynamics. Technical Report 96/19, Oxford University Computing Laboratory, 1996.

7. B. Fischer. *Polynomial Based Iteration Methods for Symmetric Linear Systems*. John Wiley & Sons, Ltd., Chicester, UK, 1996.
8. R. W. Freund. Quasi-kernel polynomials and their use in non-Hermitian matrix iterations. *J. Comput. Appl. Math.*, 43; 135–158, 1992.
9. R. W. Freund, G. H. Golub and N. M. Nachtigal. *Iterative Solution of Linear Systems*, pages 57–100. Acta Numerica 1992. Cambridge University Press, 1992.
10. S. Goossens and D. Roose. Ritz and Harmonic Ritz values and the convergence of FOM and GMRES. Technical Report TW257, Department of Computer Science, K. U. Leuven, 1997.
11. A. Greenbaum. *Iterative Methods for Solving Linear Systems*. SIAM, Philadelphia, PA, 1997.
12. A. Greenbaum, V. Pták and Z. Strakos. Any nonincreasing convergence curve is possible for GMRES. *SIAM J. Matrix Anal. Appl.*, 17(3), 465–469, 1996.
13. T. A. Manteuffel and J. S. Otto. On the roots of the orthogonal polynomials and residual polynomials associated with a conjugate gradient method. *Num. Lin. Alg. Appl.*, 1(5), 449–475, 1994.
14. T. A. Manteuffel and G. Starke. On hybrid iterative methods for nonsymmetric systems of linear equations. *Numer. Math.*, 73, 489–506, 1996.
15. R. B. Morgan and M. Zeng. Harmonic projection methods for large nonsymmetric eigenvalue problems. *Num. Lin. Alg. Appl.*, 5(1), 33–55, 1998.
16. N. M. Nachtigal, L. Reichel and L. N. Trefethen. A hybrid GMRES algorithm for nonsymmetric linear systems. *SIAM J. Matrix Anal. Appl.*, 13(3), 796–825, 1992.
17. C. C. Paige, B. N. Parlett and H. A. Van der Vorst. Approximate solutions and eigenvalue bounds from Krylov subspaces. *Num. Lin. Alg. Appl.*, 2(2):115–134, 1995.
18. B. N. Parlett. *The Symmetric Eigenvalue Problem*. Prentice-Hall, Englewood Cliffs, NJ, 1980.
19. Y. Saad. Projection methods for solving large sparse eigenvalue problems. In B. Kågström and A. Ruhe, editors, *Matrix Pencils*, Lecture Notes in Mathematics 973, pages 121–144. Springer-Verlag, 1983.
20. Y. Saad. *Numerical Methods for Large Eigenvalue Problems*. Manchester University Press, Manchester, UK, 1992.
21. Y. Saad. *Iterative Methods for Sparse Linear Systems*. PWS Publishing Company, Boston, MA, 1996.
22. Y. Saad and M. H. Schultz. GMRES: a generalized minimal residual algorithm for solving nonsymmetric linear systems. *SIAM J. Sci. Stat. Comput.*, 7(3), 856–869, 1986.
23. P. E. Saylor and D. C. Smolarski. Implementation of an adaptive algorithm for Richardson's method. *Lin. Alg. Appl.*, 154–156, 615–646, 1991.
24. G. L. G. Sleijpen and H. A. Van der Vorst. A Jacobi–Davidson iteration method for linear eigenvalue problems. *SIAM J. Matrix Anal. Appl.*, 17(2), 401–425, 1996.
25. K. C. Toh and L. N. Trefethen. Calculation of pseudospectra by the Arnoldi iteration. *SIAM J. Sci. Comput.*, 17(1), 1–15, 1996.
26. H. F. Walker and L. Zhou. A Simpler GMRES. *Num. Lin. Alg. Appl.*, 1(6), 571–581, 1994.

