On the convergence of Q-OR and Q-MR Krylov methods for solving linear systems

Gérard MEURANT

Joint work with J. Duintjer Tebbens

November 2013

Many Krylov methods have been proposed over the years for solving linear systems

Most of them can be classified as quasi-orthogonal (Q-OR) or quasi-minimum residual (Q-MR)

Q-OR: FOM, BiCG, Hessenberg, ...

Q-MR: GMRES, truncated GMRES, QMR, CMRH, ...

Whatever their definition, these methods share many fundamental properties

See M. Eiermann and O.G. Ernst, *Geometric aspects in the theory of Krylov subspace methods*, Acta Numerica, v 10 n 10 (2001), pp. 251–312

They differ by the basis of the Krylov space that is constructed:

- orthogonal for FOM/GMRES,
- bi-orthogonal for BiCG/QMR,
- based on an LU factorization for Hessenberg/CMRH

What do we know about GMRES?

Let

$$K = \begin{pmatrix} b & Ab & A^2b & \cdots & A^{n-1}b \end{pmatrix}$$

be the Krylov matrix that we assume of full rank. Then

$$K = VU$$

with V orthogonal (or unitary) and U upper triangular with positive real diagonal entries

The matrix $H = V^*AV$ is upper Hessenberg

We have

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

where C is the companion matrix for the eigenvalues of A

Let x_k^G (resp. x_k^F) be the iterates for GMRES (resp. FOM) and the residual vectors $r_k^G = b - Ax_k^G$, $r_k^F = b - Ax_k^F$

We assume $x_0 = 0$ and $\|b\| = 1$

We know that

- every residual norm convergence curve is possible for GMRES (and FOM)

- $|(U^{-1})_{1,k}| = 1/\|r_{k-1}^F\|$

- $\|r_k^G\|^2 = 1/(M_{k+1}^{-1})_{1,1}$ with $M_{k+1} = U_{k+1}^* U_{k+1}$

- one can construct matrices A with a prescribed spectrum and right-hand sides b such that GMRES yields a prescribed decreasing residual norm convergence curve

- we have two parametrizations of this class of matrices

Moreover we can express the GMRES residual norms as functions of the eigenvalues and eigenvectors of A

Let A be a diagonalizable matrix with $A = X\Lambda X^{-1}$. Then

$$\|r_k^M\|^2 = \sigma_{k+1}^N / \sigma_k^D$$

with

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

$$\sigma_1^D = \sum_{i=1}^n \left| \sum_{j=1}^n X_{i,j} c_j \lambda_j \right|^2$$

and

$$\sigma_k^D = \sum_{I_k} \left| \sum_{J_k} \det(X_{I_k,J_k}) c_{j_1} \cdots c_{j_k} \lambda_{j_1} \cdots \lambda_{j_k} \prod_{j_1 \le j_l < j_p \le j_k} (\lambda_{j_p} - \lambda_{j_l}) \right|^2, \quad k > 1$$

where the summations are over all sets of indices $I_{k+1}, J_{k+1}, I_k, J_k$ defined as I_ℓ to be a set of ℓ indices $(i_1, i_2, \ldots, i_\ell)$ such that $1 \le i_1 < \cdots < i_\ell \le n$, X_{I_ℓ,J_ℓ} is the submatrix of X whose row and column indices are defined by I_ℓ and J_ℓ and $c = X^{-1}b$

If the matrix A is normal we have simpler formulas

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

$$\sigma_1^D = \sum_{i=1}^{n} |c_j|^2 \, |\lambda_j|^2$$

and

$$\sigma_k^D = \sum_{I_k} |c_{j_1}|^2 \cdots |c_{j_k}|^2 \, |\lambda_{j_1}|^2 \cdots |\lambda_{j_k}|^2 \prod_{j_1 \le j_l < j_p \le j_k} \left|(\lambda_{j_p} - \lambda_{j_l})\right|^2, \; k > 1$$

For all these properties see

A. Greenbaum and Z. Strakoš, *Matrices that generate the same Krylov residual spaces*, in Recent advances in iterative methods, G.H. Golub, A. Greenbaum and M. Luskin, eds., Springer, (1994), pp. 95–118

A. Greenbaum, V. Pták and Z. Strakoš, *Any nonincreasing convergence curve is possible for GMRES*, SIAM J. Matrix Anal. Appl., v 17 (1996), pp. 465–469

M. Arioli, V. Pták and Z. Strakoš, *Krylov sequences of maximal length and convergence of GMRES*, BIT Numerical Mathematics, v 38 n 4 (1998), pp. 636–643

J. Duintjer Tebbens and G. Meurant, *Any Ritz value behavior is possible for Arnoldi and for GMRES*, SIAM J. Matrix Anal. Appl., v 33 n 3 (2012), pp. 958–978

J. Duintjer Tebbens and G. Meurant, *GMRES convergence and the Jordan canonical form*, in preparation, (2013)

Q-OR and Q-MR methods

Our aim is to see if some of these properties can be extended to Q-OR and Q-MR methods

We assume that we have a basis V of the Krylov space (with columns of unit norm) such that $K = VU$ with V nonsingular and U upper triangular

We define $H = UCU^{-1}$. As a consequence $AV = VH$. The iterates are

$$x_k = V_k y^{(k)}$$

where V_k is the matrix of the k first columns of V. The residual r_k is

$$V_k e_1 - AV_k y^{(k)} = V_k(e_1 - H_k y^{(k)}) - h_{k+1,k} y_k^{(k)} v_{k+1} = V_{k+1}(e_1 - \underline{H}_k y^{(k)})$$

The Q-OR method is defined (provided that H_k is nonsingular) by

$$H_k y^{(k)} = e_1$$

This annihilates the first term in the residual

In the Q-MR method $y^{(k)}$ is computed as the solution of the least squares problem

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

where $\underline{H}_k$ is $(k+1) \times k$. The vector $z_k^M = e_1 - \underline{H}_k y^{(k)}$ is referred as the quasi-residual. The residual vector is $r_k^M = V_{k+1} z_k^M$

Generally, the two problems are solved using Givens rotations with sines s_j. It is known that

$$\|z_k^M\| = |s_1 s_2 \cdots s_k|$$

Moreover we have a relation between the Q-OR residual norms and the Q-MR quasi-residual norms

$$\frac{1}{\|r_k^O\|^2} = \frac{1}{\|z_k^M\|^2} - \frac{1}{\|z_{k-1}^M\|^2}$$

Properties of Q-OR and Q-MR methods

From these results we can show by induction that

$$|(U^{-1})_{1,k}| = \frac{1}{\|r_{k-1}^O\|}$$

A consequence of this result is the following

Let $M_{k+1} = U_{k+1}^* U_{k+1}$. Then

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

The difference with GMRES is that we only have the norm of the quasi-residual

Let A be a diagonalizable matrix with $A = X\Lambda X^{-1}$ and $Z = V^{-1}X$. Then

$$\|z_k^M\|^2 = \sigma_{k+1}^N / \sigma_k^D$$

with

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

$$\sigma_1^D = \sum_{i=1}^{n} \left| \sum_{j=1}^{n} Z_{i,j} c_j \lambda_j \right|^2$$

and

$$\sigma_k^D = \sum_{I_k} \left| \sum_{J_k} \det(Z_{I_k,J_k}) c_{j_1} \cdots c_{j_k} \lambda_{j_1} \cdots \lambda_{j_k} \prod_{j_1 \le j_l < j_p \le j_k} (\lambda_{j_p} - \lambda_{j_l}) \right|^2 , \quad k > 1$$

where the summations are over all sets of indices $I_{k+1}, J_{k+1}, I_k, J_k$ defined as I_ℓ to be a set of ℓ indices $(i_1, i_2, \ldots, i_\ell)$ such that $1 \le i_1 < \cdots < i_\ell \le n$, Z_{I_ℓ,J_ℓ} is the submatrix of Z whose row and column indices are defined by I_ℓ and J_ℓ and $c = X^{-1}b$

This result arises from $\|z_k^M\|^2 = 1/(M_{k+1}^{-1})_{1,1}$ and

$$\begin{aligned} M &= U^*U = K^*V^{-*}V^{-1}K \\ &= \begin{pmatrix} c & \Lambda c & \cdots & \Lambda^{n-1}c \end{pmatrix}^* Z^*Z \begin{pmatrix} c & \Lambda c & \cdots & \Lambda^{n-1}c \end{pmatrix} \end{aligned}$$

It yields

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

where D_c is diagonal and $\mathcal{V}_{k+1}$ is an $n \times (k+1)$ Vandermonde matrix

We compute the $(1,1)$ entry of the inverse using Cramer's rule and the Cauchy-Binet determinant formula

Construction of linear systems with a prescribed convergence curve

Can we construct linear systems with a prescribed convergence curve and a prescribed spectrum for Q-OR and Q-MR methods?

For FOM/GMRES this is easy since we just have to construct an upper triangular matrix U^{-1} with the inverses of the FOM residual norms on the first row. Then we take

$$A = VUCU^{-1}V^*, \qquad b = Ve_1$$

where C is the companion matrix of the given eigenvalues and V is any unitary matrix

Things are more difficult for some Q-OR/Q-MR methods

BiCG

We would like to find a matrix

$$H = \begin{pmatrix} \gamma_1 & \beta_2 & 0 & 0 & 0 \\ \rho_2 & \gamma_2 & \beta_3 & 0 & 0 \\ 0 & \ddots & \ddots & \ddots & 0 \\ 0 & 0 & \rho_{n-1} & \gamma_{n-1} & \beta_n \\ 0 & 0 & 0 & \rho_n & \gamma_n \end{pmatrix} = UCU^{-1}$$

such that the first row of U^{-1} is prescribed as $\begin{pmatrix} 1 & g_1 & \cdots & g_{n-1} \end{pmatrix}$ with $g_j \neq 0$

Let $\omega_2, \ldots, \omega_n$ be arbitrary chosen entries of the last column of U^{-1} with $\omega_n \neq 0$ and $\omega_1 = g_{n-1}$

The last column of $U^{-1}H = CU^{-1}$ yields

$$\begin{pmatrix} g_{n-2}\beta_n + g_{n-1}\gamma_n \\ \vdots \\ \omega_n\gamma_n \end{pmatrix} = \begin{pmatrix} -\alpha_0\omega_n \\ \vdots \\ \omega_{n-1} - \alpha_{n-1}\omega_n \end{pmatrix}$$

We use the first and last equations

$$\begin{pmatrix} g_{n-2} & g_{n-1} \\ 0 & \omega_n \end{pmatrix} \begin{pmatrix} \beta_n \\ \gamma_n \end{pmatrix} = \begin{pmatrix} -\alpha_0\omega_n \\ \omega_{n-1} - \alpha_{n-1}\omega_n \end{pmatrix}$$

The solution of this 2×2 non singular linear system yields γ_n, β_n
From the other equations that we discarded we can compute the unknown entries $\nu_{j,n-1}$ of column $n-1$ of U^{-1}

Then we go on with column $n-1$

We have three unknowns $\beta_{n-1}, \gamma_{n-1}$ and ρ_n
We first take the first and the last two equations
This gives us a linear system with an upper triangular system

$$\begin{pmatrix} g_{n-3} & g_{n-2} & g_{n-1} \\ 0 & \nu_{n-1,n-1} & \omega_{n-1} \\ 0 & 0 & \omega_n \end{pmatrix} \begin{pmatrix} \beta_{n-1} \\ \gamma_{n-1} \\ \rho_n \end{pmatrix} = \begin{pmatrix} 0 \\ \nu_{n-2,n-1} \\ \nu_{n-1,n-1} \end{pmatrix}$$

And so on...

So far we don't know how to completely handle the case with zero entries on the first row of U^{-1}

This algorithm can be extended to a larger upper bandwidth

Conclusion

Many properties of FOM/GMRES are extended to general Q-OR/Q-MR methods

We express the Q-MR quasi-residual norms as functions of the eigenvalues, the eigenvectors, the right-hand side and the basis of the Krylov space

We have a parametrization of the class of matrices with a prescribed spectrum and a prescribed Q-OR/Q-MR convergence curve

In particular we can construct examples with a BiCG (finite) convergence curve