The roots of GMRES polynomials need not influence GMRES residual norms

Jurjen Duintjer Tebbens

joint work with

Gérard Meurant and Kui Du

Numerical Linear Algebra with Applications (NL2A)
CIRM Luminy, France October 24, 2016.

Outline

Outline

1. Eigenvalues and convergence of Krylov subspace methods
2. Some recent results
3. Harmonic Ritz values

Outline

1. Eigenvalues and convergence of Krylov subspace methods
2. Some recent results
3. Harmonic Ritz values

Eigenvalues and convergence of Krylov subspace methods

The convergence of Krylov subspace methods for linear systems

$$Ax = b, \qquad A \in \mathbb{C}^{n \times n}, b \in \mathbb{C}^n$$

with a normal matrix A is sometimes said to be governed by eigenvalues:

Eigenvalues and convergence of Krylov subspace methods

The convergence of Krylov subspace methods for linear systems

$$Ax = b, \qquad A \in \mathbb{C}^{n\times n}, b \in \mathbb{C}^n$$

with a normal matrix A is sometimes said to be governed by eigenvalues:

- the eigenvalue distribution decides about the possibly worst rate of convergence,

Eigenvalues and convergence of Krylov subspace methods

The convergence of Krylov subspace methods for linear systems

$$Ax = b, \qquad A \in \mathbb{C}^{n\times n}, b \in \mathbb{C}^n$$

with a normal matrix A is sometimes said to be governed by eigenvalues:

- the eigenvalue distribution decides about the possibly worst rate of convergence,
- eigenvalues close to zero hamper convergence,

Eigenvalues and convergence of Krylov subspace methods

The convergence of Krylov subspace methods for linear systems

$$Ax = b, \qquad A \in \mathbb{C}^{n\times n}, b \in \mathbb{C}^n$$

with a normal matrix A is sometimes said to be governed by eigenvalues:

- the eigenvalue distribution decides about the possibly worst rate of convergence,
- eigenvalues close to zero hamper convergence,
- clustering of eigenvalues stimulates convergence.

Eigenvalues and convergence of Krylov subspace methods

For instance, in the MINRES method, residual norms satisfy the minimization property

$$\|r_k\| = \min_{p \in \pi_k} \|p(A)b\|, \qquad \pi_k \text{ the degree } k \text{ polynomials with } \pi_k(0) = 1.$$

Eigenvalues and convergence of Krylov subspace methods

For instance, in the MINRES method, residual norms satisfy the minimization property

$$\|r_k\| = \min_{p \in \pi_k} \|p(A)b\|, \qquad \pi_k \text{ the degree } k \text{ polynomials with } \pi_k(0) = 1.$$

Plugging in the spectral decomposition $A = W \Lambda W^*$,

$$\|r_k\| = \min_{p \in \pi_k} \|p(A)b\| = \min_{p \in \pi_k} \|p\,(\Lambda) W^* b\|.$$

Eigenvalues and convergence of Krylov subspace methods

For instance, in the MINRES method, residual norms satisfy the minimization property

$$\|r_k\| = \min_{p \in \pi_k} \|p(A)b\|, \qquad \pi_k \text{ the degree } k \text{ polynomials with } \pi_k(0) = 1.$$

Plugging in the spectral decomposition $A = W \Lambda W^*$,

$$\|r_k\| = \min_{p \in \pi_k} \|p(A)b\| = \min_{p \in \pi_k} \|p(\Lambda) W^* b\|.$$

Thus residual norms are fully determined by two quantities:

1. eigenvalues,
2. components of the right-hand side in the eigenvector basis.

Eigenvalues and convergence of Krylov subspace methods

The equality

$$\|r_k\| = \min_{p \in \pi_k} \|p(\Lambda) W^* b\|$$

leads to well-known bounds like

Eigenvalues and convergence of Krylov subspace methods

The equality

$$\|r_k\| = \min_{p\,\in\pi_k} \|p\,(\Lambda)W^*b\|$$

leads to well-known bounds like

- $$\frac{\|r_k\|}{\|b\|} \leq \min_{p\in\pi_k} \max_{i=1,\ldots,n} |p_k(\lambda_i)|,$$

  which is sharp (for every k separately), and

Eigenvalues and convergence of Krylov subspace methods

The equality

$$\|r_k\| = \min_{p\,\in\pi_k} \|p\,(\Lambda)W^* b\|$$

leads to well-known bounds like

- $$\frac{\|r_k\|}{\|b\|} \leq \min_{p\in\pi_k} \max_{i=1,\dots,n} |p_k(\lambda_i)|,$$

  which is sharp (for every k separately), and

- $$\frac{\|x - x_k\|_A}{\|x - x_0\|_A} \leq 2\left(\frac{\sqrt{\kappa(A)} - 1}{\sqrt{\kappa(A)} + 1}\right)^k$$

  for the CG method.

Eigenvalues and convergence of Krylov subspace methods

The solution of the minimization problem

$$\|r_k\|^2 = \min_{p \in \pi_k} \|p(\Lambda) W^* b\|^2$$

can be expressed in closed-form [Bellalij and Sadok, 2011], [DT, Meurant, Sadok and Strakoš, 2014]:

Eigenvalues and convergence of Krylov subspace methods

The solution of the minimization problem

$$\|r_k\|^2 = \min_{p \in \pi_k} \|p(\Lambda) W^* b\|^2$$

can be expressed in closed-form [Bellalij and Sadok, 2011], [DT, Meurant, Sadok and Strakoš, 2014]:

$$\|r_k\|^2 = \frac{\sum_{I_{k+1}} \left[\prod_{j=1}^{k+1} \gamma_{i_j}\right] \prod_{\substack{i_1 \le i_\ell < i_j \le i_{k+1} \\ i_\ell, i_j \in I_{k+1}}} |\lambda_{i_j} - \lambda_{i_\ell}|^2}{\sum_{I_k} \left[\prod_{j=1}^{k} \gamma_{i_j} |\lambda_{i_j}|^2\right] \prod_{\substack{i_1 \le i_\ell < i_j \le i_k \\ i_\ell, i_j \in I_k}} |\lambda_{i_j} - \lambda_{i_\ell}|^2},$$

where $\gamma_{i_j} = |e_{i_j}^T c|^2$, $c = W^* b$ and $\sum_{I_k}$ denote summation over all possible sets I_k of k indices $i_1, i_2, \ldots, i_k$ such that $1 \le i_1 < \cdots < i_k \le n$.

Eigenvalues and convergence of Krylov subspace methods

With non-normal matrices A, convergence need not be governed by eigenvalues, the probably most convincing evidence being

Eigenvalues and convergence of Krylov subspace methods

With non-normal matrices A, convergence need not be governed by eigenvalues, the probably most convincing evidence being

Theorem [Greenbaum, Pták & Strakoš, 1996]. *Let*

$$\|b\|_2 = f_0 \geq f_1 \geq f_2 \cdots \geq f_{n-1} > 0$$

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

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

be any set of nonzero complex numbers.

Eigenvalues and convergence of Krylov subspace methods

With non-normal matrices A, convergence need not be governed by eigenvalues, the probably most convincing evidence being

Theorem [Greenbaum, Pták & Strakoš, 1996]. *Let*

$$\|b\|_2 = f_0 \geq f_1 \geq f_2 \cdots \geq f_{n-1} > 0$$

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 the GMRES method applied to A and b satisfy

$$\|r_k\|_2 = f_k, \quad 0 \leq k \leq n, \quad \text{and} \quad \text{eig}(A) = \{\lambda_1, \ldots, \lambda_n\}.$$

Eigenvalues and convergence of Krylov subspace methods

In [Arioli, Pták & Strakoš, 1998] a full parametrization of the class generating prescribed residual norms and eigenvalues was given. It was based on orthogonal bases for the Krylov residual subspaces $A\mathcal{K}_k(A,b)$.

Eigenvalues and convergence of Krylov subspace methods

In [Arioli, Pták & Strakoš, 1998] a full parametrization of the class generating prescribed residual norms and eigenvalues was given. It was based on orthogonal bases for the Krylov residual subspaces $A\mathcal{K}_k(A, b)$.

In [DT & Meurant 2013] a parametrization based on orthogonal bases for the Krylov subspaces $\mathcal{K}_k(A, b)$ is described:

Eigenvalues and convergence of Krylov subspace methods

In [Arioli, Pták & Strakoš, 1998] a full parametrization of the class generating prescribed residual norms and eigenvalues was given. It was based on orthogonal bases for the Krylov residual subspaces $A\mathcal{K}_k(A,b)$.

In [DT & Meurant 2013] a parametrization based on orthogonal bases for the Krylov subspaces $\mathcal{K}_k(A,b)$ is described:

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

$$A = VHV^*, \qquad H \text{ upper Hessenberg.}$$

Eigenvalues and convergence of Krylov subspace methods

In [Arioli, Pták & Strakoš, 1998] a full parametrization of the class generating prescribed residual norms and eigenvalues was given. It was based on orthogonal bases for the Krylov residual subspaces $A\mathcal{K}_k(A, b)$.

In [DT & Meurant 2013] a parametrization based on orthogonal bases for the Krylov subspaces $\mathcal{K}_k(A, b)$ is described:

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

$$A = VHV^*, \qquad H \text{ upper Hessenberg.}$$

- To force the desired eigenvalues, H is of the form

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

where C is the companion matrix for the prescribed spectrum.

Eigenvalues and convergence of Krylov subspace methods

Let

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

Eigenvalues and convergence of Krylov subspace methods

Let

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

- To force GMRES residual norms $f(0) \geq \cdots \geq f(n-1) > 0$, the first row g^T of U can be chosen as

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

Eigenvalues and convergence of Krylov subspace methods

Let

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

- To force GMRES residual norms $f(0) \geq \cdots \geq f(n-1) > 0$, the first row g^T of U can be chosen as

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

Summarizing,

$$A = V \begin{bmatrix} & g^T & \\ 0 & & T \end{bmatrix}^{-1} C \begin{bmatrix} & g^T & \\ 0 & & T \end{bmatrix} V^*, \qquad b = Ve_1,$$

for some non-singular, upper triangular matrix $T \in \mathbb{C}^{(n-1)\times(n-1)}$.

Eigenvalues and convergence of Krylov subspace methods

Thus for assessing the quality of a preconditioner $\mathcal{P}$ when GMRES is applied to

$$\mathcal{P}Ax = \mathcal{P}b, \qquad \mathcal{P}A \quad \text{non-symmetric},$$

this means that analysis of the spectrum of $\mathcal{P}A$ alone is not enough.

Eigenvalues and convergence of Krylov subspace methods

Thus for assessing the quality of a preconditioner $\mathcal{P}$ when GMRES is applied to

$$\mathcal{P}Ax = \mathcal{P}b, \qquad \mathcal{P}A \quad \text{non-symmetric},$$

this means that analysis of the spectrum of $\mathcal{P}A$ alone is not enough.

For instance in constraint preconditioning, the fact that spec($\mathcal{P}A$) is, say,

$$\text{spec } (\mathcal{P}A) = \{1, \frac{1}{2} \pm \frac{\sqrt{(5)}}{2}\}$$

does not suffice to guarantee fast convergence of GMRES when $\mathcal{P}A$ is non-symmetric. What is needed additionally, is the fact that the eigenvalues have belong to Jordan blocks of small size.

Eigenvalues and convergence of Krylov subspace methods

Not either need eigenvalues close to zero hamper convergence.

Eigenvalues and convergence of Krylov subspace methods

Not either need eigenvalues close to zero hamper convergence.

This somewhat undermines the theoretical foundations of deflation methods, which aim at speeding up convergence by elimination of eigenspaces corresponding to small (or other presumed unfavorable) eigenvalues.

Eigenvalues and convergence of Krylov subspace methods

Not either need eigenvalues close to zero hamper convergence.

This somewhat undermines the theoretical foundations of deflation methods, which aim at speeding up convergence by elimination of eigenspaces corresponding to small (or other presumed unfavorable) eigenvalues.

One may wonder whether counterintuitive examples are always of an academic character? For instance, they may be far from normal and not satisfy any sparsity pattern.

Eigenvalues and convergence of Krylov subspace methods

Not either need eigenvalues close to zero hamper convergence.

This somewhat undermines the theoretical foundations of deflation methods, which aim at speeding up convergence by elimination of eigenspaces corresponding to small (or other presumed unfavorable) eigenvalues.

One may wonder whether counterintuitive examples are always of an academic character? For instance, they may be far from normal and not satisfy any sparsity pattern.

Practical examples with a tridiagonal matrix where it is hard to explain GMRES convergence based on eigenvalue distribution are given by for instance some convection-diffusion model problems that have been studied by many authors, e.g. [Fischer, Ramage, Sylvester & Wathen, 1999], [Ernst, 2000], [Elman & Ramage, 2001, 2002], [Liesen & Strakoš, 2005]:

Eigenvalues and convergence of Krylov subspace methods

Consider the convection-diffusion equation

$$\begin{aligned} -\nu\nabla^2 u + w\cdot\nabla u = 0 \qquad &\text{in} \qquad \Omega = (0,1)\times(0,1), \\ u = g \qquad &\text{on} \qquad \partial\Omega, \end{aligned}$$

where we use

Eigenvalues and convergence of Krylov subspace methods

Consider the convection-diffusion equation

$$\begin{aligned} -\nu\nabla^2 u + w\cdot\nabla u = 0 \qquad &\text{in} \qquad \Omega = (0,1)\times(0,1),\\ u = g \qquad &\text{on} \qquad \partial\Omega, \end{aligned}$$

where we use

- $\nu = 0.01$: scalar diffusion coefficient

Eigenvalues and convergence of Krylov subspace methods

Consider the convection-diffusion equation

$$
\begin{aligned}
-\nu\nabla^2 u + w \cdot \nabla u = 0 \qquad &\text{in} \qquad \Omega = (0,1)\times(0,1),\\
u = g \qquad &\text{on} \qquad \partial\Omega,
\end{aligned}
$$

where we use

- $\nu = 0.01$: scalar diffusion coefficient
- $w = [0,1]^T$: velocity field (wind)

Eigenvalues and convergence of Krylov subspace methods

Consider the convection-diffusion equation

$$
\begin{aligned}
-\nu\nabla^2 u + w\cdot\nabla u = 0 \qquad & \text{in} \qquad \Omega = (0,1)\times(0,1),\\
u = g \qquad & \text{on} \qquad \partial\Omega,
\end{aligned}
$$

where we use

- $\nu = 0.01$: scalar diffusion coefficient
- $w = [0,1]^T$: velocity field (wind)
- an N by N grid with spacing $h = 1/(N+1)$ where $N = 16$, i.e. the system matrix has size 256.

Eigenvalues and convergence of Krylov subspace methods

Consider the convection-diffusion equation

$$\begin{aligned} -\nu \nabla^2 u + w \cdot \nabla u &= 0 \qquad \text{in} \qquad \Omega = (0,1) \times (0,1), \\ u &= g \qquad \text{on} \qquad \partial\Omega, \end{aligned}$$

where we use

- $\nu = 0.01$: scalar diffusion coefficient
- $w = [0,1]^T$: velocity field (wind)
- an N by N grid with spacing $h = 1/(N+1)$ where $N = 16$, i.e. the system matrix has size 256.
- bilinear finite element nodal basis functions ϕ_j, $j = 1, 2, \ldots$

Eigenvalues and convergence of Krylov subspace methods

Consider the convection-diffusion equation

$$\begin{aligned} -\nu\nabla^2 u + w\cdot\nabla u = 0 \qquad &\text{in} \qquad \Omega = (0,1)\times(0,1), \\ u = g \qquad &\text{on} \qquad \partial\Omega, \end{aligned}$$

where we use

- $\nu = 0.01$: scalar diffusion coefficient
- $w = [0,1]^T$: velocity field (wind)
- an N by N grid with spacing $h = 1/(N+1)$ where $N = 16$, i.e. the system matrix has size 256.
- bilinear finite element nodal basis functions ϕ_j, $j = 1, 2, \ldots$
- Raithby boundary conditions

Eigenvalues and convergence of Krylov subspace methods

Consider the convection-diffusion equation

$$\begin{aligned} -\nu\nabla^2 u + w \cdot \nabla u = 0 \qquad & \text{in} \qquad \Omega = (0,1)\times(0,1), \\ u = g \qquad & \text{on} \qquad \partial\Omega, \end{aligned}$$

where we use

- $\nu = 0.01$: scalar diffusion coefficient
- $w = [0,1]^T$: velocity field (wind)
- an N by N grid with spacing $h = 1/(N+1)$ where $N = 16$, i.e. the system matrix has size 256.
- bilinear finite element nodal basis functions ϕ_j, $j = 1, 2, \ldots$
- Raithby boundary conditions

With Streamline Upwind Petrov Galerkin discretization (SUPG) [Brooks & Hughes, 1979], the coefficient matrix A is of the form

$$A_{i,j} = \nu\langle\nabla\phi_j, \nabla\phi_i\rangle + \langle w\cdot\nabla\phi_j, \phi_i\rangle + \frac{\delta\, h}{\|w\|}\langle w\cdot\nabla\phi_j, w\cdot\nabla\phi_i\rangle,$$

where $0 < \delta < 1$ and $\langle\cdot,\cdot\rangle$ denotes the L^2 inner product on Ω.

Eigenvalues and convergence of Krylov subspace methods

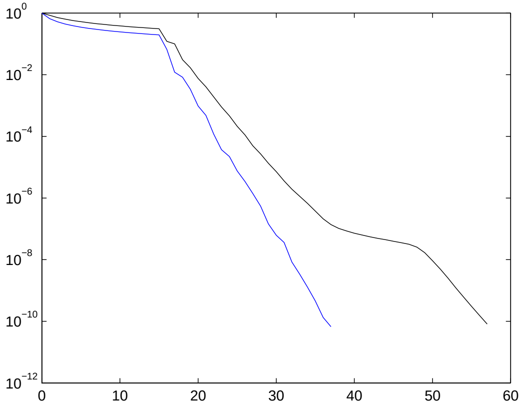


Figure : GMRES residual norms for the convection-diffusion problem with SUPG stabilization parameters $\delta = 0.7$ (blue) and $\delta = 0.2$ (black).

Eigenvalues and convergence of Krylov subspace methods

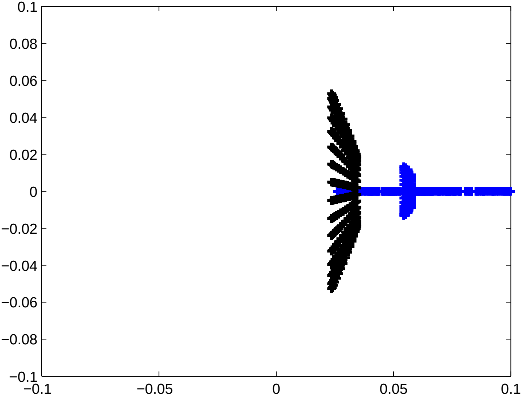


Figure : Spectra of system matrices for the convection-diffusion problem with SUPG stabilization parameters $\delta = 0.7$ (blue) and $\delta = 0.2$ (black).

Eigenvalues and convergence of Krylov subspace methods

A further aspect of eigenvalue dominance is the link between superlinear convergence and Ritz value convergence:

Eigenvalues and convergence of Krylov subspace methods

A further aspect of eigenvalue dominance is the link between superlinear convergence and Ritz value convergence:

- In the CG method for Hermitian positive definite matrices, a converged Ritz value often implies an accelerated phase of convergence of the A-norm of the error, see, e.g., [van der Sluis & van der Vorst, 1986].

Eigenvalues and convergence of Krylov subspace methods

A further aspect of eigenvalue dominance is the link between superlinear convergence and Ritz value convergence:

- In the CG method for Hermitian positive definite matrices, a converged Ritz value often implies an accelerated phase of convergence of the A-norm of the error, see, e.g., [van der Sluis & van der Vorst, 1986].
- An analogue result for the GMRES method suggests a similar phenomenon provided A is close to normal - the involved bounds contain $\kappa(W)$ [van der Vorst & Vuik, 1993].

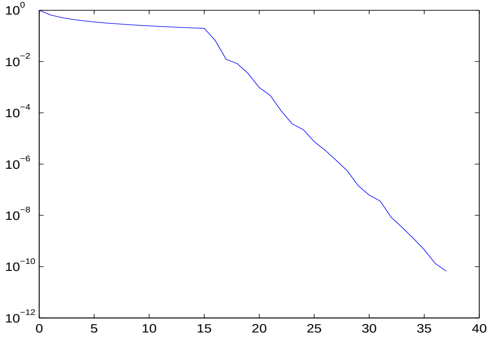
Eigenvalues and convergence of Krylov subspace methods

A further aspect of eigenvalue dominance is the link between superlinear convergence and Ritz value convergence:

- In the CG method for Hermitian positive definite matrices, a converged Ritz value often implies an accelerated phase of convergence of the A-norm of the error, see, e.g., [van der Sluis & van der Vorst, 1986].
- An analogue result for the GMRES method suggests a similar phenomenon provided A is close to normal - the involved bounds contain $\kappa(W)$ [van der Vorst & Vuik, 1993].

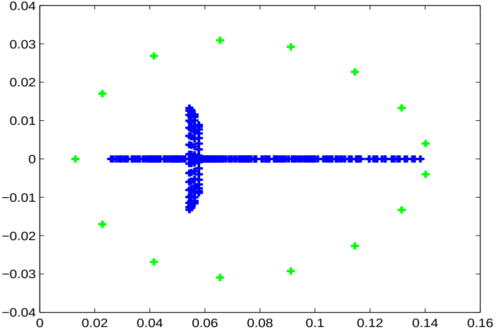
For the higly non-normal case, let us consider the convection-diffusion problem with SUPG stabilization parameter $\delta = 0.7$.

Eigenvalues and convergence of Krylov subspace methods



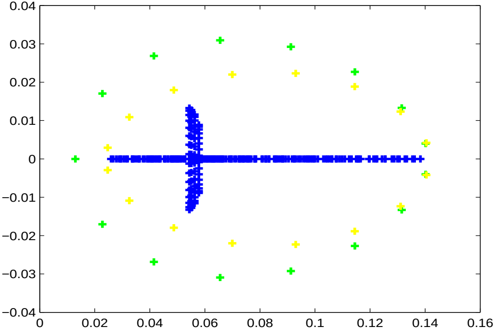
GMRES residual norms for the convection-diffusion problem with SUPG stabilization parameters $\delta = 0.7$.

Eigenvalues and convergence of Krylov subspace methods



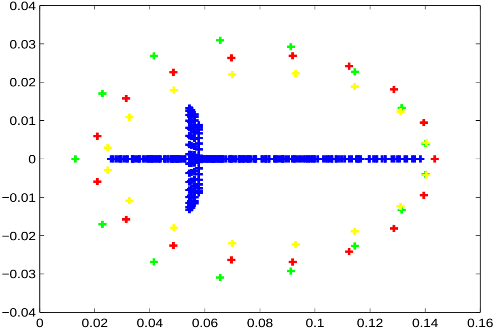
Blue: Eigenvalues of $A \in \mathbb{R}^{256\times 256}$. Green: Ritz values at iteration 15.

Eigenvalues and convergence of Krylov subspace methods



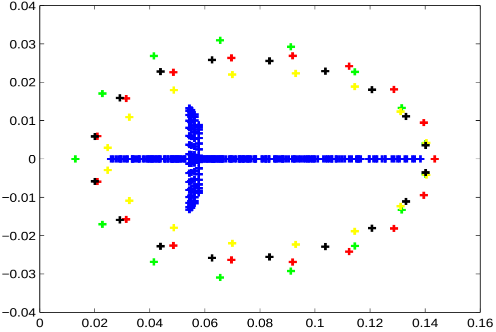
Blue: Eigenvalues of $A \in \mathbb{R}^{256\times 256}$. Yellow: Ritz values at iteration 16.

Eigenvalues and convergence of Krylov subspace methods



Blue: Eigenvalues of $A \in \mathbb{R}^{256 \times 256}$. Red: Ritz values at iteration 17.

Eigenvalues and convergence of Krylov subspace methods



Blue: Eigenvalues of $A \in \mathbb{R}^{256\times 256}$. Black: Ritz values at iteration 18.

Eigenvalues and convergence of Krylov subspace methods

In the next section we address the following two questions:

Eigenvalues and convergence of Krylov subspace methods

In the next section we address the following two questions:

- What do we know about the relation between Ritz value convergence and GMRES convergence for general non-normal matrices ?

Eigenvalues and convergence of Krylov subspace methods

In the next section we address the following two questions:

- What do we know about the relation between Ritz value convergence and GMRES convergence for general non-normal matrices ?
- Do similar results exist for Krylov subspace methods for non-normal matrices other than GMRES?

Outline

Some recent results

An analogue of the Greenbaum, Pták, Strakoš theorem for the FOM method is straightforward using the relation

$$\frac{1}{\|r_k^{FOM}\|^2} = \frac{1}{\|r_k^{GMRES}\|^2} - \frac{1}{\|r_{k-1}^{GMRES}\|^2} \; :$$

Some recent results

An analogue of the Greenbaum, Pták, Strakoš theorem for the FOM method is straightforward using the relation

$$\frac{1}{\|r_k^{FOM}\|^2} = \frac{1}{\|r_k^{GMRES}\|^2} - \frac{1}{\|r_{k-1}^{GMRES}\|^2} :$$

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

$$A = VHV^*, \qquad H \text{ upper Hessenberg.}$$

Some recent results

An analogue of the Greenbaum, Pták, Strakoš theorem for the FOM method is straightforward using the relation

$$\frac{1}{\|r_k^{FOM}\|^2} = \frac{1}{\|r_k^{GMRES}\|^2} - \frac{1}{\|r_{k-1}^{GMRES}\|^2} :$$

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

$$A = VHV^*, \qquad H \text{ upper Hessenberg.}$$

- To force the desired eigenvalues, H is of the form

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

where C is the companion matrix for the prescribed spectrum.

Some recent results

An analogue of the Greenbaum, Pták, Strakoš theorem for the FOM method is straightforward using the relation

$$\frac{1}{\|r_k^{FOM}\|^2} = \frac{1}{\|r_k^{GMRES}\|^2} - \frac{1}{\|r_{k-1}^{GMRES}\|^2} :$$

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

  $$A = VHV^*, \qquad H \text{ upper Hessenberg.}$$

- To force the desired eigenvalues, H is of the form

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

  where C is the companion matrix for the prescribed spectrum.

- To force FOM residual norms $f(0), \ldots, f(n-1)$, $f(i) > 0$, the first row g^T of U can be chosen as

  $$g_k = \frac{1}{f(k-1)}, \quad k = 1, \ldots, n.$$

Some recent results

The similarity transformation matrix U in

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

satisfies

$$U = [e_1, He_1, \dots, H^{n-1}e_1]^{-1}$$

Some recent results

The similarity transformation matrix U in

$$H = U^{-1} C U, \qquad U \text{ nonsingular upper triangular,}$$

satisfies

$$U = [e_1, He_1, \ldots, H^{n-1} e_1]^{-1}$$

and is the change of basis matrix in the transition from the Krylov matrix

$$K = [b, Ab, \ldots, A^{n-1} b]$$

to the orthogonal basis V,

$$KU = V.$$

Some recent results

The similarity transformation matrix U in

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

satisfies

$$U = [e_1, He_1, \ldots, H^{n-1}e_1]^{-1}$$

and is the change of basis matrix in the transition from the Krylov matrix

$$K = [b, Ab, \ldots, A^{n-1}b]$$

to the orthogonal basis V,

$$KU = V.$$

In fact, even if V is not orthogonal, but in some Krylov subspace method $KU = V$ holds, the first row of U gives the residual norm of the Krylov subspace method working with V. For example, when V is bi-orthogonal to a basis for $\mathcal{K}(A^*, s)$, it gives Bi-CG residual norms:

Some recent results

To force desired eigenvalues and Bi-CG residual norms we can

- Choose a nonsingular matrix V with normalized columns and put $b = Ve_1$ and
$$A = VTV^{-1}, \qquad T \text{ tridiagonal.}$$

Some recent results

To force desired eigenvalues and Bi-CG residual norms we can

- Choose a nonsingular matrix V with normalized columns and put $b = Ve_1$ and
$$A = VTV^{-1}, \qquad T \text{ tridiagonal.}$$

- Try to find a tridiagonal T allowing the decomposition
$$T = U^{-1}CU, \qquad U \text{ nonsingular upper triangular,}$$
where C is the companion matrix for the prescribed spectrum and

Some recent results

To force desired eigenvalues and Bi-CG residual norms we can

- Choose a nonsingular matrix V with normalized columns and put $b = Ve_1$ and
$$A = VTV^{-1}, \qquad T \text{ tridiagonal.}$$

- Try to find a tridiagonal T allowing the decomposition
$$T = U^{-1}CU, \qquad U \text{ nonsingular upper triangular,}$$
where C is the companion matrix for the prescribed spectrum and where the first row g^T of U has entries
$$g_k = \frac{1}{f(k-1)}, \quad k = 1, \ldots, n,$$
if $f(0), \ldots, f(n-1)$, $f(i) > 0$ are the prescribed Bi-CG residual norms.

Some recent results

It can be shown that tridiagonal T allowing the decomposition

$$T = U^{-1}CU, \qquad U \text{ nonsingular upper triangular,}$$

can be found by applying a Bi-Lanczos process to C [DT & Meurant, 2016].

Some recent results

It can be shown that tridiagonal T allowing the decomposition

$$T = U^{-1}CU, \qquad U \text{ nonsingular upper triangular,}$$

can be found by applying a Bi-Lanczos process to C [DT & Meurant, 2016].

Thus *any* Bi-CG convergence history is possible with *any* spectrum.

Some recent results

It can be shown that tridiagonal T allowing the decomposition

$$T = U^{-1}CU, \qquad U \text{ nonsingular upper triangular},$$

can be found by applying a Bi-Lanczos process to C [DT & Meurant, 2016].

Thus *any* Bi-CG convergence history is possible with *any* spectrum.

And by using the relation

$$\frac{1}{\|r_k^{Bi-CG}\|^2} = \frac{1}{\|r_k^{QMR}\|^2} - \frac{1}{\|r_{k-1}^{QMR}\|^2},$$

any QMR convergence history is possible with *any* spectrum.

Some recent results

It can be shown that tridiagonal T allowing the decomposition

$$T = U^{-1}CU, \qquad U \text{ nonsingular upper triangular},$$

can be found by applying a Bi-Lanczos process to C [DT & Meurant, 2016].

Thus *any* Bi-CG convergence history is possible with *any* spectrum.

And by using the relation

$$\frac{1}{\|r_k^{Bi-CG}\|^2} = \frac{1}{\|r_k^{QMR}\|^2} - \frac{1}{\|r_{k-1}^{QMR}\|^2},$$

any QMR convergence history is possible with *any* spectrum.

The same can be done for the OR/MR pair of methods Hessenberg/CMRH based on Hessenberg LU factorization [Sadok, 1999], [Heyouni & Sadok, 1998] .

Some recent results

It can be shown that tridiagonal T allowing the decomposition

$$T = U^{-1}CU, \qquad U \text{ nonsingular upper triangular,}$$

can be found by applying a Bi-Lanczos process to C [DT & Meurant, 2016].

Thus *any* Bi-CG convergence history is possible with *any* spectrum.

And by using the relation

$$\frac{1}{\|r_k^{Bi-CG}\|^2} = \frac{1}{\|r_k^{QMR}\|^2} - \frac{1}{\|r_{k-1}^{QMR}\|^2},$$

any QMR convergence history is possible with *any* spectrum.

The same can be done for the OR/MR pair of methods Hessenberg/CMRH based on Hessenberg LU factorization [Sadok, 1999], [Heyouni & Sadok, 1998] .

In fact, it can be done for any Krylov subspace method for non-normal matrices that is classified as a OR or MR-type method.

Some recent results

In the parametrization for GMRES/FOM

$$A = V \begin{bmatrix} & g^T & \\ 0 & & T \end{bmatrix}^{-1} C \begin{bmatrix} & g^T & \\ 0 & & T \end{bmatrix} V^*, \qquad b = Ve_1,$$

the non-singular, upper triangular matrix $T \in \mathbb{C}^{(n-1)\times(n-1)}$ is a free parameter.

Some recent results

In the parametrization for GMRES/FOM

$$A = V \left[\begin{array}{cc} & g^T \\ 0 & T \end{array} \right]^{-1} C \left[\begin{array}{cc} & g^T \\ 0 & T \end{array} \right] V^*, \qquad b = Ve_1,$$

the non-singular, upper triangular matrix $T \in \mathbb{C}^{(n-1)\times(n-1)}$ is a free parameter.

In the GMRES and FOM methods it can be used to prescribe all Ritz values:

Some recent results

In the parametrization for GMRES/FOM

$$A = V \begin{bmatrix} & g^T & \\ 0 & & T \end{bmatrix}^{-1} C \begin{bmatrix} & g^T & \\ 0 & & T \end{bmatrix} V^*, \qquad b = Ve_1,$$

the non-singular, upper triangular matrix $T \in \mathbb{C}^{(n-1)\times(n-1)}$ is a free parameter.

In the GMRES and FOM methods it can be used to prescribe all Ritz values:

To force at iteration k the Ritz values $\rho_1^{(k)} \dots \rho_k^{(k)}$, $k = 1, \dots, n-1$, the entries $t_{i,k}$ in the kth column of T must satisfy

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

Some recent results

This shows that:

- Any Ritz value behavior can be prescribed for the Arnoldi method, in all its iterations

Some recent results

This shows that:

- Any Ritz value behavior can be prescribed for the Arnoldi method, in all its iterations
- Any Ritz value behavior is possible with any GMRES/FOM residual norm history.

Some recent results

This shows that:

- Any Ritz value behavior can be prescribed for the Arnoldi method, in all its iterations
- Any Ritz value behavior is possible with any GMRES/FOM residual norm history.

Thus theoretical foundations of deflation methods using Ritz value approximations are even weaker:

Some recent results

This shows that:

- Any Ritz value behavior can be prescribed for the Arnoldi method, in all its iterations
- Any Ritz value behavior is possible with any GMRES/FOM residual norm history.

Thus theoretical foundations of deflation methods using Ritz value approximations are even weaker:

- The used Ritz values need not approximate eigenvalues at all,

Some recent results

This shows that:

- Any Ritz value behavior can be prescribed for the Arnoldi method, in all its iterations
- Any Ritz value behavior is possible with any GMRES/FOM residual norm history.

Thus theoretical foundations of deflation methods using Ritz value approximations are even weaker:

- The used Ritz values need not approximate eigenvalues at all,
- Even if they do approximate eigenvalues, these need not influence GMRES residual norms.

Some recent results

This shows that:

- Any Ritz value behavior can be prescribed for the Arnoldi method, in all its iterations
- Any Ritz value behavior is possible with any GMRES/FOM residual norm history.

Thus theoretical foundations of deflation methods using Ritz value approximations are even weaker:

- The used Ritz values need not approximate eigenvalues at all,
- Even if they do approximate eigenvalues, these need not influence GMRES residual norms.

But many deflation methods use instead harmonic Ritz values.

Some recent results

This shows that:

- Any Ritz value behavior can be prescribed for the Arnoldi method, in all its iterations
- Any Ritz value behavior is possible with any GMRES/FOM residual norm history.

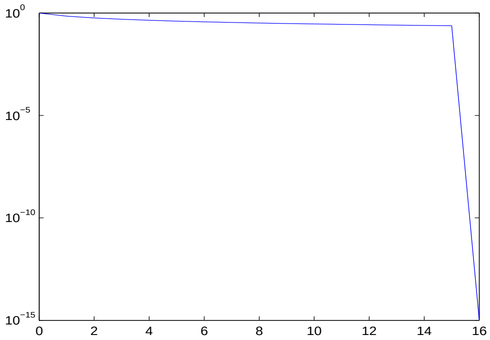
Thus theoretical foundations of deflation methods using Ritz value approximations are even weaker:

- The used Ritz values need not approximate eigenvalues at all,
- Even if they do approximate eigenvalues, these need not influence GMRES residual norms.

But many deflation methods use instead harmonic Ritz values.

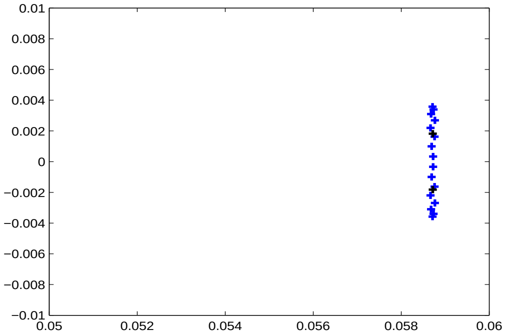
Let us illustrate with a small 1D convection-diffusion problem.

Some recent results



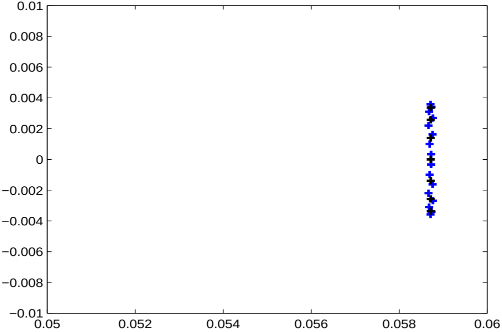
GMRES residual norms for a 1D convection-diffusion problem.

Some recent results



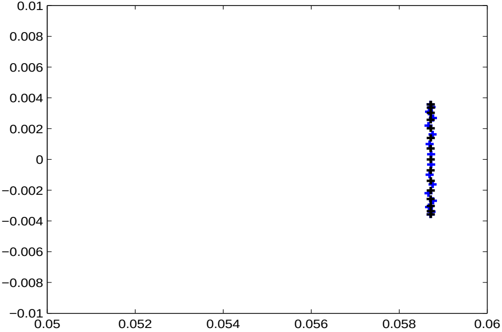
Blue: Eigenvalues of $A \in \mathbb{R}^{16\times 16}$. Black: Ritz values at iteration 2.

Some recent results



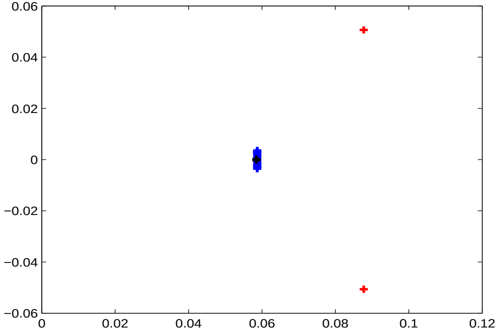
Blue: Eigenvalues of $A \in \mathbb{R}^{16\times 16}$. Black: Ritz values at iteration 7.

Some recent results



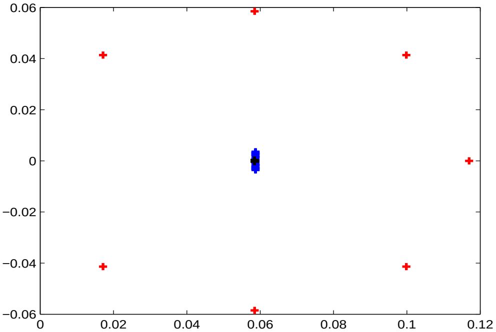
Blue: Eigenvalues of $A \in \mathbb{R}^{16\times 16}$. Black: Ritz values at iteration 15.

Some recent results



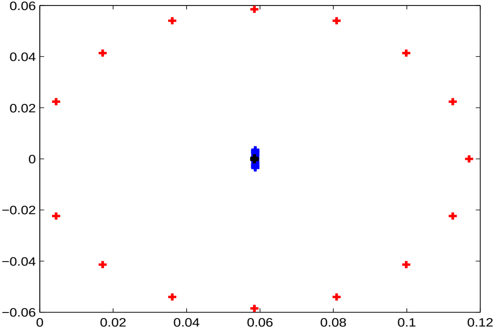
Blue: Eigenvalues of $A \in \mathbb{R}^{16\times 16}$. Black: Ritz values at iteration 2. Red: Harmonic Ritz values at iteration 2.

Some recent results



Blue: Eigenvalues of $A \in \mathbb{R}^{16\times 16}$. Black: Ritz values at iteration 7. Red: Harmonic Ritz values at iteration 7.

Some recent results



Blue: Eigenvalues of $A \in \mathbb{R}^{16\times 16}$. Black: Ritz values at iteration 15.
Red: Harmonic Ritz values at iteration 15.

Outline

Harmonic Ritz values

What do we know about the relation between harmonic Ritz value convergence and GMRES convergence ?

Harmonic Ritz values

What do we know about the relation between harmonic Ritz value convergence and GMRES convergence ?

Of course, harmonic Ritz values are the roots of GMRES polynomials:

Harmonic Ritz values

What do we know about the relation between harmonic Ritz value convergence and GMRES convergence ?

Of course, harmonic Ritz values are the roots of GMRES polynomials:

If at the kth iteration, GMRES generates residual vector r_k and harmonic Ritz values $\theta_1^{(k)}, \ldots, \theta_k^{(k)}$, then

$$\|r_k\| = \|p_k(A)b\|,$$

where

$$p_k(z) = \prod_{i=1}^{k} \left(1 - \frac{z}{\theta_i^{(k)}}\right).$$

Harmonic Ritz values

What do we know about the relation between harmonic Ritz value convergence and GMRES convergence ?

Of course, harmonic Ritz values are the roots of GMRES polynomials:

If at the kth iteration, GMRES generates residual vector r_k and harmonic Ritz values $\theta_1^{(k)}, \ldots, \theta_k^{(k)}$, then

$$\|r_k\| = \|p_k(A)b\|,$$

where

$$p_k(z) = \prod_{i=1}^{k} \left(1 - \frac{z}{\theta_i^{(k)}}\right).$$

Thus a close relation might be expected.

Harmonic Ritz values

First, we need to characterize admissible harmonic Ritz value sets.

Harmonic Ritz values

First, we need to characterize admissible harmonic Ritz value sets.

Theorem ([Du, DT & Meurant, 2017?])

Let $\Theta^{(k)}$ *denote the* k*-tuple of the harmonic Ritz values at step* k*:*

$$\Theta^{(k)} = (\theta_1^{(k)}, \theta_2^{(k)}, \ldots, \theta_k^{(k)}).$$

If GMRES stagnates from step $k+1$ *to step* $k+m$ *, i.e.,*

$$\|\mathbf{r}_k\| = \|\mathbf{r}_{k+1}\| = \cdots = \|\mathbf{r}_{k+m}\|,$$

then, for $i = 1 : m$*, the* $(k+i)$*-tuple of the harmonic Ritz values at step* $k+i$ *is*

$$\Theta^{(k+i)} = (\theta_1^{(k)}, \theta_2^{(k)}, \ldots, \theta_k^{(k)}, \infty, \cdots, \infty).$$

Harmonic Ritz values

First, we need to characterize admissible harmonic Ritz value sets.

Theorem ([Du, DT & Meurant, 2017?])

Let $\Theta^{(k)}$ *denote the* k*-tuple of the harmonic Ritz values at step* k*:*

$$\Theta^{(k)} = (\theta_1^{(k)}, \theta_2^{(k)}, \ldots, \theta_k^{(k)}).$$

If GMRES stagnates from step $k+1$ *to step* $k+m$ *, i.e.,*

$$\|\mathbf{r}_k\| = \|\mathbf{r}_{k+1}\| = \cdots = \|\mathbf{r}_{k+m}\|,$$

then, for $i = 1 : m$*, the* $(k+i)$*-tuple of the harmonic Ritz values at step* $k+i$ *is*

$$\Theta^{(k+i)} = (\theta_1^{(k)}, \theta_2^{(k)}, \ldots, \theta_k^{(k)}, \infty, \cdots, \infty).$$

Proof: Follows from $p_k(z) = p_{k+1}(z) = \cdots = p_{k+m}(z)$.

Harmonic Ritz values

If H_k is the kth leading principal submatrix of H, the harmonic Ritz values are the eigenvalues of

$$\hat{H}_k = H_k + h_{k+1,k}^2 H_k^{-*} e_k e_k^T.$$

Harmonic Ritz values

If H_k is the kth leading principal submatrix of H, the harmonic Ritz values are the eigenvalues of

$$\hat{H}_k = H_k + h_{k+1,k}^2 H_k^{-*} e_k e_k^T.$$

H_k being upper Hessenberg, it can be decomposed as

$$H_k = U_k^{-1} C^{(k)} U_k,$$

where

Harmonic Ritz values

If H_k is the kth leading principal submatrix of H, the harmonic Ritz values are the eigenvalues of

$$\hat{H}_k = H_k + h_{k+1,k}^2 H_k^{-*} e_k e_k^T.$$

H_k being upper Hessenberg, it can be decomposed as

$$H_k = U_k^{-1} C^{(k)} U_k,$$

where

- $C^{(k)}$ is the companion matrix corresponding to the *ordinary* Ritz values in the kth iteration of GMRES.

Harmonic Ritz values

If H_k is the kth leading principal submatrix of H, the harmonic Ritz values are the eigenvalues of

$$\hat{H}_k = H_k + h_{k+1,k}^2 H_k^{-*} e_k e_k^T.$$

H_k being upper Hessenberg, it can be decomposed as

$$H_k = U_k^{-1} C^{(k)} U_k,$$

where

- $C^{(k)}$ is the companion matrix corresponding to the *ordinary* Ritz values in the kth iteration of GMRES.
- U_k is the kth leading principal submatrix of U in

$$H = U^{-1} C U.$$

Harmonic Ritz values

We have the following relation between the companion matrices for H_k and $\hat{H}_k$:

Theorem ([Du, DT & Meurant, 2017?])

Let $H_k = U_k^{-1} C^{(k)} U_k$ *be non-singular. The matrix*

$$\hat{H}_k = H_k + h_{k+1,k}^2 H_k^{-*} e_k e_k^T$$

whose eigenvalues are the harmonic Ritz values at step k *can be written as*

$$U_k^{-1} \hat{C}^{(k)} U_k,$$

where

$$\hat{C}^{(k)} = C^{(k)} + \frac{1}{u_{k+1,k+1}^2 \overline{e_1^T C^{(k)} e_k}} U_k U_k^* e_1 e_k^T.$$

Harmonic Ritz values

We have the following relation between the companion matrices for H_k and $\hat{H}_k$:

Theorem ([Du, DT & Meurant, 2017?])

Let $H_k = U_k^{-1} C^{(k)} U_k$ *be non-singular. The matrix*

$$\hat{H}_k = H_k + h_{k+1,k}^2 H_k^{-*} e_k e_k^T$$

whose eigenvalues are the harmonic Ritz values at step k *can be written as*

$$U_k^{-1} \hat{C}^{(k)} U_k,$$

where

$$\hat{C}^{(k)} = C^{(k)} + \frac{1}{u_{k+1,k+1}^2 \overline{e_1^T C^{(k)} e_k}} U_k U_k^* e_1 e_k^T.$$

Thus with $\hat{C}^{(k)}$ prescribed, we can attempt to construct $U^{(k)}$ while keeping the first row of U fixed.

Harmonic Ritz values

The construction of the upper triangular matrix U can be done as follows:

Harmonic Ritz values

The construction of the upper triangular matrix U can be done as follows:

Given $f(0) \geq \cdots \geq f(n-1) > 0$ and an admissible harmonic Ritz value set $\Theta = \{\Theta^{(1)},\ \Theta^{(2)},\ \ldots,\ \Theta^{(n)}\}$,

Harmonic Ritz values

The construction of the upper triangular matrix U can be done as follows:

Given $f(0) \geq \cdots \geq f(n-1) > 0$ and an admissible harmonic Ritz value set $\Theta = \{\Theta^{(1)},\ \Theta^{(2)},\ \ldots,\ \Theta^{(n)}\}$,

1. $u_{1,1} = 1$
2. If $f(k) < f(k-1)$, let $\theta_1^{(k)}, \ldots, \theta_k^{(k)}$ be the roots of the polynomial $z^k + \beta_{k-1} z^{k-1} + \cdots + \beta_1 z + \beta_0$.

Harmonic Ritz values

The construction of the upper triangular matrix U can be done as follows:

Given $f(0) \geq \cdots \geq f(n-1) > 0$ and an admissible harmonic Ritz value set $\Theta = \{\Theta^{(1)},\ \Theta^{(2)},\ \ldots,\ \Theta^{(n)}\}$,

1. $u_{1,1} = 1$
2. If $f(k) < f(k-1)$, let $\theta_1^{(k)}, \ldots, \theta_k^{(k)}$ be the roots of the polynomial $z^k + \beta_{k-1} z^{k-1} + \cdots + \beta_1 z + \beta_0$. Then put

$$
\begin{aligned}
u_{1,k+1} &= \frac{\beta_0}{|\beta_0|} \sqrt{1/f(k)^2 - 1/f(k-1)^2}, \\
u_{k+1,k+1} &= \frac{1/f(k)^2 - 1/f(k-1)^2 + e_1^T U_k U_k^* e_1}{|\beta_0| \sqrt{1/f(k)^2 - 1/f(k-1)^2}}, \\
u_{j,k+1} &= \beta_{j-1} u_{k+1,k+1} - \frac{e_j^T U_k U_k^* e_1}{\bar{u}_{1,k+1}}, \quad j = 2, \ldots, k.
\end{aligned}
$$

Harmonic Ritz values

The construction of the upper triangular matrix U can be done as follows:

Given $f(0) \geq \cdots \geq f(n-1) > 0$ and an admissible harmonic Ritz value set $\Theta = \{\Theta^{(1)},\ \Theta^{(2)},\ \ldots,\ \Theta^{(n)}\}$,

1. $u_{1,1} = 1$
2. If $f(k) < f(k-1)$, let $\theta_1^{(k)}, \ldots, \theta_k^{(k)}$ be the roots of the polynomial $z^k + \beta_{k-1} z^{k-1} + \cdots + \beta_1 z + \beta_0$. Then put
$$\begin{aligned} u_{1,k+1} &= \frac{\beta_0}{|\beta_0|} \sqrt{1/f(k)^2 - 1/f(k-1)^2}, \\ u_{k+1,k+1} &= \frac{1/f(k)^2 - 1/f(k-1)^2 + e_1^T U_k U_k^* e_1}{|\beta_0| \sqrt{1/f(k)^2 - 1/f(k-1)^2}}, \\ u_{j,k+1} &= \beta_{j-1} u_{k+1,k+1} - \frac{e_j^T U_k U_k^* e_1}{\bar{u}_{1,k+1}}, \quad j = 2, \ldots, k. \end{aligned}$$
3. If $f(k) = f(k-1)$, let $u_{1,k+1} = 0$, $u_{k+1,k+1} > 0$ and $u_{j,k+1}$, $j = 2, \ldots, k$ arbitrary complex.

Conclusion

Conclusion: Any GMRES residual norm history is possible with any admissible harmonic Ritz values.

Conclusion

Conclusion: Any GMRES residual norm history is possible with any admissible harmonic Ritz values.

Future work: Attempt to find theoretical reasons for the fact that deflation methods work in spite of these results.

Thank you for your attention.

Related publications

- A. Greenbaum, V. Pták and Z. Strakoš, Any nonincreasing convergence curve is possible for GMRES, **SIAM J. Matrix Anal. Appl.**, 17 (1996), pp. 465–469.
- M. Arioli, V. Pták and Z. Strakoš, Krylov sequences of maximal length and convergence of GMRES, **BIT Num. Maths.**, 38 (1996), 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.**, 33 (2012), pp. 958–978.
- J. Duintjer Tebbens and G. Meurant, Prescribing the behavior of early terminating GMRES and Arnoldi iterations, **Numer. Algorithms**, 65 (2014), pp. 69–90.
- J. Duintjer Tebbens, G. Meurant, H. Sadok and Z. Strakoš, On investigating GMRES convergence using unitary matrices, **Lin. Alg. Appl.**, 450 (2014), pp. 83–107.
- G. Meurant and J. Duintjer Tebbens, The role eigenvalues play in forming GMRES residual norms with non-normal matrices, **Numer. Algorithms**, 68 (2015), pp. 143–165.
- J. Duintjer Tebbens and G. Meurant, On the convergence of QOR and QMR Krylov methods for solving nonsymmetric linear systems, **BIT Num. Maths.**, 56 (2016), pp. 77–97.