Do Ritz values influence the behavior of restarted GMRES?

Jurjen Duintjer Tebbens

joint work with

Gérard Meurant

Institute of Computer Science, Academy of Sciences of the Czech Republic

June 27, Glasgow 25th Biennial Conference on Numerical Analysis 2013

Motivation

We consider the solution of linear systems

$$\mathbf{A}x = b$$

where $\mathbf{A} \in \mathbb{C}^{n \times n}$ is **non-normal and nonsingular**, by the Generalized Minimal Residual (GMRES) method [Saad & Schultz 1986].

As this is a Krylov subspace method based on long recurrences, we will focus on restarted GMRES; GMRES(m) will denote GMRES restarted after every mth iteration.

Without loss of generality, $\|b\|=1$, $x_0=0$.

The kth residual norm satisfies

Optimality property

$$||r_k|| = \min_{x \in \mathcal{K}_k(\mathbf{A}, b)} ||b - Ax||,$$

where the minimization is over all elements of the kth Krylov subspace,

$$\mathcal{K}_k(\mathbf{A}, b) \equiv \operatorname{span}\{b, \mathbf{A}b, \dots, \mathbf{A}^{k-1}b\}.$$

- Residual norms do not increase, but they can stagnate in GMRES(m)
- Residuals can be written as polynomials in A times b,

$$r_k = p(A)b$$
 with $||r_k|| = \min_{p \in \pi_k} ||p(A)b||$,

where π_k is the set of polynomials of degree k taking the value one in the origin.

Influence of spectral properties

Let the Jordan normal form of A be

$$A = XJX^{-1},$$

then the kth residual norm can be written as

$$||r_k|| = \min_{p \in \pi_k} ||Xp(J)X^{-1}b||.$$

This shows that the convergence of GMRES, measured by the residual norm, depends on

- ullet the eigenvalues contained in J
- \bullet the eigenvectors (or principal vectors with non-diagonalizable input matrices) contained in X
- components of the right-hand side in the eigenvector basis.

Limited influence of eigenvalues alone

The next classical result shows that convergence needs not depend on the eigenvalues alone:

Theorem 1 [Greenbaum & Pták & Strakoš 1996] Let

$$||b|| = f_0 \ge f_1 \ge f_2 \dots \ge f_{n-1} > 0$$

be any non-increasing sequence of real positive values and let

$$\lambda_1,\ldots,\lambda_n$$

be any set of nonzero complex numbers. Then there exists a class of matrices $A \in \mathbb{C}^{n \times n}$ and right-hand sides $b \in \mathbb{C}^n$ such that the residual vectors $r^{(k)}$ generated by GMRES method satisfy

$$\|r^{(k)}\| = f_k, \quad 0 \le k \le n, \quad \text{and} \quad \operatorname{spectrum}(A) = \{\lambda_1, \dots, \lambda_n\}.$$

Influence of Ritz values

We recently extended this result with the fact that GMRES convergence needs not be dependent on Ritz values either, except that a zero Ritz value implies stagnation:

Theorem 2 [DT & Meurant 2012] In addition to the assumptions of Theorem 1, let also n(n-1)/2 complex values

be given and assume that $f_{k-1} = f_k$ if and only if

$$0 \in \{\theta_1^{(k)}, \dots, \theta_k^{(k)}\}.$$

Influence of Ritz values

Then there exists a class of matrices $A\in\mathbb{C}^{n\times n}$ and right-hand sides $b\in\mathbb{C}^n$ such that the residual vectors $r^{(k)}$ generated by GMRES method satisfy

$$||r^{(k)}|| = f_k, \quad 0 \le k \le n,$$
 spectrum $(A) = \{\lambda_1, \dots, \lambda_n\},\$

and GMRES generates in the kth iteration (for all $k \leq n$) the Ritz values

$$\theta_1^{(k)}, \ldots, \theta_k^{(k)}.$$

Influence of Ritz values

Then there exists a class of matrices $A\in\mathbb{C}^{n\times n}$ and right-hand sides $b\in\mathbb{C}^n$ such that the residual vectors $r^{(k)}$ generated by GMRES method satisfy

$$||r^{(k)}|| = f_k, \quad 0 \le k \le n,$$
 spectrum $(A) = \{\lambda_1, \dots, \lambda_n\},$

and GMRES generates in the kth iteration (for all $k \leq n$) the Ritz values

$$\theta_1^{(k)}, \dots, \theta_k^{(k)}.$$

- Thus, in every iteration, we can prescribe the Ritz values and simultaneously the GMRES residual norm.
- This also shows that the Arnoldi method for eigenproblems can generate arbitrary Ritz values in all intermediate iterations.

Consequences for restarted GMRES?

- It seems possible to prescribe the harmonic Ritz values in the Arnoldi method as well [Meurant, personal communication].
- ullet Prescribing GMRES residual norms and harmonic Ritz values simultaneously is unlikely to be possible harmonic Ritz values are the roots of the GMRES polynomials $r_k=p(A)b$.

Consequences for restarted GMRES?

- It seems possible to prescribe the harmonic Ritz values in the Arnoldi method as well [Meurant, personal communication].
- ullet Prescribing GMRES residual norms and harmonic Ritz values simultaneously is unlikely to be possible harmonic Ritz values are the roots of the GMRES polynomials $r_k=p(A)b$.

The purpose of this talk is:

- To investigate whether residual norms, eigenvalues and Ritz values can be prescribed in restarted GMRES as well.
- To point out possible consequences for (analysis of) preconditioning and other popular acceleration strategies for GMRES(m).

Outline

- 1 Prescribing residual norms and Ritz values in GMRES(m)
- 2 Without stagnation at the end of cycles
- 3 Allowing stagnation at the end of cycles
- 4 Brief discussion
- Conclusions

Here is how one can prescribe Ritz values and residual norms in full GMRES [DT & Meurant 2013]:

Here is how one can prescribe Ritz values and residual norms in full GMRES [DT & Meurant 2013]:

ullet Choose a unitary matrix V and put $b=Ve_1$ and

$$A = VHV^*$$
, H upper Hessenberg.

Here is how one can prescribe Ritz values and residual norms in full GMRES [DT & Meurant 2013]:

ullet Choose a unitary matrix V and put $b=Ve_1$ and

$$A = VHV^*$$
, H upper Hessenberg.

ullet To force the desired eigenvalues, H will be of the form

$$H = U^{-1}CU$$
, U nonsingular upper triangular,

where ${\cal C}$ is the companion matrix for the prescribed spectrum.

Here is how one can prescribe Ritz values and residual norms in full GMRES [DT & Meurant 2013]:

• Choose a unitary matrix V and put $b = Ve_1$ and

$$A = VHV^*$$
, H upper Hessenberg.

ullet To force the desired eigenvalues, H will be of the form

$$H = U^{-1}CU$$
, U nonsingular upper triangular,

where C is the companion matrix for the prescribed spectrum.

ullet To force the desired residual norms, the first row g^T of U has entries

$$g_1 = \frac{1}{f(0)}, \qquad g_k = \frac{\sqrt{f(k-2)^2 - f(k-1)^2}}{f(k-2)f(k-1)}, \qquad k = 2, \dots, n.$$

Let

$$A = V(U^{-1}CU)V^*, \qquad b = Ve_1.$$

Let

$$A = V(U^{-1}CU)V^*, \qquad b = Ve_1.$$

ullet To force the desired Ritz values, the remaining submatrix T of

$$U = \left[\begin{array}{cc} g^T \\ 0 & T \end{array} \right]$$

has entries satisfying

$$\prod_{i=1}^{k} (\lambda - \rho_i^{(k)}) = g_{k+1} + \sum_{i=1}^{k} t_{i,k} \lambda^i.$$

Let

$$A = V(U^{-1}CU)V^*, \qquad b = Ve_1.$$

ullet To force the desired Ritz values, the remaining submatrix T of

$$U = \left[\begin{array}{cc} g^T \\ 0 & T \end{array} \right]$$

has entries satisfying

$$\prod_{i=1}^{k} (\lambda - \rho_i^{(k)}) = g_{k+1} + \sum_{i=1}^{k} t_{i,k} \lambda^i.$$

Is prescribing these values possible in restarted GMRES?

Prescribing residual norms in restarted GMRES was considered in the paper [Vecharinsky & Langou 2011]. It assumes a rather special situation in $\mathsf{GMRES}(\mathsf{m})$:

- During every restart cycle, all residual norms stagnate except for the very last iteration inside the cycle.
- In this very last iteration it is assumed that the residual norm is strictly decreasing.

Prescribing residual norms in restarted GMRES was considered in the paper [Vecharinsky & Langou 2011]. It assumes a rather special situation in GMRES(m):

- During every restart cycle, all residual norms stagnate except for the very last iteration inside the cycle.
- In this very last iteration it is assumed that the residual norm is strictly decreasing.

Theorem 3 [Vecharinsky & Langou 2011]. Let n complex nonzero numbers $\lambda_1,\ldots,\lambda_n$ and k positive decreasing numbers

$$f(0) > f(1) > \dots > f(k-1) > 0,$$

be given. With the assumptions 1. and 2. above, let the very last residual at the end of the jth cycle be denoted by \bar{r}_j . If km < n, then:

ullet There exists a matrix A of order n with a right hand side such that GMRES(m) generates residual norms at the end of cycles satisfying

$$\|\bar{r}_j\| = f(j), \qquad j = 0, 1, \dots, k.$$

• The matrix A has the eigenvalues $\lambda_1, \ldots, \lambda_n$.

• There exists a matrix A of order n with a right hand side such that GMRES(m) generates residual norms at the end of cycles satisfying

$$\|\bar{r}_i\| = f(j), \qquad j = 0, 1, \dots, k.$$

• The matrix A has the eigenvalues $\lambda_1, \ldots, \lambda_n$.

In fact, to prescribe all residual norms and all Ritz values in GMRES(m), it suffices that $(m+1) \times m$ Hessenberg matrices of the individual restart cycles have the form described before, i.e. that the kth Hessenberg matrix is

$$\hat{H}_{m}^{(k)} = \begin{bmatrix} g_{1}^{(k)} & \dots & g_{m+1}^{(k)} \\ 0 & T_{m}^{(k)} \end{bmatrix}^{-1} \begin{bmatrix} 0 \\ I_{m} \end{bmatrix} \begin{bmatrix} g_{1}^{(k)} & \dots & g_{m}^{(k)} \\ 0 & T_{m-1}^{(k)} \end{bmatrix},$$

where $g^{(k)}$ determines the convergence curve and the columns of T_{m-1} determine the Ritz values.

Outline

- Prescribing residual norms and Ritz values in GMRES(m)
- Without stagnation at the end of cycles
- 3 Allowing stagnation at the end of cycles
- Brief discussion
- Conclusions

First, we assume restart cycles do not stagnate in their last iteration.

Theorem 5 [DT & Meurant 2013?] Let

$$\hat{H}_m^{(1)}, \dots, \hat{H}_m^{(k)} \in \mathbb{C}^{(m+1) \times m}$$

be k unreduced upper Hessenberg matrices with positive subdiagonal and let km < n. If $A \in \mathbb{C}^{n \times n}$ is a matrix and $b \in \mathbb{C}^n$ a nonzero vector, the following assertions are equivalent:

- 1. The kth cycle of GMRES(m) applied to A and b does not stagnate in its last iteration and generates the Hessenberg matrix $\hat{H}_m^{(k)}$.
- 2. The matrix A and the vector b have the form

$$A = VHV^*, \qquad b = Ve_1,$$

where V is unitary, H is upper Hessenberg and the columns (k-1)m+1 till km corresponding to the kth cycle are of the form:

$$H\left[e_{(k-1)m+1},\dots,e_{km}\right] = \begin{bmatrix} & (\prod_{i=2}^{k-1}\zeta_1^{(i)})z^{(1)}e_1^T\hat{H}_m^{(k)} \\ & \vdots \\ & \zeta_1^{(k-1)}z^{(k-2)}e_1^T\hat{H}_m^{(k)} \\ & \hat{h}^{(k)} & z^{(k-1)}e_1^T\hat{H}_m^{(k)}\begin{bmatrix} 0 \\ I_{m-1} \end{bmatrix} \\ & 0 & \begin{bmatrix} 0 & I_m \end{bmatrix}\hat{H}_m^{(k)}\begin{bmatrix} 0 \\ I_{m-1} \end{bmatrix} \end{bmatrix}, \quad \textit{where}$$

$$\begin{split} z^{(i)} &= \left(I_{m+1} - \hat{H}_m^{(i)}(\hat{H}_m^{(i)})^\dagger\right) e_1 / \left\| \left(I_{m+1} - \hat{H}_m^{(i)}(\hat{H}_m^{(i)})^\dagger\right) e_1 \right\|, \quad 1 \leq i \leq k-1, \\ \hat{h}^{(k)} &= [\hat{h}_1^{(k)}, \dots, \hat{h}_{m+1}^{(k)}]^T = \frac{1}{\zeta_{m+1}^{(k-1)}} \left(h_{1,1}^{(k)} z^{(k-1)} - \hat{H}_m^{(k-1)}[\zeta_1^{(k-1)}, \dots, \zeta_m^{(k-1)}]^T\right) \end{split}$$
 and

$$\hat{h}_{m+2}^{(k)} = \frac{h_{2,1}^{(k)}}{\zeta_{m+1}^{(k-1)}}.$$

Thus we know how to generate, by the right choice of columns of H, arbitrary Hessenberg matrices during all restarts. Therefore we may prescribe not only GMRES residual norms inside cycles and Ritz values but also other values (singular values, harmonic Ritz values ...).

Remark: Prescribing k restarts under the condition km < n means that in the parametrization

$$A = VHV^*, \qquad b = ||b||Ve_1,$$

we put conditions on the first km < n columns of H only. The last column can be chosen arbitrarily. It can be checked (see, e.g., [Parlett & Strang 2008], that any nonzero spectrum of A is possible with an appropriate choice of the last column.

Outline

- Prescribing residual norms and Ritz values in GMRES(m)
- 2 Without stagnation at the end of cycles
- 3 Allowing stagnation at the end of cycles
- 4 Brief discussion
- Conclusions

Let m iterations of the initial cycle give the Arnoldi decomposition

$$AV_m^{(1)} = V_{m+1}^{(1)} \hat{H}_m^{(1)}, \quad V_{m+1}^{(1)^*} V_{m+1}^{(1)} = I_{m+1}.$$

Let m iterations of the initial cycle give the Arnoldi decomposition

$$AV_m^{(1)} = V_{m+1}^{(1)} \hat{H}_m^{(1)}, \quad V_{m+1}^{(1)^*} V_{m+1}^{(1)} = I_{m+1}.$$

The m iterations of the second cycle give the Arnoldi decomposition

$$AV_m^{(2)} = V_{m+1}^{(2)} \hat{H}_m^{(2)}, \quad V_{m+1}^{(2)^*} V_{m+1}^{(2)} = I_{m+1},$$

where if $r_m^{(1)}$ is the residual vector at the end of the first cycle,

$$V_{m+1}^{(2)}e_1 = \frac{r_m^{(1)}}{\|r_m^{(1)}\|} \equiv V_{m+1}^{(1)}z^{(1)}.$$

Let \boldsymbol{m} iterations of the initial cycle give the Arnoldi decomposition

$$AV_m^{(1)} = V_{m+1}^{(1)} \hat{H}_m^{(1)}, \quad V_{m+1}^{(1)^*} V_{m+1}^{(1)} = I_{m+1}.$$

The m iterations of the second cycle give the Arnoldi decomposition

$$AV_m^{(2)} = V_{m+1}^{(2)} \hat{H}_m^{(2)}, \quad V_{m+1}^{(2)^*} V_{m+1}^{(2)} = I_{m+1},$$

where if $r_m^{(1)}$ is the residual vector at the end of the first cycle,

$$V_{m+1}^{(2)}e_1 = \frac{r_m^{(1)}}{\|r_m^{(1)}\|} \equiv V_{m+1}^{(1)}z^{(1)}.$$

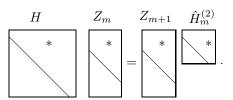
How do we construct the columns of H ? We know that the columns $1,\ldots,m$ of H are

$$H\left[\begin{array}{c}I_m\\0\end{array}\right] = \left[\begin{array}{c}\hat{H}_m^{(1)}\\0\end{array}\right].$$

Lemma 1. The matrix $\hat{H}_m^{(2)}$ is the Hessenberg matrix generated by m iterations of Arnoldi with input matrix H and initial vector $\begin{bmatrix} z^{(1)T} & 0 \end{bmatrix}^T$, i.e.

$$HZ_m = Z_{m+1}\hat{H}_m^{(2)}, \quad Z_{m+1}e_1 = \begin{bmatrix} z^{(1)} \\ 0 \end{bmatrix}, \quad Z_{m+1}^*Z_{m+1} = I_{m+1}.$$
 (1)

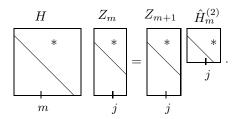
Can we construct the columns $m+1, m+2, \ldots, 2m$ of H such that (1) is satisfied with a prescribed Hessenberg matrix $\hat{H}_m^{(2)}$? This will depend on the number of non-zeroes in $\begin{bmatrix} z^{(1)T} & 0 \end{bmatrix}^T$ because



Lemma 2. Let $r_m^{(1)} = V_{m+1}^{(1)} z^{(1)}$. Then for an integer j the last j-1 entries of $z^{(1)}$ are zero if and only if the last j residual norms are equal, i.e.

$$||r_0^{(1)}|| \ge ||r_1^{(1)}|| \ge \dots \ge ||r_{m-j}^{(1)}|| > ||r_{m-j+1}^{(1)}|| = \dots = ||r_m^{(1)}||.$$

Then the Arnoldi decomposition $HZ_m=Z_{m+1}\hat{H}_m^{(2)}$ looks like



Therefore, with j-1 stagnation steps at the end of the first restart cycle:

- the first j-1 columns of the Hessenberg matrix of the second cycle $\hat{H}_m^{(2)}$ are fully determined by $\hat{H}_m^{(1)}$ and $z^{(1)}$ they cannot be prescribed.
- We can also prove that the first row of $\hat{H}_m^{(2)}$ is zero on its first j-1 positions, i.e. they correspond to iterations with stagnation!

Therefore, with j-1 stagnation steps at the end of the first restart cycle:

- the first j-1 columns of the Hessenberg matrix of the second cycle $\hat{H}_m^{(2)}$ are fully determined by $\hat{H}_m^{(1)}$ and $z^{(1)}$ they cannot be prescribed.
- We can also prove that the first row of $\hat{H}_m^{(2)}$ is zero on its first j-1 positions, i.e. they correspond to iterations with stagnation!

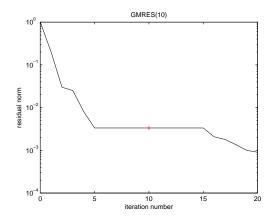
Corollary If the last j-1 residual norms stagnate in the initial cycle, i.e.

$$||r_0^{(1)}|| \ge ||r_1^{(1)}|| \ge \dots \ge ||r_{m-j}^{(1)}|| > ||r_{m-j+1}^{(1)}|| = \dots = ||r_m^{(1)}||$$

then the first j-1 residual norms stagnate in the second cycle,

$$||r_0^{(2)}|| = ||r_1^{(2)}|| = \dots = ||r_{i-1}^{(2)}||.$$

Hence stagnation in one cycle is literally mirrored in the next cycle!



Outline

- Prescribing residual norms and Ritz values in GMRES(m)
- 2 Without stagnation at the end of cycles
- 3 Allowing stagnation at the end of cycles
- 4 Brief discussion
- 5 Conclusions

Preconditioning restarted GMRES

The previous results have a number of theoretical implications for strategies to accelerate restarted GMRES like preconditioning.

Any convergence speed of restarted GMRES is possible with any spectrum, therefore:

- A preconditioner that clusters eigenvalues needs not accelerate GMRES(m).
- Additional spectral information is necessary to guarantee acceleration.
- An important example is constraint preconditioning, where the few distinct eigenvalues of the preconditioned matrix belong to small Jordan blocks.

Acceleration of restarted GMRES

Our results also have consequences for spectral acceleration techniques (often called deflation techniques, but deflation needs not exploit spectral quantities, see, e.g. [Nabben & Vuik 2004, 2006, 2008]):

- The suspicion is that outlying eigenvalues, mostly eigenvalues close to zero, hamper convergence
- Eigenvalue approximations are obtained from the Ritz or harmonic Ritz values generated during the GMRES(m) process
- The corresponding eigenvectors (or invariant subspaces) are used to eliminate the influence of convergence hampering eigenvalues
- This can be done through preconditioning, augmentation of the Krylov subspaces, projecting away invariant subspaces or a combination of these.

Acceleration of restarted GMRES

Any nonzero Ritz values can be generated by restarted GMRES, therefore:

- There is no guarantee that spectral acceleration techniques will find good approximate eigenvalues from Ritz values.
- The same appears to hold for harmonic Ritz values
- Additionally, and again, any convergence speed of GMRES(m) is possible with any spectrum
- Therefore, eigenvalues close to zero need not hamper convergence at all

Note that we showed that a zero Ritz value *does* imply stagnation. At the end of a cycle it even also implies stagnation at the beginning of the next cycle.

Outline

- 1 Prescribing residual norms and Ritz values in GMRES(m)
- 2 Without stagnation at the end of cycles
- 3 Allowing stagnation at the end of cycles
- 4 Brief discussion
- Conclusions

ullet Any prescribed non-stagnating residual norms and nonzero Ritz values are possible for the first n iterations of restarted GMRES, with any spectrum of A.

- ullet Any prescribed non-stagnating residual norms and nonzero Ritz values are possible for the first n iterations of restarted GMRES, with any spectrum of A.
- When we prescribe stagnation at the end of a cycle, we must also prescribe it at the beginning of the next cycle. The same holds for zero Ritz values at the end of a cycle.

- ullet Any prescribed non-stagnating residual norms and nonzero Ritz values are possible for the first n iterations of restarted GMRES, with any spectrum of A.
- When we prescribe stagnation at the end of a cycle, we must also prescribe it at the beginning of the next cycle. The same holds for zero Ritz values at the end of a cycle.
- Many spectral acceleration techniques are very efficient in practice, but lack a sound theoretical explication for their success.

- ullet Any prescribed non-stagnating residual norms and nonzero Ritz values are possible for the first n iterations of restarted GMRES, with any spectrum of A.
- When we prescribe stagnation at the end of a cycle, we must also prescribe it at the beginning of the next cycle. The same holds for zero Ritz values at the end of a cycle.
- Many spectral acceleration techniques are very efficient in practice, but lack a sound theoretical explication for their success.
- Questions for future work include:

- ullet Any prescribed non-stagnating residual norms and nonzero Ritz values are possible for the first n iterations of restarted GMRES, with any spectrum of A.
- When we prescribe stagnation at the end of a cycle, we must also prescribe it at the beginning of the next cycle. The same holds for zero Ritz values at the end of a cycle.
- Many spectral acceleration techniques are very efficient in practice, but lack a sound theoretical explication for their success.
- Questions for future work include:
 - Are comparable results possible for Krylov subspace methods with short recurrences (Bi-CG, Bi-CGStab,...)?

- ullet Any prescribed non-stagnating residual norms and nonzero Ritz values are possible for the first n iterations of restarted GMRES, with any spectrum of A.
- When we prescribe stagnation at the end of a cycle, we must also prescribe it at the beginning of the next cycle. The same holds for zero Ritz values at the end of a cycle.
- Many spectral acceleration techniques are very efficient in practice, but lack a sound theoretical explication for their success.
- Questions for future work include:
 - Are comparable results possible for Krylov subspace methods with short recurrences (Bi-CG, Bi-CGStab,...)?
 - ullet What can be said for GMRES(m) after iteration number n ?

Related papers

- A. Greenbaum and Z. Strakoš, [Matrices that generate the same Krylov residual spaces, IMA Vol. Math. Appl., 60 (1994), pp. 95–118.]
- A. Greenbaum, V. Pták and Z. Strakoš, [Any nonincreasing convergence curve is possible for GMRES, SIMAX, 17 (1996), pp. 465–469.]
- M. Arioli, V. Pták and Z. Strakoš, [Krylov sequences of maximal length and convergence of GMRES, BIT, 38 (1996), pp. 636–643.]
- J. Duintjer Tebbens and G. Meurant, [Any Ritz value behavior is possible for Arnoldi and for GMRES, SIMAX, 33 (2012), pp. 958–978.]
- J. Duintjer Tebbens and G. Meurant, [Prescribing the behavior of early terminating GMRES and Arnoldi iterations, Numer. Algorithms, online first February 2013]

Thank you for your attention!