GMRES convergence for non-normal matrices

Gérard MEURANT

Joint work with J. Duintjer Tebbens
(using ideas from H. Sadok)

June 2013

We solve the linear system

$$Ax = b$$

with GMRES [Saad and Schultz (1986)]

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

The question we would like to address is:

what determines GMRES convergence?

This is a continuation of the 2012 SD talk which was dealing with normal matrices

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 the basis vectors are linearly independent. Then,

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

and H is (unreduced) upper Hessenberg

The GMRES iterates $x_k = V_k y_k$ are computed by solving

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

Let

$$K = \begin{pmatrix} b & Ab & \cdots & A^{n-1}b \end{pmatrix}$$

be the non singular Krylov matrix and

$$K = VU$$

with U upper triangular and non singular. Then

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

C being the companion matrix of A

[This is a consequence of $AK = KC$]

We have the following result for the GMRES residual norms [proved by many people: Stewart, Zitko, Ipsen, Liesen, Rozložník and Strakoš and Sadok]

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

where $M = U^*U = K^*K$ and M_{k+1} its principal submatrix of order $k+1$

We will use two simple tools:

- Cramer's rule (1750 but known before that)
- The Cauchy-Binet formula (1812) for $\det(AB)$ with A and B rectangular

Diagonalizable matrices

Let $A = X\Lambda X^{-1}$ and $c = X^{-1}b$. Then

$$K = X\begin{pmatrix} c & \Lambda c & \cdots & \Lambda^{n-1}c \end{pmatrix}$$

Therefore

$$M = \begin{pmatrix} c & \Lambda c & \cdots & \Lambda^{n-1}c \end{pmatrix}^* X^* X \begin{pmatrix} c & \Lambda c & \cdots & \Lambda^{n-1}c \end{pmatrix}$$

and

$$M_{k+1} = \mathcal{V}_{k+1}^* D_{\bar{c}} X^* X D_c \mathcal{V}_{k+1}$$

with D_c diagonal with c_i as diagonal entries and ...

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

an $n \times (k+1)$ Vandermonde matrix

Using Cramer's rule, we have

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

Let $F = XD_c\mathcal{V}_{k+1}$, an $n \times (k+1)$ matrix. Then $M_{k+1} = F^*F$ and by the Cauchy-Binet formula

$$\det(M_{k+1}) = \sum_{I_{k+1}} |\det(F_{I_{k+1},:})|^2$$

where I_{k+1} is a set of $k+1$ row indices $(i_1, \quad i_2, \quad \ldots, i_{k+1})$ such that $1 \leq i_1 < \cdots < i_{k+1} \leq n$ and $F_{I_{k+1},:}$ is the submatrix of F whose row indices belong to I_{k+1}

$$F_{I_{k+1},:} = (XD_c)_{I_{k+1},:}\mathcal{V}_{k+1}, \quad [(k+1) \times n] * [n \times (k+1)]$$

Hence we can again apply the Cauchy-Binet formula (this is what is different from the normal case - paper by Duintjer Tebbens, GM, Sadok and Strakoš, submitted)

$$\det(F_{I_{k+1},:}) = \sum_{J_{k+1}} \det(X_{I_{k+1},J_{k+1}}) c_{j_1} \cdots c_{j_{k+1}} \det(\mathcal{V}(\lambda_{j_1}, \dots, \lambda_{j_{k+1}}))$$

with

$$\mathcal{V}(\lambda_{j_1}, \dots, \lambda_{j_{k+1}}) = \begin{pmatrix} 1 & \lambda_{j_1} & \cdots & \lambda_{j_1}^k \\ 1 & \lambda_{j_2} & \cdots & \lambda_{j_2}^k \\ \vdots & \vdots & & \vdots \\ 1 & \lambda_{j_{k+1}} & \cdots & \lambda_{j_{k+1}}^k \end{pmatrix}$$

Similarly

$$M_{2:k+1,2:k+1} = \mathcal{V}_k^* \overline{\Lambda} D_{\bar{c}} X^* X D_c \Lambda \mathcal{V}_k$$

Hence we can apply the same technique for computing $\det(M_{2:k+1,2:k+1})$

The norm of the residual

Then $\|r_k\|^2 = N/D$ with

$$N = \sum_{I_{k+1}} \left| \sum_{J_{k+1}} \det(X_{I_{k+1},J_{k+1}}) c_{j_1} \cdots c_{j_{k+1}} \prod_{j_1 \le j_l < j_p \le j_{k+1}} (\lambda_{j_p} - \lambda_{j_l}) \right|^2$$

and

$$D = \sum_{I_k} \left| \sum_{J_k} \det(X_{I_k,J_k}) c_{j_1} \cdots c_{j_k} \lambda_{j_1} \cdots \lambda_{j_k} \prod_{j_1 \le j_l < j_p \le j_k} (\lambda_{j_p} - \lambda_{j_l}) \right|^2$$

where the summation is over all such possible ordered sets of indices

For $k = 1$ the denominator is a bit different
$D = \sum_{i=1}^n \left| \sum_{j=1}^n X_{i,j} c_j \lambda_j \right|^2$

Bounds for the residual norms

Using a result by Bellalij, Jbilou and Sadok we can prove that

$$\|r_k\|^2 \leq \frac{\|X\|^2}{e_1^T(\mathcal{V}_{k+1}^* D_{\bar{c}} D_c \mathcal{V}_{k+1}^*)^{-1} e_1}$$

It yields

$$\|r_k\|^2 \leq \|X\|^2 \frac{\sum_{1 \leq i_1 < \cdots < i_{k+1} \leq n} \prod_{j=1}^{k+1} \omega_{i_j} | \prod_{i_1 \leq i_\ell < i_j \leq i_{k+1}} (\lambda_{i_j} - \lambda_{i_\ell})|^2}{\sum_{1 \leq i_1 < \cdots < i_k \leq n} \prod_{j=1}^{k} \omega_{i_j} |\lambda_{i_j}| \, | \prod_{i_1 \leq i_\ell < i_j \leq i_k} (\lambda_{i_j} - \lambda_{i_\ell})|^2}$$

where $\omega_j = |c_j|^2$ with $c = X^{-1} b$

We can also obtain a similar lower bound where the multiplying factor is $\sigma_{min}(X)^2$

Hence, the residual norms depend on the (differences of the) eigenvalues, the eigenvectors and the right-hand side (through $X^{-1}b$)

Can we extend these results to non-diagonalizable matrices using the Jordan canonical form?

Let us consider a small example

An example with two Jordan blocks

Let $A = SJS^{-1}$ of order 5, $c = S^{-1}b$ and J defined as

$$J = \begin{pmatrix} \lambda & 1 & & & \\ & \lambda & 1 & & \\ & & \lambda & & \\ & & & \mu & 1 \\ & & & & \mu \end{pmatrix}$$

$\lambda, \mu \neq 0$
Then

$$M = K^*K \quad \textit{and} \;\; K = K(A, b) = SK(J, c)$$

We don't have a diagonal matrix any longer. Let us compute $K(J, c)$

$$K(J,c) = \begin{pmatrix} c_1 & \lambda c_1 + c_2 & \lambda^2 c_1 + 2\lambda c_2 + c_3 & \lambda^3 c_1 + 3\lambda^2 c_2 + 3\lambda c_3 & \lambda^4 c_1 + 4\lambda^3 c_2 + 6\lambda^2 c_3 \\ c_2 & \lambda c_2 + c_3 & \lambda^2 c_2 + 2\lambda c_3 & \lambda^3 c_2 + 3\lambda^2 c_3 & \lambda^4 c_2 + 4\lambda^3 c_3 \\ c_3 & \lambda c_3 & \lambda^2 c_3 & \lambda^3 c_3 & \lambda^4 c_3 \\ c_4 & \mu c_4 + c_5 & \mu^2 c_4 + 2\mu c_5 & \mu^3 c_4 + 3\mu^2 c_5 & \mu^4 c_4 + 4\mu^3 c_5 \\ c_5 & \mu c_5 & \mu^2 c_5 & \mu^3 c_5 & \mu^4 c_5 \end{pmatrix}$$

We can separate the coefficients and the eigenvalues

$$K(J,c) = CT = \begin{pmatrix} c_1 & c_2 & c_3 & 0 & 0 \\ c_2 & c_3 & 0 & 0 & 0 \\ c_3 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & c_4 & c_5 \\ 0 & 0 & 0 & c_5 & 0 \end{pmatrix} \begin{pmatrix} 1 & \lambda & \lambda^2 & \lambda^3 & \lambda^4 \\ 0 & 1 & 2\lambda & 3\lambda^2 & 4\lambda^3 \\ 0 & 0 & 1 & 3\lambda & 6\lambda^2 \\ 1 & \mu & \mu^2 & \mu^3 & \mu^4 \\ 0 & 1 & 2\mu & 3\mu^2 & 4\mu^3 \end{pmatrix}$$

Let the Cholesky factorization of S^*S be $S^*S = \mathcal{R}^*\mathcal{R}$ where $\mathcal{R}$ is upper triangular and let $Y = \mathcal{R}C$. Then

$$M = K^*K = K(J,c)^*S^*SK(J,c) = T^*C^*\mathcal{R}^*\mathcal{R}CT = T^*Y^*YT$$

The matrix Y has the following structure,

$$Y = \begin{pmatrix} x & x & x & x & x \\ x & x & 0 & x & x \\ x & 0 & 0 & x & x \\ 0 & 0 & 0 & x & x \\ 0 & 0 & 0 & x & 0 \end{pmatrix}$$

Then, using the same techniques as before, the GMRES residual norms are given by

$$\|r_k\|^2 = \frac{\sum_{I_{k+1}} \left| \sum_{J_{k+1}} \det(Y_{I_{k+1},J_{k+1}}) \det(T_{J_{k+1},[1:k+1]}) \right|^2}{\sum_{I_k} \left| \sum_{J_k} \det(Y_{I_k,J_k}) \det(T_{J_k,[2:k+1]}) \right|^2}$$

The main (open) problem is to compute the determinants of submatrices of T

This can be done for small examples

For our previous example and for $k < 3$, we have determinants of submatrices of T equal to 1

But, for $k = 3$ we have

Determinants of $T_{J_4,[1:4]}$

Indices in J_4	value
{1,2,3,4}	$(\mu - \lambda)^3$
{1,2,3,5}	$3(\mu - \lambda)^2$
{1,2,4,5}	$(\mu - \lambda)^4$
{1,3,4,5}	$-2(\mu - \lambda)^3$
{2,3,4,5}	$3(\mu - \lambda)^2$

We remark that all the terms in $\det(M_4)$ have $\mu - \lambda$ as a factor

This was not the case for $k = 2$

For the general case the block of n_i rows of T corresponding to the eigenvalue λ_i is

$$\begin{pmatrix} 1 & \lambda_i & \lambda_i^2 & \cdots & \lambda_i^{n_i-1} & \cdots & \lambda_i^{n-1} \\ 0 & 1 & 2\lambda_i & \cdots & \binom{n_i-1}{1}\lambda_i^{n_i-2} & \cdots & \binom{n-1}{1}\lambda_i^{n-2} \\ 0 & 0 & 1 & \cdots & \binom{n_i-1}{2}\lambda_i^{n_i-3} & \cdots & \binom{n-2}{2}\lambda_i^{n-3} \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 1 & \cdots & \binom{n-1}{n_i-1}\lambda_i^{n-n_i} \end{pmatrix}$$

It seems difficult to compute analytically all the determinants we need. But some results can be obtained. For instance,

Proposition

If $k < \max_i(n_i)$ *there are determinants of submatrices in* $T_{[1:k+1],[1:k+1]}$ *that are equal to* 1

Scaled Jordan blocks

See Ipsen (1998), Tichý, Liesen and Faber (2007)

This is a case that can be handled completely using the same machinery as before

$$A = \begin{pmatrix} \lambda & \eta & & & \\ & \lambda & \eta & & \\ & & \ddots & \ddots & \\ & & & \lambda & \eta \\ & & & & \lambda \end{pmatrix}$$

Then $A = SJS^{-1}$

$$J = \begin{pmatrix} \lambda & 1 & & & \\ & \lambda & 1 & & \\ & & \ddots & \ddots & \\ & & & \lambda & 1 \\ & & & & \lambda \end{pmatrix}, \quad S = \begin{pmatrix} 1 & & & & \\ & 1/\eta & & & \\ & & 1/\eta^2 & & \\ & & & \ddots & \\ & & & & 1/\eta^{n-1} \end{pmatrix}$$

Theorem

Let b be an n-vector of unit norm and B the Hankel matrix of order n defined by $(b_1, b_2, \dots, b_n, 0, \dots, 0)$. The GMRES residual norms when solving $Ax = b$, with $x_0 = 0$ and A a scaled Jordan block of order n defined by λ and η, satisfy

$$\|r_k\|^2 = \frac{\sum_{I_{k+1}} |\det(B_{I_{k+1},[1:k+1]})|^2}{D_k}$$

with

$$\begin{aligned} D_k &= \sum_{I_k} |(\lambda/\eta)^k \det(B_{I_k,\mathcal{I}_1^k}) + (\lambda/\eta)^{k-1} \det(B_{I_k,\mathcal{I}_2^k}) + \cdots \\ &+ (\lambda/\eta) \det(B_{I_k,\mathcal{I}_k^k}) + \det(B_{I_k,\mathcal{I}_{k+1}^k})|^2 \end{aligned}$$

where $\mathcal{I}_j^k$, $j = 1, \dots, k+1$ are the sets of indices with k elements in the ordered combinations of $k+1$ elements enumerated in lexicographic ordering

Generalization to other Krylov methods

Most of the results are still valid for a QMR-like method using any ascending basis of the Krylov space

$$K = VU$$

V being any non singular matrix whose columns span the Krylov space with U upper triangular. Then we *define*

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

with C the companion matrix of $A \Rightarrow AV = VH$

$x_k = V_k y$ with y minimizing the norm of the quasi-residual z_k

$$\|e_1 - \underline{H}_k y\|$$

for $x_0 = 0, \|b\| = 1$

Examples: QMR (Freund and Nachtigal), CMRH (Sadok)

In QMR one computes a bi-orthogonal basis (with $K(A^T, \tilde{b})$)

In CMRH one computes an Hessenberg basis $V = P^T L$ with P a permutation matrix and L lower triangular (based on LU factorization of K)

We can obtain an expression for the norm of the quasi-residual for diagonalizable matrices using $U = V^{-1} K$

Let $M = U^*U \neq K^*K$ and M_{k+1} a principal matrix of M. Then

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

and $\|z_k\|^2 = N/D$ with

$$N = \sum_{I_{k+1}} \left| \sum_{J_{k+1}} \det(Z_{I_{k+1},J_{k+1}}) c_{j_1} \cdots c_{j_{k+1}} \prod_{j_1 \le j_l < j_p \le j_{k+1}} (\lambda_{j_p} - \lambda_{j_l}) \right|^2,$$

$$D = \sum_{I_k} \left| \sum_{J_k} \det(Z_{I_k,J_k}) c_{j_1} \cdots c_{j_k} \lambda_{j_1} \cdots \lambda_{j_k} \prod_{j_1 \le j_l < j_p \le j_k} (\lambda_{j_p} - \lambda_{j_l}) \right|^2,$$

with $Z = V^{-1}X$, X eigenvector matrix

Conclusion

We have shown how the GMRES residual norms depend on the eigenvalues, eigenvectors and the right-hand side for diagonalizable matrices

The case of the Jordan canonical form is more complicated with only partial results

Most of the results can be extended to quasi-minimal residual methods