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

Serge Goossens   *Dirk Roose*

Report TW 257, April 1997

Katholieke Universiteit Leuven
Department of Computer Science
Celestijnenlaan 200A – B-3001 Heverlee (Belgium)

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

Serge Goossens　　*Dirk Roose*

Report TW 257, April 1997

Department of Computer Science, K.U.Leuven

Abstract

FOM and GMRES are Krylov subspace iterative methods for solving nonsymmetric linear systems. The Ritz values are approximate eigenvalues, which can be computed cheaply within these algorithms. In this paper, we generalise the concept of Harmonic Ritz values, introduced by Paige et al. for symmetric matrices, to nonsymmetric matrices. We show that the zeroes of the residual polynomials of FOM and GMRES are the Ritz and Harmonic Ritz values respectively. We present an upper bound for the difference between the matrices from which the Ritz and Harmonic Ritz values are computed. The differences between these values allows us to describe breakdown of FOM and stagnation of GMRES.

Keywords : Krylov subspaces, FOM, Ritz Values, GMRES, Harmonic Ritz values.
AMS(MOS) Classification : Primary : 65F10, Secondary : 65F15, 65F50.

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

Abstract

FOM and GMRES are Krylov subspace iterative methods for solving nonsymmetric linear systems. The Ritz values are approximate eigenvalues, which can be computed cheaply within these algorithms. In this paper, we generalise the concept of Harmonic Ritz values, introduced by Paige et al. for symmetric matrices, to nonsymmetric matrices. We show that the zeroes of the residual polynomials of FOM and GMRES are the Ritz and Harmonic Ritz values respectively. We present an upper bound for the difference between the matrices from which the Ritz and Harmonic Ritz values are computed. The differences between these values allows us to describe breakdown of FOM and stagnation of GMRES.

KEY WORDS Krylov subspaces, FOM, Ritz values, GMRES, 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 nonsingular matrix $A \in \mathbb{R}^{n\times n}$ and a vector $r_0 \in \mathbb{R}^n$, the Krylov subspaces

$$\mathcal{K}_m(A, r_0) = \text{SPAN}\{r_0, Ar_0, A^2 r_0, \ldots, A^{(m-1)} r_0\} \tag{1}$$

* Serge.Goossens@cs.kuleuven.ac.be http://www.cs.kuleuven.ac.be/∼serge/
† Dirk.Roose@cs.kuleuven.ac.be

for $m = 1, 2, \ldots, n$ form a nested sequence of subspaces. In this paper we will assume that the dimension of the Krylov subspace $\mathcal{K}_m(A, r_0)$ is m. If the dimension of $\mathcal{K}_m(A, r_0)$ is smaller than m, then this subspace is invariant under A and the approximations discussed in this paper are exact. For the iterative solution of linear systems this means that $\mathcal{K}_m(A, r_0)$ contains the exact solution and for the computation of eigenvalues this means that $\mathcal{K}_m(A, r_0)$ contains the invariant subspaces corresponding to some eigenvalues of A.

The Arnoldi basis is an orthonormal basis $\{v_1, v_2, \ldots, v_m\}$ obtained by orthogonalising the Krylov (power) basis $\{r_0, Ar_0, A^2r_0, \ldots, A^{(m-1)}r_0\}$. The computational kernel of FOM and GMRES is the Arnoldi process which computes the orthonormal basis for the Krylov subspace $\mathcal{K}_m(A, r_0)$ and yields the fundamental relation (4). Originally Arnoldi [2] introduced this procedure to reduce a dense matrix into Hessenberg form and remarked that this process could be used to compute approximations if it was stopped before completion. Nowadays it's widely known that good approximations can be computed if the Arnoldi process is carried out with a well-preconditioned operator. Let the matrix $V_m \in \mathbb{R}^{n\times m}$ be defined as

$$V_m = \begin{pmatrix} v_1 & v_2 & \ldots & v_m \end{pmatrix}. \tag{2}$$

Since the Arnoldi basis is orthonormal, V_m is an orthogonal matrix. In the orthogonalisation process the scalars $h_{i,j}$ are computed so that the square upper Hessenberg matrix $H_m \in \mathbb{R}^{m\times m}$

$$H_m = \begin{pmatrix} h_{1,1} & h_{1,2} & \ldots & h_{1,m-1} & h_{1,m} \\ h_{2,1} & h_{2,2} & \ldots & h_{2,m-1} & h_{2,m} \\ & h_{3,2} & \ldots & h_{3,m-1} & h_{3,m} \\ & & \ddots & \vdots & \vdots \\ & & & h_{m,m-1} & h_{m,m} \end{pmatrix} \tag{3}$$

satisfies the fundamental relation

$$AV_m = V_m H_m + h_{m+1,m} v_{m+1} e_m^H = V_{m+1} \bar{H}_m. \tag{4}$$

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 $\begin{pmatrix} 0 & \ldots & 0 & h_{m+1,m} \end{pmatrix}$. From (4) we derive the following expression for H_m:

$$H_m = V_m^H A V_m. \tag{5}$$

If the matrix A is symmetric, the fundamental relation (4) can be replaced by a 3-term recurrence relation, since in this case H_m is a symmetric tridiagonal matrix.

In the next section we briefly describe FOM and GMRES, two well-known Krylov subspace iterative methods for solving nonsymmetric linear systems. The eigenvalues of H_m are called Ritz values and approximate the eigenvalues of A. We prove in section 3 that these Ritz values are the zeroes of the FOM residual polynomial. We give an outline of the Simpler GMRES algorithm of Walker and Zhou because in the subsequent theory we use the upper triangular matrix which is computed in this algorithm. A discussion on orthogonalisation concludes section 4. In section 5 we define a transformation matrix which is also used in the proofs. We generalise the concept of Harmonic Ritz values, introduced by Paige et al. for symmetric matrices, to nonsymmetric matrices and prove in section 6 that they are approximate eigenvalues according to the minimal residual criterion. We present an upper bound for the difference between the matrices from which the Ritz and Harmonic Ritz values are computed. In section 7 we prove that the zeroes of the GMRES residual polynomials are the Harmonic Ritz values. In section 8 we relate breakdown of FOM and stagnation of GMRES to the differences between the Ritz and Harmonic Ritz values. Numerical results for a problem which causes nearly breakdown and stagnation are presented in section 9. In section 10 we give numerical results for a 2D convection dominated convection-diffusion problem. Some remarks concerning related work conclude this paper.

2 Approximate Solutions from Krylov Subspaces

Krylov subspace iterative methods such as Arnoldi's method for linear systems [2], also known as the full orthogonalisation method (FOM), and the generalised minimal residual method (GMRES) proposed by Saad and Schultz [12] are well known methods for the approximate solution of large sparse nonsymmetric linear systems. Both algorithms are outlined in Saad's book [11]. A theoretical comparison of FOM and GMRES is given by Brown [3]. In this paper we assume that the initial guess x_0 to the solution of the linear system $Ax = b$ is zero and thus the initial residual r_0 equals the right-hand side b. Unless the initial guess x_0 is the exact solution to the linear system, the norm $\beta = \|r_0\|_2 = \|b\|_2$ of the residual is nonzero. We assume that the system has been scaled so that the initial residual has unit length by setting $\tilde{x} = x/\beta$ and $\tilde{b} = b/\beta$.

Since both FOM and GMRES select an approximate solution from the Krylov subspace this approximation can be written as

$$x = \varphi_{m-1}(A)r_0 \in \mathcal{K}_m(A, r_0), \tag{6}$$

where $\varphi_{m-1}(\lambda) = \gamma_{m-1}\lambda^{m-1} + \cdots + \gamma_1\lambda + \gamma_0 \in \mathbb{P}_{m-1}$ is a real polynomial of degree $(m-1)$. The residual corresponding to this approximate solution is

$$r = b - Ax = (I - A\varphi_{m-1}(A))\, r_0 = \tilde{\varphi}_m(A)r_0 \in \mathcal{K}_{m+1}(A, r_0) \tag{7}$$

where

$$\tilde{\varphi}_m(\lambda) = 1 - \lambda\varphi_{m-1}(\lambda). \tag{8}$$

The residual can thus be written as the residual polynomial $\tilde{\varphi}_m(\lambda)$ evaluated for $\lambda = A$ acting on the initial residual r_0. From (8) it is clear that residual polynomials are normalised by $\tilde{\varphi}_m(0) = 1$. The main topic addressed in this paper is the computation of the zeroes of the residual polynomial $\tilde{\varphi}_m(\lambda)$ without explicitly constructing it.

FOM selects the approximate solution $x^{\mathrm{FOM}} = \varphi_{m-1}^{\mathrm{FOM}}(A)r_0 \in \mathcal{K}_m(A, r_0)$ such that the residual $r^{\mathrm{FOM}} = \tilde{\varphi}_m^{\mathrm{FOM}}(A)r_0$ is orthogonal to $\mathcal{K}_m(A, r_0)$

$$r^{\mathrm{FOM}} = \tilde{\varphi}_m^{\mathrm{FOM}}(A)r_0 \perp \mathcal{K}_m(A, r_0). \tag{9}$$

Since V_m spans $\mathcal{K}_m(A, r_0)$ this can be formulated as

$$V_m^H (r_0 - AV_m y_m) = 0 \Leftrightarrow H_m y_m = e_1, \tag{10}$$

which shows that a linear system with H_m must be solved to compute the FOM approximate solution. In this paper e_j denotes column j of the identity matrix.

GMRES selects the approximate solution $x^{\text{GMRES}} = \varphi_{m-1}^{\text{GMRES}}(A)r_0 \in \mathcal{K}_m(A, r_0)$ which minimises the norm of the residual $\|r^{\text{GMRES}}\|_2$

$$\min_{x \in \mathcal{K}_m(A,r_0)} \|\tilde{\varphi}_m^{\text{GMRES}}(A)r_0\|_2. \tag{11}$$

The matrix V_{m+1} is orthogonal and spans the subspaces $\text{SPAN}\{r_0\}$ and $\text{SPAN}\{AV_m\}$. Hence we can write

$$\|r^{\text{GMRES}}\|_2 = \|V_{m+1}^H (r_0 - AV_m y_m)\|_2 = \|e_1 - \bar{H}_m y_m\|_2. \tag{12}$$

This overdetermined linear system can be solved in the least squares sense using the normal equations:

$$\bar{H}_m^H \bar{H}_m y_m = \bar{H}_m^H e_1 = H_m^H e_1. \tag{13}$$

If H_m is nonsingular, (13) can be rewritten as

$$H_m^{-H} \bar{H}_m^H \bar{H}_m y_m = e_1. \tag{14}$$

Equation (14) can be simplified by observing that $\bar{H}_m^H \bar{H}_m = H_m^H H_m + h_{m+1,m}^2 e_m e_m^H$. We define the vector $f_m = H_m^{-H} e_m$ and note that the matrix in (14) is a rank one update of H_m

$$H_m^{-H} \bar{H}_m^H \bar{H}_m = H_m + h_{m+1,m}^2 f_m e_m^H. \tag{15}$$

The linear system that must be solved to compute the GMRES approximate solution can thus be written as:

$$\left(H_m + h_{m+1,m}^2 f_m e_m^H\right) y_m = e_1. \tag{16}$$

The minimal residual criterion is equivalent to the requirement that the residual is orthogonal to the image of the Krylov subspace $A\mathcal{K}_m(A, r_0)$

$$r^{\text{GMRES}} = \tilde{\varphi}_m^{\text{GMRES}}(A)r_0 \perp A\mathcal{K}_m(A, r_0). \tag{17}$$

Plugging (4) into (17) also yields (13) which shows that orthogonalising the residual with respect to $A\mathcal{K}_m(A, r_0)$ is equivalent to the minimal residual requirement.

3 Ritz Values and FOM Residual Polynomial

The classical Galerkin approach for computing approximate eigenpairs has been discussed by several authors, e.g. in a book by Saad [10]. 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 $\mathcal{K}_m(A, r_0)$

$$(Ax - \mu x) \perp \mathcal{K}_m(A, r_0) \Leftrightarrow V_m^H (AV_m y_m - \mu V_m y_m) = 0. \tag{18}$$

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

$$H_m y_m = \mu y_m. \tag{19}$$

By definition 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".

Now we prove that 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 zeroes of the FOM residual polynomial or equivalently that the FOM approximation polynomial $\varphi_{m-1}^{\mathrm{FOM}}(\lambda)$ interpolates λ^{-1} in the Ritz values.

Lemma 3.1 *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 the following expression*

$$\chi_m(A) r_0 = V_m \chi_m(H_m) e_1 + \gamma_m \zeta_m v_{m+1} \tag{20}$$

where $\zeta_m = h_{m+1,m} \ldots h_{3,2} h_{2,1}$.

Proof We define $u_1 = e_1$ and $u_j = H_m u_{j-1} = H_m^{j-1} u_1$ for $j = 2, 3, \ldots, m$. Since H_m is upper Hessenberg u_j has nonzero entries only in its first j positions, consequently $e_m^H u_j = 0$ for $j < m$. Using (4) we have for $j < m$

$$\begin{aligned}
A V_m e_1 &= \left(V_m H_m + h_{m+1,m} v_{m+1} e_m^H\right) e_1 = V_m H_m e_1 = V_m u_2 & (21)\\
A^j V_m e_1 &= A V_m u_j = V_m H_m u_j = V_m H_m^j e_1 = V_m u_{j+1} & (22)\\
A^m V_m e_1 &= A V_m u_m = \left(V_m H_m + h_{m+1,m} v_{m+1} e_m^H\right) u_m & \\
&= V_m H_m u_m + h_{m+1,m} v_{m+1} e_m^H u_m = V_m H_m^m e_1 + h_{m+1,m} v_{m+1} e_m^H u_m & (23)
\end{aligned}$$

A straightforward computation shows that $h_{m+1,m} e_m^H u_m = \zeta_m$ and we can conclude that

$$\chi_m(A) V_m e_1 = V_m \chi_m(H_m) e_1 + \gamma_m \zeta_m v_{m+1}. \tag{24}$$

This completes the proof since $r_0 = V_m e_1$. ■

The last nonzero entry in u_{j+1} is in position $j+1$ and equals $\zeta_j = h_{j+1,j} \ldots h_{3,2} h_{2,1} \neq 0$. Hence from (22) we can conclude that

$$A^j V_m e_1 = V_m u_{j+1} = \zeta_j v_{j+1} + \hat{v}_{j+1} \tag{25}$$

with $\hat{v}_{j+1} \in \mathrm{SPAN}\{v_1, v_2, \ldots, v_j\}$ for $j = 1, 2, \ldots, m$.

Lemma 3.2 *All nonzero 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 nonzero* $\alpha \neq 0$*, where* $\psi_m(\lambda) = \det(\lambda I - H_m)$ *is the characteristic polynomial of* H_m.

Proof Let the polynomial $\chi_j(\lambda) = \tilde{\gamma}_j \lambda^j + \cdots + \tilde{\gamma}_1 \lambda + \tilde{\gamma}_0$ have strict degree $j < m$. We know from (25) that $\chi_j(A) r_0$ has a nonzero component $\tilde{\gamma}_j \zeta_j v_{j+1}$ along $v_{j+1} \in \mathcal{K}_m(A, r_0)$ and thus cannot be orthogonal to $\mathcal{K}_m(A, r_0)$. Since $\psi_m(\lambda) = \det(\lambda I - H_m) = \gamma_m \lambda^m + \cdots + \gamma_1 \lambda + \gamma_0$ is the characteristic polynomial of H_m, we have by the Cayley-Hamilton theorem that $\psi_m(H_m) = 0$. Setting the polynomial $\chi_m(A) = \alpha \psi_m(A)$ in (20) we deduce that $\alpha \psi_m(A) r_0 = \alpha \gamma_m \zeta_m v_{m+1}$ is orthogonal to $\mathcal{K}_m(A, r_0)$. Any polynomial $\varphi_m(\lambda)$ of degree m that is not a scalar multiple of $\psi_m(\lambda)$ can be written as $\varphi_m(\lambda) = \alpha \psi_m(\lambda) + \chi_j(\lambda)$ with $\chi_j(\lambda)$ a nonzero polynomial of degree $j < m$. We have that

$$\varphi_m(A) r_0 = \alpha \psi_m(A) r_0 + \chi_j(A) r_0 = \alpha \gamma_m \zeta_m v_{m+1} + \chi_j(A) r_0 \tag{26}$$

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

Theorem 3.1 *The FOM residual polynomial is a multiple of the characteristic polynomial of the Hessenberg matrix H_m.*

Proof Since the FOM residual polynomial $\tilde{\varphi}_m^{\mathrm{FOM}}(\lambda) = \gamma_m \lambda^m + \cdots + \gamma_1 \lambda + \gamma_0$ has degree m (20) yields an expression for $r^{\mathrm{FOM}} \in \mathcal{K}_{m+1}(A, r_0)$

$$r^{\mathrm{FOM}} = \tilde{\varphi}_m^{\mathrm{FOM}}(A) r_0 = V_m \tilde{\varphi}_m^{\mathrm{FOM}}(H_m) e_1 + \gamma_m \zeta_m v_{m+1}. \tag{27}$$

Thus $\tilde{\varphi}_m^{\mathrm{FOM}}(H_m) = 0$ is necessary to have $r^{\mathrm{FOM}} \perp \mathcal{K}_m(A, r_0)$. By lemma 3.2 we know that the FOM residual polynomial $\tilde{\varphi}_m^{\mathrm{FOM}}(\lambda) = \alpha \psi_m(\lambda)$ has to be a scalar multiple of the characteristic polynomial of H_m in order to eliminate all the components of the residual in $\mathcal{K}_m(A, r_0)$. If H_m is nonsingular then $\psi_m(0) = -\det H_m \neq 0$ and α can be determined from $\tilde{\varphi}_m^{\mathrm{FOM}}(0) = \alpha \psi_m(0) = 1$. We obtain the following expression for the residual polynomial

$$\tilde{\varphi}_m^{\mathrm{FOM}}(\lambda) = \frac{\psi_m(\lambda)}{\psi_m(0)}. \tag{28}$$

This completes the proof. ■

The requirement that H_m has to be nonsingular is related to possible breakdown of FOM. The existence of H_m^{-1} is a necessary and sufficient condition for the existence and the uniqueness of x_m^{FOM} and as a consequence also for the existence and the uniqueness of the corresponding residual r_m^{FOM} as can be seen from (10).

The FOM residual polynomial $\tilde{\varphi}_m^{\mathrm{FOM}}(\lambda)$ is uniquely defined by the normalisation $\tilde{\varphi}_m^{\mathrm{FOM}}(0) = 1$ and by the fact that the m Ritz values $\vartheta_i^{(m)}$ are its zeroes

$$\tilde{\varphi}_m^{\mathrm{FOM}}(\lambda) = \prod_{i=1}^{m} \left(1 - \frac{\lambda}{\vartheta_i^{(m)}} \right). \tag{29}$$

4 The Simpler GMRES Algorithm

Walker and Zhou [15] 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 factorisation of an upper Hessenberg matrix. Their algorithm is based on (17) and is equivalent to the original GMRES algorithm. We briefly outline Simpler GMRES in algorithm 4.1.

Algorithm 4.1 (Simpler GMRES)

Initialisation: $\tilde{r} = z_0 = r_0$
Arnoldi process: *for* $j = 1, 2, \ldots, m$
  compute the matrix-vector product $\tilde{z} = Az_{j-1}$
  Modified Gram-Schmidt orthogonalisation*: for* $i = 1, 2, \ldots, j-1$
    compute the matrix element $\rho_{i,j} = z_i^H \tilde{z}$
    orthogonalise $\tilde{z} = \tilde{z} - \rho_{i,j} z_i$
  compute the norm $\rho_{j,j} = \|\tilde{z}\|_2$
  define the vector $z_j = \dfrac{\tilde{z}}{\rho_{j,j}}$
Projection: *for* $j = 1, 2, \ldots, m$
  compute the inner product $\xi_j = z_j^H \tilde{r}$
  orthogonalise $\tilde{r} = \tilde{r} - \xi_j z_j$
Residual vector: $r_m = \tilde{r}$

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 $R_m \in \mathbb{R}^{m\times m}$ is defined by its elements $\rho_{i,j}$

$$R_m = \begin{pmatrix} \rho_{1,1} & \rho_{1,2} & \cdots & \rho_{1,m} \\ & \rho_{2,2} & \cdots & \rho_{2,m} \\ & & \ddots & \vdots \\ & & & \rho_{m,m} \end{pmatrix}. \tag{30}$$

We have assumed that the dimension of $\mathcal{K}_m(A, r_0)$ is m. Since A is nonsingular this implies that $A\mathcal{K}_m(A, r_0)$ also has dimension m. Hence the Arnoldi process successfully generates a new vector z_j given the set $\{z_1, z_2, \ldots, z_{j-1}\}$. We have that

$$Az_{j-1} = \sum_{i=1}^{j} \rho_{i,j} z_i \tag{31}$$

where the $\rho_{i,j}$ are unique and $\rho_{j,j} \neq 0$. Consequently R_m is nonsingular. The determinant of R_m is equal to the product of the diagonal elements

$$\det R_m = \prod_{j=1}^{m} \rho_{j,j}. \tag{32}$$

We define the matrix $Z_m \in \mathbb{R}^{n\times m}$ by its columns

$$Z_m = \begin{pmatrix} z_0 & z_1 & \dots & z_{m-1} \end{pmatrix}. \tag{33}$$

This is not an orthogonal matrix since z_0 is not used in the orthogonalisation process, but all columns have unit length. From the construction of the vectors z_j, it is clear that $\{z_0, z_1, \dots, z_{m-1}\}$ is a basis for $\mathcal{K}_m(A, r_0)$

$$\mathcal{K}_m(A, r_0) = \text{SPAN}\{z_0, z_1, \dots, z_{m-1}\} = \text{SPAN}\{Z_m\}. \tag{34}$$

The orthogonal matrix $W_m \in \mathbb{R}^{n\times m}$ is defined by its columns

$$W_m = \begin{pmatrix} z_1 & z_2 & \dots & z_m \end{pmatrix}. \tag{35}$$

The upper triangular matrix R_m, computed in the Arnoldi process, satisfies

$$AZ_m = W_m R_m. \tag{36}$$

This is the fundamental relation in Simpler GMRES. Note that the only difference with (4) is that the initial residual is not used in the orthogonalisation process. It is clear that $\{z_1, z_2, \dots, z_m\}$ is an orthonormal basis for $A\mathcal{K}_m(A, r_0)$

$$A\mathcal{K}_m(A, r_0) = \text{SPAN}\{z_1, z_2, \dots, z_m\} = \text{SPAN}\{W_m\}. \tag{37}$$

This implies that the orthogonalisation of the residual vector r^{GMRES} with respect to $A\mathcal{K}_m(A, r_0)$ as required by the minimal residual condition (17) can be done by orthogonalising against z_j for $j = 1, 2, \dots, m$. The residual vector r_m satisfies $W_m^H r_m = 0$ and

$$z_0 = r_0 = r_m + \sum_{j=1}^{m} \xi_j z_j = r_m + W_m w_m \tag{38}$$

where the vector w_m is defined by the scalars ξ_j that have been computed during the orthogonalisation:

$$w_m = \begin{pmatrix} \xi_1 & \xi_2 & \dots & \xi_m \end{pmatrix}^H. \tag{39}$$

The vector y_m is obtained by solving the nonsingular, upper triangular linear system

$$R_m y_m = w_m. \tag{40}$$

The corresponding approximate solution is then given by $Z_m y_m$.

In all the implementations used in this paper, the orthogonalisations in the Arnoldi process were done by a Modified Gram-Schmidt (MGS) algorithm with reorthogonalisation so that the resulting V_m was orthogonal to within machine precision. It is known from a result by Kahan ("twice is enough") that additional orthogonalisations are superfluous. The algorithm and the analysis due to Kahan can be found in a book by Parlett [8]. Instead of MGS with reorthogonalisation a different orthogonalisation technique could be used. From a numerical point of view

Householder reflections are very reliable and the computational complexity is equal to MGS with reorthogonalisation. An implementation of the Arnoldi process with Householder orthogonalisations is given by Walker [14]. The numerical stability of GMRES has been discussed by Drkošová et al. [4]. They showed that loss of orthogonality does not seem to be very significant as long as the linear independence of the Arnoldi vectors is preserved, but pointed out that loss of orthogonality usually leads to loss of linear independence. Their conclusion is that this is a good reason to preserve the orthogonality. Arioli and Fassino [1] stressed that the orthogonality of the Arnoldi basis is a crucial point when spectral similarity properties have to be conserved. In practice full reorthogonalisation is seldom used. A satisfactory compromise between computational complexity and orthogonality to within machine precision is MGS with selective reorthogonalisation. In this case reorthogonalisation is done if and only if $\|v_{j+1}\|_2 \leq \delta \|Av_j\|_2$ where δ is a nonzero positive number (a value suggested in Kahan's analysis is $\delta = 0.83 - \varepsilon$, where $\varepsilon > 0$ is a tiny positive number). This is motivated by the fact that severe cancellation might have occurred if $\|v_{j+1}\|_2$ is significantly smaller than $\|Av_j\|_2$.

5 The Transformation Matrix

We assume that ξ_m in (39) is nonzero. We will later show that $\xi_m = 0$ is equivalent to stagnation of GMRES. 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

$$T_m w_m = e_1 \tag{41}$$

and shift e_{j-1} to e_j

$$T_m e_{j-1} = e_j \tag{42}$$

for $j = 2, 3, \ldots, m$. The components of w_m are given in (39) and hence it is straightforward to write down the explicit representation of the transformation matrix T_m:

$$T_m = \begin{pmatrix} 0 & & & & & 1/\xi_m \\ 1 & 0 & & & & -\xi_1/\xi_m \\ & 1 & 0 & & & -\xi_2/\xi_m \\ & & \ddots & \ddots & & \vdots \\ & & & 1 & 0 & -\xi_{m-2}/\xi_m \\ & & & & 1 & -\xi_{m-1}/\xi_m \end{pmatrix}. \tag{43}$$

The determinant of the transformation matrix T_m can easily be computed

$$\det T_m = (-1)^{m+1}\frac{1}{\xi_m}. \tag{44}$$

Hence T_m is nonsingular. 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

$$\tilde{R}_m = R_m T_m = \begin{pmatrix} \rho_{1,2} & \rho_{1,3} & \cdots & \rho_{1,m} & \tau_1 \\ \rho_{2,2} & \rho_{2,3} & \cdots & \rho_{2,m} & \tau_2 \\ & \rho_{3,3} & \cdots & \rho_{3,m} & \tau_3 \\ & & \ddots & \vdots & \vdots \\ & & & \rho_{m,m} & \tau_m \end{pmatrix} \tag{45}$$

only requires the computation of the m scalars τ_i ($i = 1, 2, \ldots, m$). We obtain an expression for the scalars τ_i by making the inner product of row i of R_m with the last column of T_m and by exploiting the fact that R_m is upper triangular:

$$\tau_1 = \frac{1}{\xi_m}\left(\rho_{1,1} - \sum_{j=2}^{m} \rho_{1,j}\xi_{j-1}\right) \tag{46}$$

$$\tau_i = \frac{-1}{\xi_m}\sum_{j=i}^{m} \rho_{i,j}\xi_{j-1} \quad i = 2, 3, \ldots, m. \tag{47}$$

The determinant of $\tilde{R}_m$ can easily be computed using (32) and (44):

$$\det \tilde{R}_m = \det R_m \det T_m = (-1)^{m+1}\frac{1}{\xi_m}\prod_{j=1}^{m} \rho_{j,j}. \tag{48}$$

However if $\xi_m = 0$, the last vector z_m is orthogonal to the current residual. This means that GMRES stagnates and the residual polynomial is not changed. Hence we get exactly the same results with the previous Krylov subspace.

6 Harmonic Ritz Values

By definition the *Harmonic Ritz values* $\tilde{\vartheta}_i^{(m)}$ *are the reciprocals of the (ordinary) Ritz values of* A^{-1} *computed from* $A\mathcal{K}_m(A, r_0)$. An approximate eigenvector $x = W_m y_m$ is sought in $A\mathcal{K}_m(A, r_0)$. The (ordinary) Ritz criterion is to have the residual of the eigenvalue equation orthogonal to this subspace

$$\left(A^{-1}x - \mu x\right) \perp A\mathcal{K}_m(A, r_0) \Leftrightarrow W_m^H \left(A^{-1}W_m y_m - \mu W_m y_m\right) = 0. \tag{49}$$

From (36) we derive the following expression for the action of A^{-1} on W_m

$$A^{-1}W_m = Z_m R_m^{-1}. \tag{50}$$

The projected eigenvalue problem (49) becomes

$$W_m^H Z_m R_m^{-1} y_m = \mu y_m. \tag{51}$$

Using (33) and (35), an expression for $W_m^H Z_m$ is obtained when z_0 is replaced by (38)

$$W_m^H Z_m = \begin{pmatrix} \xi_1 & 1 & & & \\ \xi_2 & 0 & 1 & & \\ \vdots & & \ddots & \ddots & \\ \xi_{m-1} & & & 0 & 1 \\ \xi_m & & & & 0 \end{pmatrix} = T_m^{-1}. \quad (52)$$

With this result we can simplify (51) to

$$T_m^{-1} R_m^{-1} y_m = \mu y_m. \quad (53)$$

Equation (53) clearly shows that the eigenvalues of $\tilde{R}_m$ are the reciprocals of the (ordinary) Ritz values of A^{-1} computed from $A\mathcal{K}_m(A, r_0)$

$$\tilde{R}_m y_m = \frac{1}{\mu} y_m. \quad (54)$$

At this point we can prove that the Harmonic Ritz values are eigenvalue approximations according to the minimal residual criterion.

Theorem 6.1 *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 $\mathcal{K}_m(A, r_0)$. The minimal residual criterion requires that the residual of the eigenvalue equation is orthogonal to $A\mathcal{K}_m(A, r_0)$. The columns of W_m form an orthonormal basis for this subspace as can be seen from (37)

$$(Ax - \mu x) \perp A\mathcal{K}_m(A, r_0) \Leftrightarrow W_m^H (A Z_m y_m - \mu Z_m y_m) = 0. \quad (55)$$

Using (36) we have that

$$R_m y_m = \mu W_m^H Z_m y_m. \quad (56)$$

Equation (52) allows us to simplify the projected eigenvalue problem (56) to

$$T_m R_m y_m = \mu y_m. \quad (57)$$

The eigenvalues of $T_m R_m$ are also eigenvalues of the similar matrix $R_m T_m$ as can be seen by a similarity transformation with the nonsingular matrix T_m since $T_m (R_m T_m) T_m^{-1} = T_m R_m$. The corresponding eigenvector is given by $y_m = T_m z_m$, satisfying

$$R_m T_m z_m = \mu z_m. \quad (58)$$

Hence the eigenvalue approximations according to the minimal residual criterion are the Harmonic Ritz values. ∎

The Harmonic Ritz values can also be computed from (4). An approximate eigenvector $x = V_m y_m$ is sought in $\mathcal{K}_m(A, r_0)$. The minimal residual criterion requires that the residual of the eigenvalue equation is orthogonal to $A\mathcal{K}_m(A, r_0)$

$$(Ax - \mu x) \perp A\mathcal{K}_m(A, r_0) \Leftrightarrow (AV_m)^H (AV_m y_m - \mu V_m y_m) = 0. \tag{59}$$

Using (4) we obtain the generalised eigenvalue problem

$$\bar{H}_m^H \bar{H}_m y_m = \mu H_m^H y_m. \tag{60}$$

If H_m is nonsingular, (60) can be rewritten as

$$H_m^{-H} \bar{H}_m^H \bar{H}_m y_m = \mu y_m. \tag{61}$$

Equation (15) allows us to replace (61) with the eigenvalue problem

$$\left(H_m + h_{m+1,m}^2 f_m e_m^H\right) y_m = \mu y_m. \tag{62}$$

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 (62)

$$\|h_{m+1,m}^2 f_m e_m^H\|_2 \leq \frac{h_{m+1,m}^2}{\sigma_{\min}(H_m)}. \tag{63}$$

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 (63) shows that the differences between the Harmonic Ritz values and the 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. [9] 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 values and the Harmonic Ritz values to be small in the case of a real, symmetric matrix. We show that the differences between the Ritz values and the Harmonic Ritz values can be large when GMRES stagnates.

7 GMRES Residual Polynomial

In this section we show that the GMRES residual polynomial $\tilde{\varphi}_m^{\mathrm{GMRES}}(\lambda)$ is a multiple of the characteristic polynomial of $\tilde{R}_m$. This implies that the Harmonic Ritz values are the zeroes of the GMRES residual polynomial or equivalently that the GMRES approximation polynomial $\varphi_{m-1}^{\mathrm{GMRES}}(\lambda)$ interpolates λ^{-1} in the Harmonic Ritz values.

Lemma 7.1 *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 the following expression*

$$\chi_m(A) r_0 = W_m \chi_m(\tilde{R}_m) w_m + \gamma_0 r_m. \tag{64}$$

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 nonzero 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

$$Ar_0 = z_1 \rho_{1,1} = W_m u_1 = W_m \tilde{R}_m w_m. \tag{65}$$

By induction we prove that

$$A^k r_0 = W_m u_k = W_m \tilde{R}_m^k w_m \tag{66}$$

for $k = 1, 2, \ldots, m$. Equation (65) shows that (66) is valid for $k = 1$. We assume that (66) is valid for $k = 1, 2, \ldots, j$ and prove that it is also valid for $k = j + 1$. Using (36) we have for $j < m$

$$\begin{aligned} A^{j+1} r_0 &= AA^j r_0 = AW_m u_j \\ &= Az_1 e_1^H u_j + Az_2 e_2^H u_j + \cdots + Az_j e_j^H u_j \\ &= W_m R_m e_2 e_1^H u_j + W_m R_m e_3 e_2^H u_j + \cdots + W_m R_m e_{j+1} e_j^H u_j \\ &= W_m R_m T_m u_j = W_m \tilde{R}_m u_j = W_m u_{j+1} \\ &= W_m \tilde{R}_m^{j+1} w_m. \end{aligned} \tag{67}$$

We use (66) in combination with the expansion (38) of the initial residual to find an expression for $\chi_m(A) r_0$ in terms of W_m and r_m:

$$\begin{aligned} \chi_m(A) r_0 &= \gamma_m A^m r_0 + \cdots + \gamma_1 A r_0 + \gamma_0 r_0 \\ &= \gamma_m W_m u_m + \cdots + \gamma_1 W_m u_1 + \gamma_0 (r_m + W_m w_m) \qquad (68) \\ &= \gamma_m W_m \tilde{R}_m^m w_m + \cdots + \gamma_1 W_m \tilde{R}_m w_m + \gamma_0 (r_m + W_m w_m) \\ &= W_m \chi_m(\tilde{R}_m) w_m + \gamma_0 r_m. \qquad (69) \end{aligned}$$

Equation (69) is the desired result. ■

The last nonzero 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 (66) we can conclude that

$$A^j r_0 = W_m u_j = \tilde{\zeta}_j z_j + \hat{z}_j \tag{70}$$

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

Lemma 7.2 *All nonzero 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 nonzero $\alpha \neq 0$, 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 nonzero constant term $\tilde{\gamma}_0 \neq 0$ then we can see from (68) and (70) that $\chi_j(A) r_0$ has a nonzero 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 nonzero. On the other hand if this polynomial does not have a constant term $\tilde{\gamma}_0 = 0$ then we know from (70) that $\chi_j(A)r_0$ has a nonzero component $\tilde{\gamma}_j\zeta_j z_j$ 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 (64) 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 nonzero polynomial of degree $j < m$. We have that

$$\varphi_m(A)r_0 = \alpha\tilde{\psi}_m(A)r_0 + \chi_j(A)r_0 = \alpha\gamma_0 r_m + \chi_j(A)r_0 \tag{71}$$

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

Theorem 7.1 *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 (64) yields an expression for $r^{\mathrm{GMRES}} \in \mathcal{K}_{m+1}(A, r_0)$

$$r^{\mathrm{GMRES}} = \tilde{\varphi}_m^{\mathrm{GMRES}}(A)r_0 = W_m\tilde{\varphi}_m^{\mathrm{GMRES}}(\tilde{R}_m)w_m + \gamma_0 r_m. \tag{72}$$

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 7.2 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 (48)

$$\tilde{\psi}_m(0) = -\det\tilde{R}_m = (-1)^m\frac{1}{\xi_m}\prod_{j=1}^{m}\rho_{j,j}. \tag{73}$$

We obtain the following expression for the residual polynomial

$$\tilde{\varphi}_m^{\mathrm{GMRES}}(\lambda) = \frac{\tilde{\psi}_m(\lambda)}{\tilde{\psi}_m(0)}. \tag{74}$$

This completes the proof. ■

The GMRES residual polynomial is uniquely defined by the normalisation $\tilde{\varphi}_m^{\mathrm{GMRES}}(0) = 1$ and by the fact that its m zeroes are the Harmonic Ritz values $\tilde{\vartheta}_i^{(m)}$

$$\tilde{\varphi}_m^{\mathrm{GMRES}}(\lambda) = \prod_{i=1}^{m}\left(1 - \frac{\lambda}{\tilde{\vartheta}_i^{(m)}}\right). \tag{75}$$

8 Breakdown of FOM and 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

$$A_n = \begin{pmatrix} 0 & 1 & & \\ & \ddots & \ddots & \\ & & 0 & 1 \\ 1 & & & 0 \end{pmatrix} \in \mathbb{R}^{n\times n}, \quad x_n = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} \in \mathbb{R}^n \text{ and } b_n = \begin{pmatrix} 0 \\ \vdots \\ 0 \\ 1 \end{pmatrix} \in \mathbb{R}^n. \tag{76}$$

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 (4) is reduced to

$$Av_1 = \begin{pmatrix} v_1 & v_2 \end{pmatrix} \bar{H}_1 \tag{77}$$

where $\bar{H}_1$ is given by

$$\bar{H}_1 = \begin{pmatrix} 0 \\ 1 \end{pmatrix}. \tag{78}$$

In this case FOM clearly breaks down since $H_1 = \begin{pmatrix} 0 \end{pmatrix}$ is singular. From (19) we see that FOM sets the Ritz value $\vartheta^{(1)} = 0$ equal to zero. The FOM residual polynomial

$$\tilde{\varphi}_1^{\mathrm{FOM}}(\lambda) = 1 - \frac{\lambda}{\vartheta^{(1)}} \tag{79}$$

evaluated for nonzero λ clearly shows that the norm of the residual $\|r^{\mathrm{FOM}}\|_2$ may grow without bound when FOM encounters a breakdown.

In this case GMRES clearly stagnates. From (60) we see that the Harmonic Ritz value $\tilde{\vartheta}^{(1)} = \infty$ is set equal to infinity. Hence the GMRES residual polynomial

$$\tilde{\varphi}_1^{\mathrm{GMRES}}(\lambda) = 1 - \frac{\lambda}{\tilde{\vartheta}^{(1)}} \tag{80}$$

evaluated for finite λ clearly shows the stagnation.

The fundamental relation (36) computed by Simpler GMRES is reduced to

$$Av_1 = v_2 R_1 \tag{81}$$

where the matrix $R_1 = \begin{pmatrix} 1 \end{pmatrix}$ is nonsingular since $\mathcal{K}_m(A, r_0)$ has maximal dimension. The stagnation in Simpler GMRES can be seen from the fact that the corresponding projection $\xi_1 = 0$ is zero.

We conclude that if a vector is added to $A\mathcal{K}_m(A, r_0)$ that is orthogonal to the residual, both FOM and GMRES do not make any progress since no better approximations are available. This is referred to as breakdown of FOM which means that a Ritz value equals zero and stagnation of GMRES which means that an Harmonic Ritz value equals infinity. These observations explain the peaks in the FOM residual curve when a Ritz value becomes very small. The corresponding GMRES residual curve stagnates as the Harmonic Ritz values become very large.

9 A Simple Test Case

We are concerned with the eigenvalue estimates when GMRES nearly stagnates. The matrix A_n is given in (76) 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 $\lambda^n = 1$, meaning that 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 $\bar{H}_m^*$ defined in (82) and V_m^* defined in (83). It is straightforward to verify that $h_{i,j} = \mathcal{O}(\varepsilon^2)$ for $i = 2, 3, \dots, m$ and for $j = i+1, i+2, \dots, m$ so we set these elements to zero in this analysis and approximate $\bar{H}_m$ by $\bar{H}_m^*$:

$$\bar{H}_m^* = \begin{pmatrix} 2\varepsilon & 2\varepsilon & \dots & 2\varepsilon \\ 1 & & & \\ & 1 & & \\ & & \ddots & \\ & & & 1 \end{pmatrix} \in \mathbb{R}^{(m+1)\times m}. \tag{82}$$

In order to normalise the vectors in V_m^* we compute the norm $\|v_i^*\|_2 = 1 + \varepsilon + \mathcal{O}(\varepsilon^2)$ for $i = 1, 2, \dots, m$. This norm $\|v_i^*\|_2 \approx 1$ is approximated by one. The orthogonality of the vectors v_i^* and v_j^* is verified by computing the inner product $(v_i^*)^H v_j^* = \mathcal{O}(\varepsilon^2)$ for $i \neq j$. In this analysis V_m is approximated by V_m^*:

$$V_m^* = \begin{pmatrix} \varepsilon & \varepsilon & \dots & \varepsilon \\ \vdots & \vdots & & \vdots \\ \varepsilon & \varepsilon & \dots & \varepsilon \\ \varepsilon & \varepsilon & & 1+\varepsilon \\ \varepsilon & \varepsilon & \iddots & -\varepsilon \\ \varepsilon & 1+\varepsilon & & \vdots \\ 1+\varepsilon & -\varepsilon & \dots & -\varepsilon \end{pmatrix} \in \mathbb{R}^{m\times n}. \tag{83}$$

The Ritz values $\vartheta^{(m)}$ are the eigenvalues of H_m. The characteristic polynomial of H_m^* can easily be obtained

$$\psi_m(\lambda) = -\lambda^m + 2\varepsilon\lambda^{m-1} + 2\varepsilon\lambda^{m-2} + \dots + 2\varepsilon\lambda + 2\varepsilon = 0. \tag{84}$$

Using a well known summation formula (84) can be rewritten as

$$-\lambda^m + 2\varepsilon\left(\frac{1-\lambda^m}{1-\lambda}\right) = 0. \tag{85}$$

The Ritz values $\vartheta^{(m)}$ for this test case tend to zero as ε tends to zero. Hence we can approximate $1 - \lambda^m \approx 1$ and we have that $1 - \lambda = \mathcal{O}(1)$

$$\lambda^m \approx \frac{2\varepsilon}{1-\lambda} = \mathcal{O}(2\varepsilon). \tag{86}$$

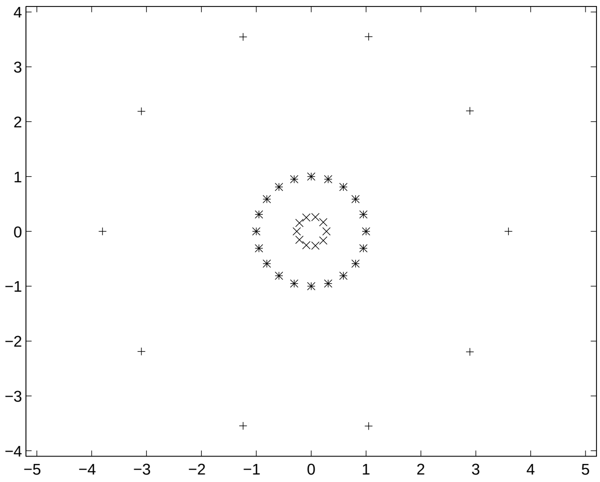

Figure 1: The Ritz ($\times$) and Harmonic Ritz ($+$) values computed from $\mathcal{K}_{10}(A_{20}, b_{20}(10^{-6}))$ for the stagnation test case.

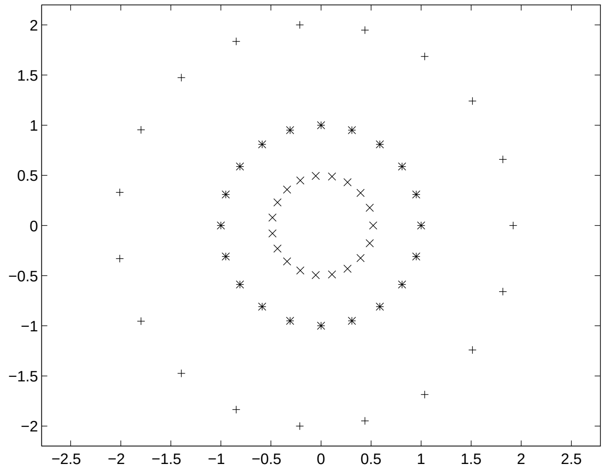

Figure 2: The Ritz ($\times$) and Harmonic Ritz ($+$) values computed from $\mathcal{K}_{19}(A_{20}, b_{20}(10^{-6}))$ for the stagnation test case.

With these approximations we see that the Ritz values $\vartheta^{(m)}$ are

$$|\vartheta^{(m)}| = \mathcal{O}\left(\sqrt[m]{2\varepsilon}\right). \tag{87}$$

Equation (86) shows that the Ritz values approximately form a regular m-polygon in the complex plane.

The Harmonic Ritz values $\tilde{\vartheta}^{(m)}$ are the eigenvalues of the generalised problem (60). Since $\bar{H}_m$ is approximated by $\bar{H}_m^*$, we can approximate $\bar{H}_m^H \bar{H}_m$ by the identity I_m:

$$\bar{H}_m^H \bar{H}_m \approx (\bar{H}_m^*)^H \bar{H}_m^* = I_m + 4\varepsilon^2 cc^H \approx I_m \tag{88}$$

where $c = \begin{pmatrix} 1 & 1 & \dots & 1 \end{pmatrix}^H \in \mathbb{R}^m$. With this approximation we can rewrite the generalised eigenvalue problem (60) from which the Harmonic Ritz values $\tilde{\vartheta}^{(m)}$ are computed as

$$y_m^H H_m = y_m^H \frac{1}{\mu^H}. \tag{89}$$

Equation (89) allows us to conclude that the Harmonic Ritz $\tilde{\vartheta}^{(m)}$ values are the reciprocals of the complex conjugates of the ordinary Ritz values $\vartheta^{(m)}$:

$$\tilde{\vartheta}^{(m)} = \frac{1}{\left(\vartheta^{(m)}\right)^H}. \tag{90}$$

From (87) and (90) we see that the Harmonic Ritz values $\tilde{\vartheta}^{(m)}$ are

$$|\tilde{\vartheta}^{(m)}| = \mathcal{O}\left(\frac{1}{\sqrt[m]{2\varepsilon}}\right). \tag{91}$$

Table 1: Estimated and computed norm of the Ritz and Harmonic Ritz values computed from $\mathcal{K}_m(A_{20}, b_{20}(10^{-6}))$ for the stagnation test case.

	Ritz Values			Harmonic Ritz Values		
m	min	est.	max	min	est.	max
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

Hence the Harmonic Ritz values $\tilde{\vartheta}^{(m)}$ go off to infinity as ε tends to zero. Table 1 shows the quality of the estimations (87) and (91), which increases with decreasing ε and decreasing m.

From (82) 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 (63) 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 will be large. Equations (82) and (83) allow us to estimate the norm of the FOM residual $\|r^{\mathrm{FOM}}\|_2 \approx 1/(2\varepsilon)$.

These results show that the differences between the Ritz values and the Harmonic Ritz values are significantly large when GMRES (nearly) stagnates. To illustrate this we show in Fig. 1 the Ritz values and the 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 times sign ($\times$). An asterisk ($*$) is used to plot the eigenvalues. Figure 2 shows the Ritz values and the Harmonic Ritz values computed from $\mathcal{K}_{19}(A_{20}, b_{20}(10^{-6}))$.

10 A Convection-Diffusion Problem

This test case was taken from [5]. Our aim is to compute the steady-state solution of a linearised 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

$$-\nu\nabla^2 u + \vec{w}\nabla u = f \text{ in } \Omega \tag{92}$$

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 unit length $\|\vec{w}\|_2 = 1$ and has a $45°$ inclination with the grid lines. The diffusion parameter was set to $\nu = 10^{-6}$, resulting in a "mesh Peclet number" of $\mathrm{Pe} = \|w\|_2 h/(2\nu) = 7812.5 > 1$, which clearly shows that the convection is dominant.

After SUPG finite element discretisation a linear system $Au = f$ is obtained in which the unknown u is the discretised version of the unknown in (92) and the vector f corresponds to the right-hand side. The nonsymmetric matrix $A = \nu L + N$ is the sum of the discretisation 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 discretised Laplacian and a matrix S corresponding to

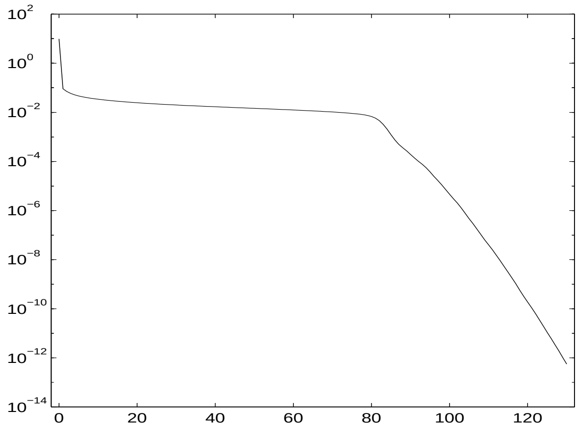

Figure 3: The norm of the residual $\|r^{\mathrm{GMRES}}\|_2$ as a function of m for the convection-diffusion problem.

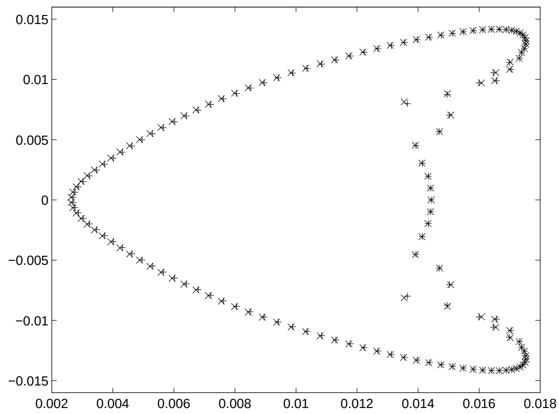

Figure 4: The Ritz ($\times$) and Harmonic Ritz ($+$) values computed from $\mathcal{K}_{130}(A, f)$ for the convection-diffusion problem.

stabilisation terms, since streamline upwinding is used. We use unpreconditioned GMRES to solve the linear system $Au = f$.

Figure 3 shows the residual norm $\|r^{\mathrm{GMRES}}\|_2$ as a function of the dimension m of $\mathcal{K}_m(A, f)$. In Fig. 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 times sign ($\times$). The Krylov subspace $\mathcal{K}_{130}(A, f)$ of dimension $m = 130$ yields an accurate solution. Hence the differences between the Ritz values and Harmonic Ritz values are small. Figure 4 shows the Ritz values and Harmonic Ritz values computed from $\mathcal{K}_{130}(A, f)$. To illustrate that the Ritz values and the Harmonic Ritz values differ significantly when GMRES (nearly) stagnates, we show in Fig. 5 the Ritz values and the Harmonic Ritz values computed from $\mathcal{K}_{41}(A, f)$. Figure 6 shows the Ritz values and the Harmonic Ritz values computed from $\mathcal{K}_{80}(A, f)$.

11 Concluding Remarks

Nachtigal et al. [7] 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.

In this paper we indicate that the zeroes of the GMRES residual polynomial can be computed by solving the eigenvalue problem (60) or (62). This is not only a cheap but also a stable procedure.

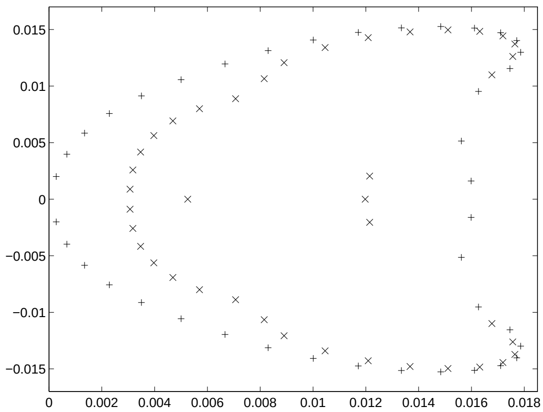

Figure 5: The Ritz ($\times$) and Harmonic Ritz ($+$) values computed from $\mathcal{K}_{41}(A, f)$ for the convection-diffusion problem.

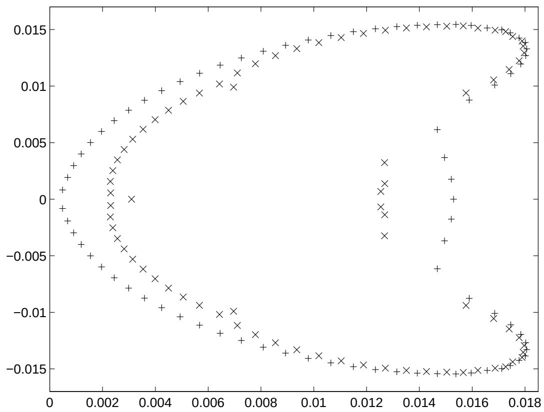

Figure 6: The Ritz ($\times$) and Harmonic Ritz ($+$) values computed from $\mathcal{K}_{80}(A, f)$ for the convection-diffusion problem.

Toh and Trefethen [13] advocate the use of $\bar{H}_m$ because it bypasses the usual consideration of Ritz values or "Arnoldi eigenvalue estimates". For highly nonnormal matrices we cannot expect the Arnoldi iteration to be effective at determining eigenvalues. On the oher hand, Greenbaum et al. [6] have shown that eigenvalues cannot be used to predict the convergence of GMRES for highly nonnormal matrices. Toh and Trefethen argue that in order to analyse the convergence of unsymmetric Krylov solvers plots of the Ritz values should be made, but they point out that this is not sufficient.

Therefore we suggest that in order to analyse the convergence of unsymmetric Krylov solvers both the Ritz spectrum and the Harmonic Ritz spectrum should be plotted. The cost of computing the Harmonic Ritz values only depends on m, which in practice is small, while the computation of pseudospectra and lemniscates is usually far too expensive.

Acknowledgements

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. The scientific responsibility rests with its authors. Serge Goossens is financed by a specialisation scholarship of the Flemish Institute for the Promotion of Scientific and Technological Research in Industry (IWT).

References

[1] M. Arioli and C. Fassino, Roundoff error analysis of algorithms based on Krylov subspace methods, *BIT*, 36 (1996), 189–206.

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

[3] P. N. Brown, A Theoretical Comparison of the Arnoldi and GMRES algorithms, *SIAM J. Sci. Stat. Comput.*, 12 (1991): 58–78.

[4] J. Drkošová, A. Greenbaum, M. Rozložník and Z. Strakoš, Numerical Stability of GMRES, *BIT*, 35 (1995), 309–330.

[5] H. C. Elman, D. J. Silvester and A. J. Wathen, *Iterative Methods for Problems in Computatioanal Fluid Dynamics*, Technical Report 96/19, Oxford University Computing Laboratory, 1996.

[6] A. Greenbaum, V. Pták and Z. Strakoš, Any nonincreasing convergence curve is possible for GMRES, *SIAM J. Matrix Anal. Applics.*, 17 (1996): 465–469.

[7] N. M. Nachtigal, L. Reichel and L. N. Trefethen, A hybrid GMRES algorithm for nonsymmetric linear systems, *SIAM J. Matrix Anal. Applics.*, 13 (1992): 796–825.

[8] B. N. Parlett, *The Symmetric Eigenvalue Problem*, Prentice-Hall, 1980.

[9] C. C. Paige, B. N. Parlett and H. A. Van der Vorst, Approximate Solutions and Eigenvalue Bounds from Krylov Subspaces, *Numerical Linear Algebra with Applications*, 2 (1995): 115–134.

[10] Y. Saad, *Numerical Methods for Large Eigenvalue Problems*, Manchester University Press, 1992.

[11] Y. Saad, *Iterative Methods for Sparse Linear Systems*, PWS Publishing Company, 1996.

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

[13] K. C. Toh and L. N. Trefethen, Calculation of Pseudospectra by the Arnoldi Iteration, *SIAM J. Sci. Comput.*, 17 (1996): 1–15.

[14] H. F. Walker, Implementation of the GMRES method using Householder transformations, *SIAM J. Sci. Stat. Comput.*, 9 (1988): 152–163.

[15] H. F. Walker and L. Zhou, A Simpler GMRES, *Numerical Linear Algebra with Applications*, 1 (1994): 571–581.