Spectral information and GMRES convergence

Gérard MEURANT

Joint work with J. Duintjer Tebbens, H. Sadok and Z. Strakoš

June 2012

We solve the linear system

$$Ax = b,$$

with GMRES (Saad and Schultz (1986))

A is a (real) matrix of order n and $x_0 = 0$, $\|b\| = 1$

The question we would like to address is:

what determines GMRES convergence?

GMRES

GMRES uses the Arnoldi process to construct an orthonormal basis of the Krylov susbpsace

$$\mathcal{K}_n(A, b) = \{b \quad Ab \quad \cdots \quad A^{n-1}b\}$$

Assume GMRES runs until iteration n. Then,

$$AV = VH, \quad V^T V = I,$$

and H is upper Hessenberg

The GMRES iterates x_k are computed by solving

$$\min_{x_k \in \mathcal{K}_k(A,b)} \|b - Ax_k\|$$

We have $(A, b) \equiv (H, e_1)$ i.e. they give the same residual norms

Many things have been proposed over the years to explain GMRES convergence:

- Eigenvalues of A
- Pseudo-eigenvalues
- Polynomial numerical hull
- ...

None of these is really satisfactory for all cases

It is generally said that GMRES convergence depends on the eigenvalues when A is normal ($A^TA = AA^T$)

Results

We will show that GMRES residual norm convergence depends on

- the eigenvalues and eigenvectors of A when the matrix is normal

- the eigenvalues and eigenvectors of an orthogonal matrix depending on the orthonormal bases of $\mathcal{K}_n(A, b)$ and $A\mathcal{K}_n(A, b)$ when the matrix may be non-normal

The results of Arioli, Greenbaum, Pták and Strakoš

They gave a parametrization of the class of matrices A and right-hand sides b giving a prescribed residual norm convergence curve (1998)

Assume we are given $n+1$ positive numbers

$$f_0 \geq f_1 \geq \cdots \geq f_{n-1} > 0, \quad f_n = 0$$

and n complex numbers $\lambda_1, \ldots, \lambda_n$ all different from 0

The following assertions are equivalent:

- The spectrum of A is $\{\lambda_1, \dots, \lambda_n\}$ and GMRES(A, b) yields residuals $r_j,\ j = 0, \dots, n-1$ such that

$$\|r_j\| = f_j, \quad j = 0, \dots, n-1$$

- The matrix A is of the form $A = WYCY^{-1}W^T$ and $b = Wh$, where W is an orthogonal matrix and

$$Y = \begin{pmatrix} h & \begin{matrix} R \\ 0 \end{matrix} \end{pmatrix}$$

R being any nonsingular upper triangular matrix of order $n-1$, h a vector such that

$$h = (\eta_1, \dots, \eta_n)^T, \quad \eta_j = (f_{j-1}^2 - f_j^2)^{1/2} = (\|r_{j-1}\|^2 - \|r_j\|^2)^{1/2}$$

and C is the companion matrix corresponding to the prescribed eigenvalues

This result is translated by many people into:
"GMRES convergence does not depend on the eigenvalues of A"

In fact, it means that if for (A, b) one has a given GMRES convergence curve, one can find another pair with the same convergence curve and any prescribed eigenvalues by changing C in the parametrization. b stays the same if we do not change W

J. Duintjer Tebbens and GM have shown that in addition to the convergence curve one can also prescribe the Ritz values at all GMRES iterations

The Ritz values (eigenvalues of H_k) are generally taken as approximations of the eigenvalues of A

This result means that there are matrices (and initial vectors) for which there is no convergence of the Ritz values

Normal matrices

A normal $\rightarrow$ H normal

Theorem

Let H be an unreduced Hessenberg matrix having real positive subdiagonal entries with eigenvalues λ_i, $i = 1, \ldots, n$ and let C be the corresponding companion matrix. Let

$$U = \begin{pmatrix} e_1 & He_1 & \cdots & H^{n-1}e_1 \end{pmatrix}$$

which is an upper triangular matrix with real positive subdiagonal entries. Then

$$H = UCU^{-1}$$

and the 2 following statements are equivalent:

Theorem (2)

1- *H is normal*

2- *There exist real positive weights ω_k with $\sum_{k=1}^{n} \omega_k = 1$ such that $M = U^T U$, where M is the moment matrix with entries defined by*

$$M_{i,j} = \sum_{k=1}^{n} \omega_k (\bar{\lambda}_k)^{i-1} \lambda_k^{j-1}$$

The equivalence of assertions 1 and 2 was first proven by Parlett (1973)

Then we can relate the GMRES residual norms and the moment matrix M

GMRES convergence for normal matrices

Theorem

Let A be a normal matrix with a spectral factorization $Z\Lambda Z^*$ with Λ diagonal and Z unitary and b a vector of unit norm
Applying GMRES to (A, b), the residual norms are given by

$$\|r_k\|^2 = \frac{1}{(M_{k+1}^{-1})_{1,1}},$$

where M is the moment matrix and M_{k+1} its principal submatrix of order $k+1$
The weights ω_k are the squares of the moduli of the components of the vector $c = Z^* b$

Now we have to compute $(M_{k+1}^{-1})_{1,1}$

$$M_{k+1} = \mathcal{V}_{k+1}^* D_\omega \mathcal{V}_{k+1}$$

with

$$\mathcal{V}_{k+1} = \begin{pmatrix} 1 & \lambda_1 & \cdots & \lambda_1^k \\ 1 & \lambda_2 & \cdots & \lambda_2^k \\ \vdots & \vdots & & \vdots \\ 1 & \lambda_n & \cdots & \lambda_n^k \end{pmatrix}$$

an $n \times k$ Vandermonde matrix and D_ω a diagonal matrix of order n with $\omega_1, \ldots, \omega_n$ on the diagonal

Using Cramer's rule, we have

$$\begin{aligned} (M_{k+1}^{-1})_{1,1} &= \frac{1}{\det(M_{k+1})} \det \begin{pmatrix} 1 & m_{1,2} & \cdots & m_{1,k} \\ 0 & & & \\ \vdots & & M_{2:k+1,2:k+1} & \\ 0 & & & \end{pmatrix} \\ &= \frac{\det(M_{2:k+1,2:k+1})}{\det(M_{k+1})} \end{aligned}$$

We use the Cauchy-Binet formula to compute the determinant of $M_{2:k+1,2:k+1}$

$$\det(M_{2:k+1,2:k+1}) = \sum_{1\leq i_1<\cdots<i_k\leq n} \left[\prod_{j=1}^{k} \omega_{i_j} |\lambda_{i_j}|^2\right] |\det(\mathcal{V}(\lambda_{i_1},\ldots,\lambda_{i_k}))|^2$$

where

$$\mathcal{V}(\lambda_{i_1},\ldots,\lambda_{i_k}) = \begin{pmatrix} 1 & \lambda_{i_1} & \cdots & \lambda_{i_1}^{k-1} \\ 1 & \lambda_{i_2} & \cdots & \lambda_{i_2}^{k-1} \\ \vdots & \vdots & & \vdots \\ 1 & \lambda_{i_k} & \cdots & \lambda_{i_k}^{k-1} \end{pmatrix}$$

The determinant is

$$\det(\mathcal{V}(\lambda_{i_1},\ldots,\lambda_{i_k})) = \prod_{i_1\leq i_\ell<i_j\leq i_k} (\lambda_{i_j} - \lambda_{i_\ell})$$

We do the same for $\det(M_{k+1})$

Theorem

Let A be a normal matrix with a spectral factorization $Z\Lambda Z^*$ with Λ diagonal (with diagonal entries λ_i) and Z unitary and b a vector of unit norm

Applying GMRES to (A, b), the residual norms are given by

$$\|r_k\|^2 = \frac{\sum_{1\le i_1<\cdots<i_{k+1}\le n}\left[\prod_{j=1}^{k+1}\omega_{i_j}\right]\prod_{i_1\le i_\ell<i_j\le i_{k+1}}|\lambda_{i_j}-\lambda_{i_\ell}|^2}{\sum_{1\le i_1<\cdots<i_k\le n}\left[\prod_{j=1}^{k}\omega_{i_j}|\lambda_{i_j}|^2\right]\prod_{i_1\le i_\ell<i_j\le i_k}|\lambda_{i_j}-\lambda_{i_\ell}|^2}$$

The weights ω_i are given by $\omega_i = |c_i|^2$ and $c = Z^* b$

This completely describes GMRES convergence for normal matrices

Convergence depends on the eigenvalues of A and on its eigenvectors and b through the weights

The general case

The matrices B such that

$$B\mathcal{K}_k(B,b) \equiv A\mathcal{K}_k(A,b), \;\; k = 1, 2, \ldots, n.$$

are called GMRES(A,b)-equivalent matrices. They give the same residual norms

Theorem (Greenbaum and Strakoš (1994))

Let W be a unitary matrix whose columns give a basis of $A\mathcal{K}_n(A,b)$ and $\mathcal{H}$ an unreduced upper Hessenberg matrix such that $AW = W\mathcal{H}$. Then, the following assertions are equivalent:

1- B *is GMRES*(A,b)*-equivalent,*

2- $B = W\tilde{R}\mathcal{H}W^*$, *where $\tilde{R}$ is any nonsingular upper triangular matrix*

$$(A, b) \rightarrow H$$
$$(A, Ab) \rightarrow \mathcal{H}$$

$$H = Q\mathcal{R}, \qquad \mathcal{H} = \mathcal{R}Q$$

where Q is an upper Hessenberg orthogonal matrix with positive real entries in the first row and $\mathcal{R}$ is upper triangular

In practice Q is never computed in GMRES. It is known implicitly through the (product of) the rotation matrices used to solve the least squares problems

The entries of Q are explicitly known as a function of the values $\eta_j, j = 1, \ldots, n$

$$h = (\eta_1, \ldots, \eta_n)^T, \quad \eta_j = (\|r_{j-1}\|^2 - \|r_j\|^2)^{1/2}$$

Theorem

The entries of the real upper Hessenberg matrix Q are explicitly known as functions of η_j's. The first row of Q is h^T

Moreover, $Q = V^T W$, where V (resp. W) is an orthogonal matrix whose columns give the orthonormal basis of $\mathcal{K}_n$ (resp. $A\mathcal{K}_n$ such that the vector $W^T b$ is real positive) given by GMRES and we have $H = Q\mathcal{H}Q^*$

The matrix Q completely describes GMRES convergence

$$A = VQ\mathcal{R}V^T, \quad b = Ve_1$$

- GMRES convergence is "contained" in Q
- The non-normality of A is "contained" in $\mathcal{R}$

From these results we can construct orthogonal matrices and right-hand sides which give the same residual norms as (A, b)

Theorem

Let V (resp. W) be the orthonormal basis of the Krylov space $\mathcal{K}_n(A, b)$ (resp. $A\mathcal{K}_n(A, b)$) such that the vector $W^T b$ is positive)

The orthogonal matrix $B = WV^T$ is GMRES(A, b)-equivalent

Moreover, the matrix $Q = V^T W$ with the right-hand side $W^T b$ also gives the same GMRES residual norm convergence curve as (A, b)

An other pair giving the same residual norms is (Q, e_1)

The residual norm convergence curve of GMRES for (A, b) is determined by the eigenvalues in the generalized eigenvalue problem

$$V^T x = \delta W^T x$$

and by the weights ω_i given by $\omega_i = |c_i|^2$ with $c = Z^* b$, Z being the matrix of the normalized generalized eigenvectors

Note that V and W depend on A and b

Then we can use the previous results for normal matrices

$$\|r_k\|^2 = \frac{\sum_{1\le i_1<\cdots<i_{k+1}\le n} \prod_{j=1}^{k+1} \omega_{i_j} \prod_{i_1\le i_\ell<i_j\le i_{k+1}} |\delta_{i_j} - \delta_{i_\ell}|^2}{\sum_{1\le i_1<\cdots<i_k\le n} \prod_{j=1}^{k} \omega_{i_j} \prod_{i_1\le i_\ell<i_j\le i_k} |\delta_{i_j} - \delta_{i_\ell}|^2}$$

where $\omega_i = |c_i|^2$ with $c = Z^* b$

The weights ω_i can also be written as the squares of the moduli of the first components of the eigenvectors of $Q = V^T W$ since the pair (Q, e_1) gives the same residual norms

GMRES convergence is governed by the distribution of the eigenvalues of $V^T W$ on the unit circle (and also the weights ω_i)

A numerical example

A convection–diffusion simple problem, see Fischer, Ramage, Silvester and Wathen (1999)

$$-\nu \Delta u + w \cdot \nabla u = 0$$

with $w = [0, 1]^T$ in $\Omega = (0, 1)^2$ with Dirichlet boundary conditions $u = g$ on $\partial\Omega$

Discretization: stabilized Petrov–Galerkin SUPG method with bilinear finite elements on a regular Cartesian mesh

$$A = \nu K \otimes M + M \otimes ((\nu + \delta h)K + C)$$

where δ is the stabilization parameter, h is the mesh size

$$M = \frac{h}{6}\mathrm{tridiag}(1,4,1), \quad K = \frac{1}{h}\mathrm{tridiag}(-1,2,-1)$$

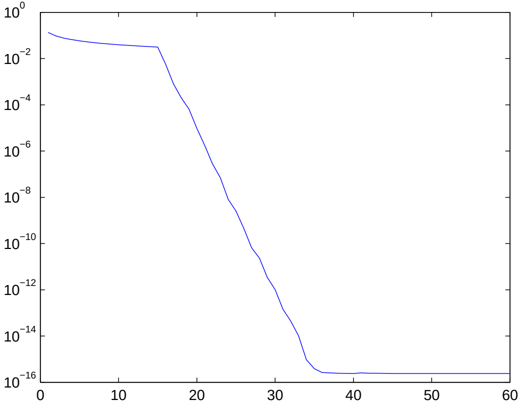
$$C = \frac{1}{2}\mathrm{tridiag}(-1,0,-1)$$

are tridiagonal matrices with constant diagonals

The matrix A is non normal

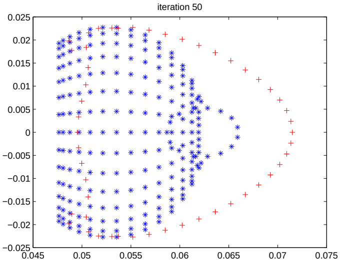
The right-hand side is example 2.1, page 1995 of Liesen and Strakoš (2005)

We use $h = 1/16$, $\nu = 0.01$ and $\delta = 0.34$. This gives a (small) linear system of order 225



Residual norms

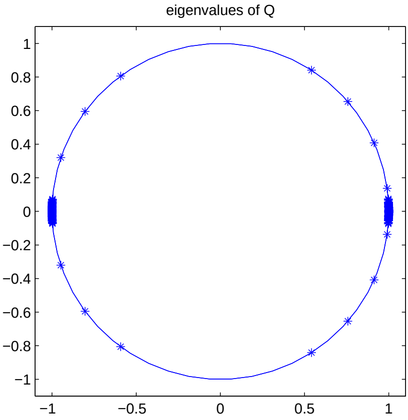
There is an initial near-stagnation for 15-16 iterations and then a good convergence



Eigenvalues of A (blue) and Ritz values (red) for $k = 50$ and $k = 100$

GMRES has converged long before any Ritz value has converged to an eigenvalue of A

As we know, the eigenvalues of A and the convergence of the Ritz values do not explain GMRES convergence for non-normal matrices



Eigenvalues of Q

We have 14 isolated eigenvalues and 2 tight clusters around 1 and -1

The isolated eigenvalues explain the initial near-stagnation

Conclusion

- When A is normal, GMRES convergence depends on the eigenvalues and eigenvectors of A

- When A is non-normal, GMRES convergence depends on the eigenvalues and eigenvectors of $Q = V^T W$

We need to study:

- Relations with the Schur parameters of Q

- Relations with Szegö polynomials

- Relations with Gauss quadrature on the unit circle