On admissible eigenvalue approximations from Krylov subspace methods for non-normal matrices

Jurjen Duintjer Tebbens

joint work with

Gérard Meurant

Crouzeix's conjecture workshop
AIM, San Jose, US, August 3, 2017.

Outline

Outline

Introduction

This talk is about some iterative methods designed for large, sparse matrices to find eigenvalues, eigenvectors and solutions of linear systems.

Introduction

This talk is about some iterative methods designed for large, sparse matrices to find eigenvalues, eigenvectors and solutions of linear systems.

More precisely, it is about Krylov subspace methods generating during the kth iteration the kth Krylov subspace

$$\mathcal{K}_k(A, v) \equiv \text{span}\{v, Av, \ldots, A^{k-1}v\}$$

for a given a nonsingular matrix and nonzero vector

$$A \in \mathbb{C}^{n\times n}, \quad v \in \mathbb{C}^n \quad (\text{w.l.o.g.} \quad \|v\| = 1).$$

Introduction

This talk is about some iterative methods designed for large, sparse matrices to find eigenvalues, eigenvectors and solutions of linear systems.

More precisely, it is about Krylov subspace methods generating during the kth iteration the kth Krylov subspace

$$\mathcal{K}_k(A, v) \equiv \text{span}\{v, Av, \ldots, A^{k-1}v\}$$

for a given a nonsingular matrix and nonzero vector

$$A \in \mathbb{C}^{n\times n}, \quad v \in \mathbb{C}^n \quad (\text{w.l.o.g.} \quad \|v\| = 1).$$

Note that

- all vectors $w \in \mathcal{K}_k(A, v) = \text{span}\{v, Av, ..., A^{k-1}v\}$ are of the form

  $$w = \alpha_0 v + \alpha_1 Av + \cdots + \alpha_{k-1}A^{k-1}v = p_{k-1}(A)v$$

  for a polynomial p_{k-1} of degree $k-1$.

Introduction

This talk is about some iterative methods designed for large, sparse matrices to find eigenvalues, eigenvectors and solutions of linear systems.

More precisely, it is about Krylov subspace methods generating during the kth iteration the kth Krylov subspace

$$\mathcal{K}_k(A,v) \equiv \text{span}\{v, Av, \ldots, A^{k-1}v\}$$

for a given a nonsingular matrix and nonzero vector

$$A \in \mathbb{C}^{n\times n}, \quad v \in \mathbb{C}^n \quad (\text{w.l.o.g.} \quad \|v\| = 1).$$

Note that

- all vectors $w \in \mathcal{K}_k(A,v) = \text{span}\{v, Av, ..., A^{k-1}v\}$ are of the form

  $$w = \alpha_0 v + \alpha_1 Av + \cdots + \alpha_{k-1}A^{k-1}v = p_{k-1}(A)v$$

  for a polynomial p_{k-1} of degree $k-1$.
- we always consider both the matrix A *and* a starting vector v and their interplay can be crucial (also if we are in properties like the field of values or matrix polynomials which depend on the matrix only).

Introduction

This talk is about some iterative methods designed for large, sparse matrices to find eigenvalues, eigenvectors and solutions of linear systems.

More precisely, it is about Krylov subspace methods generating during the kth iteration the kth Krylov subspace

$$\mathcal{K}_k(A, v) \equiv \text{span}\{v, Av, \ldots, A^{k-1}v\}$$

for a given a nonsingular matrix and nonzero vector

$$A \in \mathbb{C}^{n\times n}, \quad v \in \mathbb{C}^n \quad (\text{w.l.o.g.} \quad \|v\| = 1).$$

Note that

- all vectors $w \in \mathcal{K}_k(A, v) = \text{span}\{v, Av, ..., A^{k-1}v\}$ are of the form
  $$w = \alpha_0 v + \alpha_1 Av + \cdots + \alpha_{k-1}A^{k-1}v = p_{k-1}(A)v$$
  for a polynomial p_{k-1} of degree $k - 1$.
- we always consider both the matrix A *and* a starting vector v and their interplay can be crucial (also if we are in properties like the field of values or matrix polynomials which depend on the matrix only).

We will restrict ourselves to non-normal matrices (for which Crouzeix's conjecture has not yet been proved).

Introduction

With non-normal matrices, we can very roughly divide the available methods into those using short recurrences (lower compuational costs) and those using long recurrences (enhanced stability properties).

Introduction

With non-normal matrices, we can very roughly divide the available methods into those using short recurrences (lower compuational costs) and those using long recurrences (enhanced stability properties).

We focus on the last class which is based on the Arnoldi orthogonalization process [Arnoldi - 1951] for computing orthogonal bases of the Krylov subspaces.

Introduction

With non-normal matrices, we can very roughly divide the available methods into those using short recurrences (lower compuational costs) and those using long recurrences (enhanced stability properties).

We focus on the last class which is based on the Arnoldi orthogonalization process [Arnoldi - 1951] for computing orthogonal bases of the Krylov subspaces.

In the kth iteration of the process (without breakdown) it computes the decomposition

$$AV_k = V_{k+1}\tilde{H}_k,$$

where the columns of $V_k = [v_1, \dots, v_k]$ (the Arnoldi vectors) contain an orthogonal basis for the kth Krylov subspace $\mathcal{K}_k(A, v)$ and $\tilde{H}_k$ is rectangular upper Hessenberg.

Introduction

With non-normal matrices, we can very roughly divide the available methods into those using short recurrences (lower compuational costs) and those using long recurrences (enhanced stability properties).

We focus on the last class which is based on the Arnoldi orthogonalization process [Arnoldi - 1951] for computing orthogonal bases of the Krylov subspaces.

In the kth iteration of the process (without breakdown) it computes the decomposition

$$AV_k = V_{k+1}\tilde{H}_k,$$

where the columns of $V_k = [v_1, \ldots, v_k]$ (the Arnoldi vectors) contain an orthogonal basis for the kth Krylov subspace $\mathcal{K}_k(A, v)$ and $\tilde{H}_k$ is rectangular upper Hessenberg.

By deleting the last row we get the square matrix

$$H_k = V_k^* A V_k \in \mathbb{C}^{k\times k};$$

H_k is the orthogonal restriction of A onto $\mathcal{K}_k(A, v)$ and A is a dilation of H_k.

Introduction

Essentially,

- for eigenpair approximations of A, the Arnoldi method [Arnoldi - 1951], [Saad - 1980] uses the eigenvalues and eigenvectors of

$$H_k$$

and the first k Arnoldi vectors. The eigenvalue approximations are called Ritz values, the eigenvector approximations Ritz vectors.

Introduction

Essentially,

- for eigenpair approximations of A, the Arnoldi method [Arnoldi - 1951], [Saad - 1980] uses the eigenvalues and eigenvectors of

$$H_k$$

and the first k Arnoldi vectors. The eigenvalue approximations are called Ritz values, the eigenvector approximations Ritz vectors.

- for approximate solutions to linear systems $Ax = b$, the GMRES method [Saad, Schultz - 1986] solves least squares problems

$$\min_z \| \|b\| e_1 - \tilde{H}_k z \|$$

and the first k Arnoldi vectors.

Introduction

Essentially,

- for eigenpair approximations of A, the Arnoldi method [Arnoldi - 1951], [Saad - 1980] uses the eigenvalues and eigenvectors of

$$H_k$$

  and the first k Arnoldi vectors. The eigenvalue approximations are called Ritz values, the eigenvector approximations Ritz vectors.

- for approximate solutions to linear systems $Ax = b$, the GMRES method [Saad, Schultz - 1986] solves least squares problems

$$\min_z \| \|b\| e_1 - \tilde{H}_k z \|$$

  and the first k Arnoldi vectors.

- Both the GMRES and the Arnoldi method are very popular methods that are successful for a large variety of problem classes.

Introduction

Essentially,

- for eigenpair approximations of A, the Arnoldi method [Arnoldi - 1951], [Saad - 1980] uses the eigenvalues and eigenvectors of

$$H_k$$

  and the first k Arnoldi vectors. The eigenvalue approximations are called Ritz values, the eigenvector approximations Ritz vectors.

- for approximate solutions to linear systems $Ax = b$, the GMRES method [Saad, Schultz - 1986] solves least squares problems

$$\min_z \| \|b\| e_1 - \tilde{H}_k z \|$$

  and the first k Arnoldi vectors.

- Both the GMRES and the Arnoldi method are very popular methods that are successful for a large variety of problem classes.
- Nevertheless, convergence behavior of the two methods is not fully understood, analysis is particularly challenging with highly non-normal input matrices.

Introduction

- Often one tries to use the tools that are successful for analysis of hermitian counterparts of GMRES and Arnoldi like the Conjugate Gradients and the Lanczos method.

Introduction

- Often one tries to use the tools that are successful for analysis of hermitian counterparts of GMRES and Arnoldi like the Conjugate Gradients and the Lanczos method.

- For example, the basic tool for explaining Krylov subspace methods for hermitian linear systems is the eigenvalue distribution.

Introduction

- Often one tries to use the tools that are successful for analysis of hermitian counterparts of GMRES and Arnoldi like the Conjugate Gradients and the Lanczos method.
- For example, the basic tool for explaining Krylov subspace methods for hermitian linear systems is the eigenvalue distribution.
- However, for GMRES it is known for some time that if GMRES generates a certain residual norm history, the same history can be generated with any nonzero spectrum [Greenbaum, Strakoš - 1994].

Introduction

- Often one tries to use the tools that are successful for analysis of hermitian counterparts of GMRES and Arnoldi like the Conjugate Gradients and the Lanczos method.
- For example, the basic tool for explaining Krylov subspace methods for hermitian linear systems is the eigenvalue distribution.
- However, for GMRES it is known for some time that if GMRES generates a certain residual norm history, the same history can be generated with any nonzero spectrum [Greenbaum, Strakoš - 1994].
- Complemented with the fact that GMRES can generate arbitrary non-increasing residual norms, this gives the result that any non-increasing convergence curve is possible with any nonzero spectrum [Greenbaum , Pták, Strakoš - 1996].

Introduction

- Often one tries to use the tools that are successful for analysis of hermitian counterparts of GMRES and Arnoldi like the Conjugate Gradients and the Lanczos method.
- For example, the basic tool for explaining Krylov subspace methods for hermitian linear systems is the eigenvalue distribution.
- However, for GMRES it is known for some time that if GMRES generates a certain residual norm history, the same history can be generated with any nonzero spectrum [Greenbaum, Strakoš - 1994].
- Complemented with the fact that GMRES can generate arbitrary non-increasing residual norms, this gives the result that any non-increasing convergence curve is possible with any nonzero spectrum [Greenbaum , Pták, Strakoš - 1996].
- A complete description of the class of matrices and right hand sides with prescribed convergence and eigenvalues was given in [Arioli , Pták, Strakoš - 1998].

Convergence analysis for Krylov subspace methods

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

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

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

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

be any set of nonzero complex numbers.

Convergence analysis for Krylov subspace methods

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

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

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

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

Convergence analysis for Krylov subspace methods

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

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

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

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

For other popular methods like Bi-CG and QMR, it was proved as well that convergence behavior can be arbitrarily poor, independent from the eigenvalue distribution [D.T. & Meurant, 2016].

Introduction

Other objects have been successful in explaining GMRES for particular problems, including:

Introduction

Other objects have been successful in explaining GMRES for particular problems, including:

- The pseudo-spectrum, see e.g. [Trefethen, Embree - 2005],

Introduction

Other objects have been successful in explaining GMRES for particular problems, including:

- The pseudo-spectrum, see e.g. [Trefethen, Embree - 2005],
- the field of values, see e.g. [Eiermann - 1993],

Introduction

Other objects have been successful in explaining GMRES for particular problems, including:

- The pseudo-spectrum, see e.g. [Trefethen, Embree - 2005],
- the field of values, see e.g. [Eiermann - 1993],
- the numerical polynomial hull, see e.g. [Greenbaum - 2002],

Introduction

Other objects have been successful in explaining GMRES for particular problems, including:

- The pseudo-spectrum, see e.g. [Trefethen, Embree - 2005],
- the field of values, see e.g. [Eiermann - 1993],
- the numerical polynomial hull, see e.g. [Greenbaum - 2002],
- the Ritz values, i.e. the eigenvalues of the Hessenberg matrices generated by the underlying Arnoldi process, see e.g. [van der Vorst, Vuik - 1993].

Introduction

Other objects have been successful in explaining GMRES for particular problems, including:

- The pseudo-spectrum, see e.g. [Trefethen, Embree - 2005],
- the field of values, see e.g. [Eiermann - 1993],
- the numerical polynomial hull, see e.g. [Greenbaum - 2002],
- the Ritz values, i.e. the eigenvalues of the Hessenberg matrices generated by the underlying Arnoldi process, see e.g. [van der Vorst, Vuik - 1993].

Although in practice eigenvalues do often influence convergence of GMRES, they cannot be used as a universal tool for explaining GMRES and such a tool is unlikely to exist.

Introduction

An important tool for hermitian eigenproblems solved with Krylov subspace methods is the following interlacing property:

Introduction

An important tool for hermitian eigenproblems solved with Krylov subspace methods is the following interlacing property:

Consider a tridiagonal Jacobi matrix T_m and its leading principal submatrix T_k for some $k < m$. If the ordered eigenvalues of T_k are

$$\rho_1^{(k)} < \rho_2^{(k)} < \quad \dots \quad < \rho_k^{(k)},$$

then in every open interval between two subsequent eigenvalues

$$(\rho_{i-1}^{(k)}, \rho_i^{(k)}), \quad i = 2, \dots, k,$$

there lies at least one eigenvalue of T_m.

Introduction

An important tool for hermitian eigenproblems solved with Krylov subspace methods is the following interlacing property:

Consider a tridiagonal Jacobi matrix T_m and its leading principal submatrix T_k for some $k < m$. If the ordered eigenvalues of T_k are

$$\rho_1^{(k)} < \rho_2^{(k)} < \quad \dots \quad < \rho_k^{(k)},$$

then in every open interval between two subsequent eigenvalues

$$(\rho_{i-1}^{(k)}, \rho_i^{(k)}), \quad i = 2, \dots, k,$$

there lies at least one eigenvalue of T_m.

This interlacing property enables, among others, to prove the persistence theorem (see [Paige - 1971, 1976, 1980] or [Meurant, Strakoš - 2006]) which is crucial for controlling the convergence of Ritz values in the Lanczos method.

Introduction

- There are generalizations of the interlacing property to the non-hermitian but normal case [Fan, Pall - 1957], [Thompson - 1966], [Ericsson - 1990], [Malamud - 2005], [Carden - 2013].

Introduction

- There are generalizations of the interlacing property to the non-hermitian but normal case [Fan, Pall - 1957], [Thompson - 1966], [Ericsson - 1990], [Malamud - 2005], [Carden - 2013].

- There is no interlacing property for the principal submatrices of general non-normal matrices [de Oliveira - 1969], [Shomron, Parlett - 2009].

Introduction

- There are generalizations of the interlacing property to the non-hermitian but normal case [Fan, Pall - 1957], [Thompson - 1966], [Ericsson - 1990], [Malamud - 2005], [Carden - 2013].

- There is no interlacing property for the principal submatrices of general non-normal matrices [de Oliveira - 1969], [Shomron, Parlett - 2009].

- This makes convergence analysis of the Arnoldi method for non-normal input matrices rather delicate, just as it is for the GMRES method.

Introduction

- There are generalizations of the interlacing property to the non-hermitian but normal case [Fan, Pall - 1957], [Thompson - 1966], [Ericsson - 1990], [Malamud - 2005], [Carden - 2013].
- There is no interlacing property for the principal submatrices of general non-normal matrices [de Oliveira - 1969], [Shomron, Parlett - 2009].
- This makes convergence analysis of the Arnoldi method for non-normal input matrices rather delicate, just as it is for the GMRES method.
- The GMRES and Arnoldi methods being closely related through the Arnoldi process, can we show that arbitrary convergence behavior of Arnoldi is possible?

Introduction

- There are generalizations of the interlacing property to the non-hermitian but normal case [Fan, Pall - 1957], [Thompson - 1966], [Ericsson - 1990], [Malamud - 2005], [Carden - 2013].

- There is no interlacing property for the principal submatrices of general non-normal matrices [de Oliveira - 1969], [Shomron, Parlett - 2009].

- This makes convergence analysis of the Arnoldi method for non-normal input matrices rather delicate, just as it is for the GMRES method.

- The GMRES and Arnoldi methods being closely related through the Arnoldi process, can we show that arbitrary convergence behavior of Arnoldi is possible?

- By arbitrary behavior we mean arbitrary Ritz values for *all* iterations (we do not consider eigenvectors). Note that this involves many more conditions than prescribing one residual norm per GMRES iteration.

2. Prescribed convergence for Arnoldi's method

Notation: Let the kth Hessenberg matrix H_k generated in Arnoldi's method have the eigenvalue ρ and eigenvector y,

$$H_k\, y = \rho\, y.$$

Then

2. Prescribed convergence for Arnoldi's method

Notation: Let the kth Hessenberg matrix H_k generated in Arnoldi's method have the eigenvalue ρ and eigenvector y,

$$H_k \, y = \rho \, y.$$

Then

- ρ is a Ritz value for A
- $V_k y$ is a Ritz vector for A

2. Prescribed convergence for Arnoldi's method

Notation: Let the kth Hessenberg matrix H_k generated in Arnoldi's method have the eigenvalue ρ and eigenvector y,

$$H_k\, y = \rho\, y.$$

Then

- ρ is a Ritz value for A
- $V_k y$ is a Ritz vector for A

With the Arnoldi decomposition $AV_k = V_{k+1}\tilde{H}_k$, we obtain for the Ritz-value Ritz-vector pair $\{\rho, V_k y\}$ the residual norm:

$$\|A(V_k y) - \rho(V_k y)\| = \|A(V_k y) - V_k H_k y\| = \|V_{k+1}\tilde{H}_k y - V_k H_k y\| = h_{k+1,k}|e_k^T y|.$$

2. Prescribed convergence for Arnoldi's method

Notation: Let the kth Hessenberg matrix H_k generated in Arnoldi's method have the eigenvalue ρ and eigenvector y,

$$H_k\, y = \rho\, y.$$

Then

- ρ is a Ritz value for A
- $V_k y$ is a Ritz vector for A

With the Arnoldi decomposition $AV_k = V_{k+1}\tilde{H}_k$, we obtain for the Ritz-value Ritz-vector pair $\{\rho, V_k y\}$ the residual norm:

$$\|A(V_k y) - \rho(V_k y)\| = \|A(V_k y) - V_k H_k y\| = \|V_{k+1}\tilde{H}_k y - V_k H_k y\| = h_{k+1,k}|e_k^T y|.$$

Often for small $h_{k+1,k}|e_k^T y|$, the Arnoldi method takes $\{\rho, V_k y\}$ as an approximate eigenvalue-eigenvector pair of A. Note that a small value $h_{k+1,k}|e_k^T y|$ needs not imply that ρ is close to a true eigenvalue of A, see e.g. [Chatelin - 1993], [Godet-Thobie - 1993]; convergence analysis cannot be based on this value but focusses instead on the quality of approximate invariant subspaces [Beattie, Embree, Sorensen - 2005].

2. Prescribed convergence for Arnoldi's method

Theorem 1 [DT, Meurant - 2012]. Let the set

$$
\mathcal{R} = \{ \quad \begin{array}{l} \rho_1^{(1)}, \\ (\rho_1^{(2)}, \rho_2^{(2)}), \\ \vdots \\ (\rho_1^{(n-1)}, \ldots, \rho_{n-1}^{(n-1)}), \\ (\lambda_1, \ldots\ldots\ldots, \lambda_n) \}, \end{array}
$$

represent any choice of $n(n+1)/2$ complex Ritz values

2. Prescribed convergence for Arnoldi's method

Theorem 1 [DT, Meurant - 2012]. Let the set

$$\begin{aligned}\mathcal{R} = \{ \quad & \rho_1^{(1)}, \\ & (\rho_1^{(2)}, \rho_2^{(2)}), \\ & \vdots \\ & (\rho_1^{(n-1)}, \dots, \rho_{n-1}^{(n-1)}), \\ & (\lambda_1, \dots\dots\dots, \lambda_n) \},\end{aligned}$$

represent any choice of $n(n+1)/2$ complex Ritz values and denote by $C^{(k)}$ the companion matrix of the polynomial with roots $\rho_1^{(k)}, \dots, \rho_k^{(k)}$, i.e.

$$C^{(k)} = \begin{pmatrix} 0 & \dots & & 0 & -\alpha_0 \\ 1 & 0 & \dots & 0 & -\alpha_1 \\ & \ddots & & \vdots & \vdots \\ & & & 1 & -\alpha_{k-1} \end{pmatrix}, \qquad \prod_{j=1}^{k}(z - \rho_j^{(k)}) = z^k + \sum_{j=0}^{k-1} \alpha_j z^j.$$

2. Prescribed convergence for Arnoldi's method

If we define the unit upper triangular matrix $U(\mathcal{S})$ through

$$U(\mathcal{S}) = I_n - \left[\begin{array}{cccc} 0 & C^{(1)}e_1 & \vdots & \vdots \\ & 0 & C^{(2)}e_2 & \vdots \\ & & 0 & C^{(n-1)}e_{n-1} \\ & & & 0 \end{array} \right],$$

2. Prescribed convergence for Arnoldi's method

If we define the unit upper triangular matrix $U(\mathcal{S})$ through

$$U(\mathcal{S}) = I_n - \begin{bmatrix} 0 & C^{(1)}e_1 & \vdots & & \vdots \\ & 0 & C^{(2)}e_2 & & \vdots \\ & & 0 & & \\ & & & & C^{(n-1)}e_{n-1} \\ & & & & 0 \end{bmatrix},$$

then the upper Hessenberg matrix

$$H(\mathcal{R}) = U(\mathcal{S})^{-1}C^{(n)}U(\mathcal{S})$$

has the spectrum $\lambda_1, \ldots, \lambda_n$ and its kth leading principal submatrix has spectrum

$$\rho_1^{(k)}, \ldots, \rho_k^{(k)}, \qquad k = 1, \ldots, n-1.$$

It has unit subdiagonal.

2. Prescribed convergence for Arnoldi's method

Proof: The $k \times k$ leading principal submatrix of $H(\mathcal{R})$ is

$$[I_k, 0]\, H(\mathcal{R}) \begin{bmatrix} I_k \\ 0 \end{bmatrix} = [I_k, 0]\, U(\mathcal{S})^{-1} C^{(n)} U(\mathcal{S}) \begin{bmatrix} I_k \\ 0 \end{bmatrix}$$

$$= \quad [U_k^{-1}, \tilde{u}_{k+1}, \ldots, \tilde{u}_n] \begin{bmatrix} 0 \\ U_k \\ 0 \end{bmatrix} = [U_k^{-1}, \tilde{u}_{k+1}] \begin{bmatrix} 0 \\ U_k \end{bmatrix},$$

where U_k denotes the $k \times k$ leading principal submatrix of $U(\mathcal{S})$ and $\tilde{u}_j$ denotes the vector of the first k entries of the jth column of $U(\mathcal{S})^{-1}$ for $j > k$.

Proof: The $k \times k$ leading principal submatrix of $H(\mathcal{R})$ is

$$[I_k, 0]\, H(\mathcal{R}) \begin{bmatrix} I_k \\ 0 \end{bmatrix} = [I_k, 0]\, U(\mathcal{S})^{-1} C^{(n)} U(\mathcal{S}) \begin{bmatrix} I_k \\ 0 \end{bmatrix}$$

$$= \quad [U_k^{-1}, \tilde{u}_{k+1}, \ldots, \tilde{u}_n] \begin{bmatrix} 0 \\ U_k \\ 0 \end{bmatrix} = [U_k^{-1}, \tilde{u}_{k+1}] \begin{bmatrix} 0 \\ U_k \end{bmatrix},$$

where U_k denotes the $k \times k$ leading principal submatrix of $U(\mathcal{S})$ and $\tilde{u}_j$ denotes the vector of the first k entries of the jth column of $U(\mathcal{S})^{-1}$ for $j > k$.

Its spectrum is also the spectrum of the matrix

$$U_k [U_k^{-1}, \tilde{u}_{k+1}] \begin{bmatrix} 0 \\ U_k \end{bmatrix} U_k^{-1} = [I_k, U_k \tilde{u}_{k+1}] \begin{bmatrix} 0 \\ I_k \end{bmatrix},$$

which is a companion matrix with last column $U_k \tilde{u}_{k+1}$.

2. Prescribed convergence for Arnoldi's method

From

$$e_{k+1} = U_{k+1}U_{k+1}^{-1}e_{k+1} = \begin{bmatrix} U_k & -C^{(k)}e_k \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \tilde{u}_{k+1} \\ 1 \end{bmatrix} = \begin{bmatrix} U_k\tilde{u}_{k+1} - C^{(k)}e_k \\ 1 \end{bmatrix}$$

we obtain $U_k\tilde{u}_{k+1} = C^{(k)}e_k$. $\square$

From

$$e_{k+1} = U_{k+1}U_{k+1}^{-1}e_{k+1} = \begin{bmatrix} U_k & -C^{(k)}e_k \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \tilde{u}_{k+1} \\ 1 \end{bmatrix} = \begin{bmatrix} U_k\tilde{u}_{k+1} - C^{(k)}e_k \\ 1 \end{bmatrix}$$

we obtain $U_k\tilde{u}_{k+1} = C^{(k)}e_k$. $\square$

Remark: The matrix

$$H(\mathcal{R}) = U(\mathcal{S})^{-1}C^{(n)}U(\mathcal{S}).$$

is the *unique* upper Hessenberg matrix $H(\mathcal{R})$ with the prescribed spectrum and Ritz values and the entry one along the subdiagonal (see also [Parlett, Strang - 2008] where $H(\mathcal{R})$ is constructed in a different way).

2. Prescribed convergence for Arnoldi's method

From

$$e_{k+1} = U_{k+1}U_{k+1}^{-1}e_{k+1} = \begin{bmatrix} U_k & -C^{(k)}e_k \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \tilde{u}_{k+1} \\ 1 \end{bmatrix} = \begin{bmatrix} U_k\tilde{u}_{k+1} - C^{(k)}e_k \\ 1 \end{bmatrix}$$

we obtain $U_k\tilde{u}_{k+1} = C^{(k)}e_k$. $\square$

Remark: The matrix

$$H(\mathcal{R}) = U(\mathcal{S})^{-1}C^{(n)}U(\mathcal{S}).$$

is the *unique* upper Hessenberg matrix $H(\mathcal{R})$ with the prescribed spectrum and Ritz values and the entry one along the subdiagonal (see also [Parlett, Strang - 2008] where $H(\mathcal{R})$ is constructed in a different way).

Note that $U(\mathcal{S})$ transforms the matrix $C^{(n)}$ with all Ritz values zero to the matrix $H(\mathcal{R})$ with prescribed Ritz values. It is composed of (columns of) companion matrices and we will call $U(\mathcal{S})$ the Ritz value companion transform.

2. Prescribed convergence for Arnoldi's method

Thus the Ritz values generated in the Arnoldi method can exhibit any convergence behavior: It suffices to apply the Arnoldi process with the initial Arnoldi vector e_1 and the matrix $H(\mathcal{R})$ with arbitrarily prescribed Ritz values. Then the method generates the Hessenberg matrix $H(\mathcal{R})$ itself.

2. Prescribed convergence for Arnoldi's method

Thus the Ritz values generated in the Arnoldi method can exhibit any convergence behavior: It suffices to apply the Arnoldi process with the initial Arnoldi vector e_1 and the matrix $H(\mathcal{R})$ with arbitrarily prescribed Ritz values. Then the method generates the Hessenberg matrix $H(\mathcal{R})$ itself.

Question: Can the same prescribed Ritz values be generated with positive entries other than one on the subdiagonal?

2. Prescribed convergence for Arnoldi's method

Thus the Ritz values generated in the Arnoldi method can exhibit any convergence behavior: It suffices to apply the Arnoldi process with the initial Arnoldi vector e_1 and the matrix $H(\mathcal{R})$ with arbitrarily prescribed Ritz values. Then the method generates the Hessenberg matrix $H(\mathcal{R})$ itself.

Question: Can the same prescribed Ritz values be generated with positive entries other than one on the subdiagonal?

For $\sigma_1, \sigma_2, \ldots, \sigma_{n-1} > 0$ consider the diagonal similarity transformation

$$H \equiv \text{ diag}\,(1, \sigma_1, \sigma_1\sigma_2, \ldots, \Pi_{j=1}^{n-1}\sigma_j)\, H(\mathcal{R}) \left(\text{diag}\,(1, \sigma_1, \sigma_1\sigma_2, \ldots, \Pi_{j=1}^{n-1}\sigma_j)\right)^{-1}.$$

2. Prescribed convergence for Arnoldi's method

Thus the Ritz values generated in the Arnoldi method can exhibit any convergence behavior: It suffices to apply the Arnoldi process with the initial Arnoldi vector e_1 and the matrix $H(\mathcal{R})$ with arbitrarily prescribed Ritz values. Then the method generates the Hessenberg matrix $H(\mathcal{R})$ itself.

Question: Can the same prescribed Ritz values be generated with positive entries other than one on the subdiagonal?

For $\sigma_1, \sigma_2, \ldots, \sigma_{n-1} > 0$ consider the diagonal similarity transformation

$$H \equiv \text{diag}\,(1, \sigma_1, \sigma_1\sigma_2, \ldots, \Pi_{j=1}^{n-1}\sigma_j)\, H(\mathcal{R}) \left(\text{diag}\,(1, \sigma_1, \sigma_1\sigma_2, \ldots, \Pi_{j=1}^{n-1}\sigma_j)\right)^{-1}.$$

Then the subdiagonal of H has the entries $\sigma_1, \sigma_2, \ldots, \sigma_{n-1}$ and all leading principal submatrices of H are similar the corresponding leading principal submatrices of $H(\mathcal{R})$.

2. Prescribed convergence for Arnoldi's method

This immediately leads to a parametrization of the matrices and initial Arnoldi vectors that generate a given set of Ritz values $\mathcal{R}$:

2. Prescribed convergence for Arnoldi's method

This immediately leads to a parametrization of the matrices and initial Arnoldi vectors that generate a given set of Ritz values $\mathcal{R}$:

Theorem 2 [DT, Meurant - 2012]. Assume we are given a set of tuples

$$\begin{aligned}\mathcal{R} = \{ \quad & \rho_1^{(1)}, \\ & (\rho_1^{(2)}, \rho_2^{(2)}), \\ & \vdots \\ & (\rho_1^{(n-1)}, \ldots, \rho_{n-1}^{(n-1)}), \\ & (\lambda_1, \ldots\ldots\ldots, \lambda_n)\},\end{aligned}$$

of complex numbers and $n-1$ positive real numbers

$$\sigma_1, \ldots, \sigma_{n-1}.$$

2. Prescribed convergence for Arnoldi's method

This immediately leads to a parametrization of the matrices and initial Arnoldi vectors that generate a given set of Ritz values $\mathcal{R}$:

Theorem 2 [DT, Meurant - 2012]. Assume we are given a set of tuples

$$\begin{aligned}\mathcal{R} = \{ \quad & \rho_1^{(1)}, \\ & (\rho_1^{(2)}, \rho_2^{(2)}), \\ & \vdots \\ & (\rho_1^{(n-1)}, \dots, \rho_{n-1}^{(n-1)}), \\ & (\lambda_1, \dots\dots\dots, \lambda_n)\},\end{aligned}$$

of complex numbers and $n-1$ positive real numbers

$$\sigma_1, \dots, \sigma_{n-1}.$$

If A is a matrix of order n and v a unit nonzero n-dimensional vector, then the following assertions are equivalent:

2. Prescribed convergence for Arnoldi's method

1. The Hessenberg matrix generated by the Arnoldi method applied to A and initial Arnoldi vector v has eigenvalues $\lambda_1, \ldots, \lambda_n$, subdiagonal entries $\sigma_1, \ldots, \sigma_{n-1}$ and $\rho_1^{(k)}, \ldots, \rho_k^{(k)}$ are the eigenvalues of its kth leading principal submatrix for all $k = 1, \ldots, n-1$.

2. Prescribed convergence for Arnoldi's method

1. The Hessenberg matrix generated by the Arnoldi method applied to A and initial Arnoldi vector v has eigenvalues $\lambda_1, \ldots, \lambda_n$, subdiagonal entries $\sigma_1, \ldots, \sigma_{n-1}$ and $\rho_1^{(k)}, \ldots, \rho_k^{(k)}$ are the eigenvalues of its kth leading principal submatrix for all $k = 1, \ldots, n-1$.
2. The matrix A and initial vector v are of the form

$$A = V D_\sigma U(\mathcal{S})^{-1} C^{(n)} U(\mathcal{S}) D_\sigma^{-1} V^*, \qquad v = V e_1,$$

2. Prescribed convergence for Arnoldi's method

1. The Hessenberg matrix generated by the Arnoldi method applied to A and initial Arnoldi vector v has eigenvalues $\lambda_1, \ldots, \lambda_n$, subdiagonal entries $\sigma_1, \ldots, \sigma_{n-1}$ and $\rho_1^{(k)}, \ldots, \rho_k^{(k)}$ are the eigenvalues of its kth leading principal submatrix for all $k = 1, \ldots, n-1$.
2. The matrix A and initial vector v are of the form

$$A = V D_\sigma U(\mathcal{S})^{-1} C^{(n)} U(\mathcal{S}) D_\sigma^{-1} V^*, \qquad v = V e_1,$$

where V is unitary, $U(\mathcal{S})$ is the Ritz value companion transform,

$$D_\sigma = \text{diag}\,(1, \sigma_1, \sigma_1\sigma_2, \ldots, \Pi_{j=1}^{n-1}\sigma_j),$$

and $C^{(n)}$ is the companion matrix of the polynomial with roots $\lambda_1, \ldots, \lambda_n$.

2. Prescribed convergence for Arnoldi's method

1. The Hessenberg matrix generated by the Arnoldi method applied to A and initial Arnoldi vector v has eigenvalues $\lambda_1, \dots, \lambda_n$, subdiagonal entries $\sigma_1, \dots, \sigma_{n-1}$ and $\rho_1^{(k)}, \dots, \rho_k^{(k)}$ are the eigenvalues of its kth leading principal submatrix for all $k = 1, \dots, n-1$.
2. The matrix A and initial vector v are of the form

$$A = V D_\sigma U(\mathcal{S})^{-1} C^{(n)} U(\mathcal{S}) D_\sigma^{-1} V^*, \qquad v = V e_1,$$

where V is unitary, $U(\mathcal{S})$ is the Ritz value companion transform,

$$D_\sigma = \text{diag}\,(1, \sigma_1, \sigma_1\sigma_2, \dots, \Pi_{j=1}^{n-1}\sigma_j),$$

and $C^{(n)}$ is the companion matrix of the polynomial with roots $\lambda_1, \dots, \lambda_n$.

This also shows how little on the quality of the Ritz value ρ needs be said by

$$\|A(V_k y) - \rho(V_k y)\| = h_{k+1,k}|e_k^T y|.$$

Any distance from ρ to the spectrum of A is possible with *any* value of $h_{k+1,k}$!

2. Prescribed convergence for Arnoldi's method

Counterintuitive example 1: Convergence of interior Ritz values only:

$$\mathcal{R} = \{ \begin{array}{c} 3, \\ (3,3)\,, \\ (2,3,4)\,, \\ (3,3,3,3)\,, \\ (1,2,3,4,5)\} \,. \end{array}$$

2. Prescribed convergence for Arnoldi's method

Counterintuitive example 1: Convergence of interior Ritz values only:

$$
\begin{aligned}
\mathcal{R} = \{ \quad & 3, \\
& (3,3)\,, \\
& (2,3,4)\,, \\
& (3,3,3,3)\,, \\
& (1,2,3,4,5)\}\,.
\end{aligned}
$$

This gives the unit upper Hessenberg matrix

$$
H(\mathcal{R}) = U(\mathcal{S})^{-1} C^{(5)} U(\mathcal{S}) = \begin{bmatrix} 3 & 0 & 0 & 0 & 0 \\ 1 & 3 & 1 & 0 & 1 \\ & 1 & 3 & -1 & 0 \\ & & 1 & 3 & 5 \\ & & & 1 & 3 \end{bmatrix}.
$$

2. Prescribed convergence for Arnoldi's method

Thus these Ritz values are generated by the Arnoldi method applied to

$$A = V\,\mathsf{diag}\left(1, \sigma_1, \ldots, \Pi_{j=1}^{n-1}\sigma_j\right) \begin{bmatrix} 3 & 0 & 0 & 0 & 0 \\ 1 & 3 & 1 & 0 & 1 \\ & 1 & 3 & -1 & 0 \\ & & 1 & 3 & 5 \\ & & & 1 & 3 \end{bmatrix} \mathsf{diag}\left(1, \sigma_1, \ldots, \Pi_{j=1}^{n-1}\sigma_j\right)^{-1} V^*$$

with initial vector $v = Ve_1$ and for any unitary V and positive values $\sigma_1, \ldots, \sigma_{n-1}$.

2. Prescribed convergence for Arnoldi's method

Thus these Ritz values are generated by the Arnoldi method applied to

$$A = V \mathsf{diag}\,(1, \sigma_1, \ldots, \Pi_{j=1}^{n-1}\sigma_j) \begin{bmatrix} 3 & 0 & 0 & 0 & 0 \\ 1 & 3 & 1 & 0 & 1 \\ & 1 & 3 & -1 & 0 \\ & & 1 & 3 & 5 \\ & & & 1 & 3 \end{bmatrix} \mathsf{diag}\,(1, \sigma_1, \ldots, \Pi_{j=1}^{n-1}\sigma_j)^{-1} V^*$$

with initial vector $v = Ve_1$ and for any unitary V and positive values $\sigma_1, \ldots, \sigma_{n-1}$.

This is not a highly non-normal example, for instance with $\sigma_i \equiv 1$:

$$\|A\|\|A^{-1}\| = 9.7137,$$

and the eigenvector basis W of A has condition number

$$\|W\|\|W^{-1}\| = 4.8003.$$

2. Prescribed convergence for Arnoldi's method

Counterintuitive example 2: We can prescribe the "diverging" Ritz values

$$\mathcal{R} = \{ \quad 1, \\ (0, 2), \\ (-1, 1, 3), \\ (-2, 0, 2, 4), \\ (1, 1, 1, 1, 1)\},$$

2. Prescribed convergence for Arnoldi's method

Counterintuitive example 2: We can prescribe the "diverging" Ritz values

$$\begin{aligned}\mathcal{R} = \{ \quad & 1, \\ & (0,2)\,, \\ & (-1,1,3)\,, \\ & (-2,0,2,4)\,, \\ & (1,1,1,1,1)\},\end{aligned}$$

with corresponding unit upper Hessenberg matrix

$$H(\mathcal{R}) = U(\mathcal{S})^{-1} C^{(5)} U(\mathcal{S}) = \begin{bmatrix} 1 & 1 & 0 & -3 & 0 \\ 1 & 1 & 3 & 0 & -31 \\ & 1 & 1 & 6 & 0 \\ & & 1 & 1 & -10 \\ & & & 1 & 1 \end{bmatrix}.$$

2. Prescribed convergence for Arnoldi's method

Counterintuitive example 2: We can prescribe the "diverging" Ritz values

$$\begin{aligned}\mathcal{R} = \{ \quad & 1, \\ & (0,2)\,, \\ & (-1,1,3)\,, \\ & (-2,0,2,4)\,, \\ & (1,1,1,1,1)\},\end{aligned}$$

with corresponding unit upper Hessenberg matrix

$$H(\mathcal{R}) = U(\mathcal{S})^{-1} C^{(5)} U(\mathcal{S}) = \begin{bmatrix} 1 & 1 & 0 & -3 & 0 \\ 1 & 1 & 3 & 0 & -31 \\ & 1 & 1 & 6 & 0 \\ & & 1 & 1 & -10 \\ & & & 1 & 1 \end{bmatrix}.$$

These Ritz values are generated by Arnoldi applied to

$$A = VH(\mathcal{R})V^*, \quad v = Ve_1$$

for unitary V.

2. Prescribed convergence for Arnoldi's method

The same "diverging" Ritz values are generated with the exponentially decreasing values 2^{-1}, 2^{-2}, 2^{-3} and 2^{-4} on the subdiagonal of the Hessenberg matrix:

$$A = V \begin{bmatrix} 1 & 2 & 0 & -192 & 0 \\ 0.5 & 1 & 12 & 0 & -15872 \\ & 0.25 & 1 & 48 & 0 \\ & & 0.125 & 1 & -160 \\ & & & 0.0625 & 1 \end{bmatrix} V^*, \qquad v = Ve_1.$$

2. Prescribed convergence for Arnoldi's method

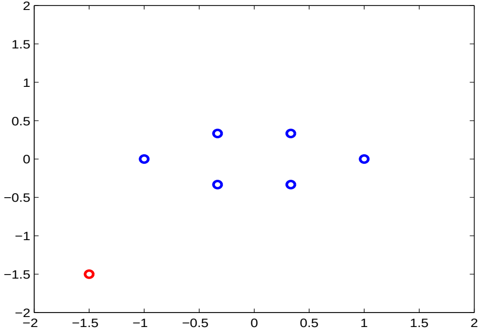
The same "diverging" Ritz values are generated with the exponentially decreasing values 2^{-1}, 2^{-2}, 2^{-3} and 2^{-4} on the subdiagonal of the Hessenberg matrix:

$$A = V \begin{bmatrix} 1 & 2 & 0 & -192 & 0 \\ 0.5 & 1 & 12 & 0 & -15872 \\ & 0.25 & 1 & 48 & 0 \\ & & 0.125 & 1 & -160 \\ & & & 0.0625 & 1 \end{bmatrix} V^*, \qquad v = Ve_1.$$

Then the rounded residual norms $\|A(V_k y) - \rho(V_k y)\| = h_{k+1,k}|e_k^T y|$ seem to indicate convergence:

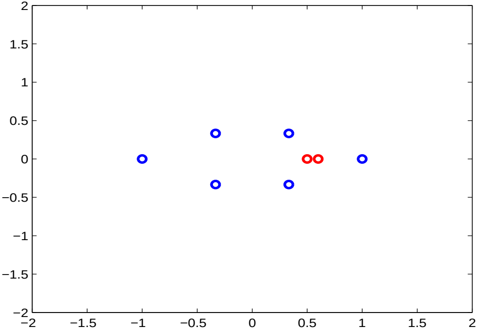
$$\begin{aligned} \{ \quad & \tfrac{1}{2}, \\ & (0.1118, 0.1118)\,, \\ & (0.011, 0.0052, 0.011)\,, \\ & (0.0006, 0.0001, 0.0001, 0.0006)\,\}\,. \end{aligned}$$

2. Prescribed convergence for Arnoldi's method



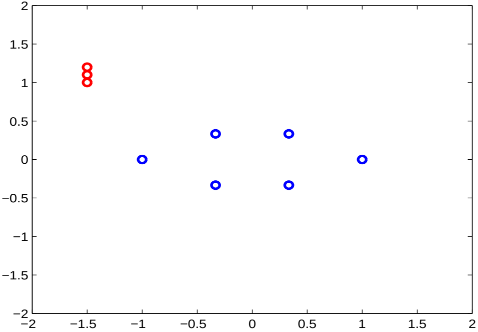
Blue dots: Eigenvalues of A. Red dots: Ritz values in the first iteration.

2. Prescribed convergence for Arnoldi's method



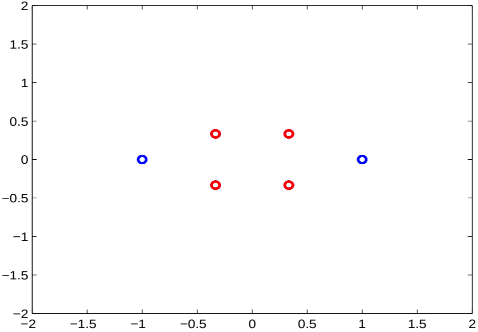
Blue dots: Eigenvalues of A. Red dots: Ritz values in the second iteration.

2. Prescribed convergence for Arnoldi's method



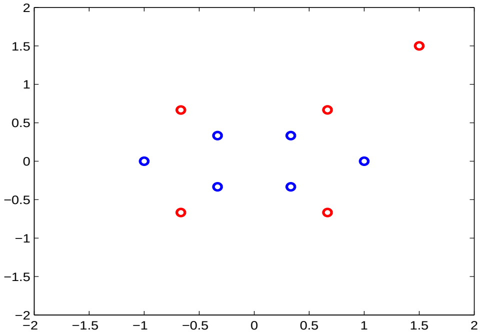
Blue dots: Eigenvalues of A. Red dots: Ritz values in the third iteration.

2. Prescribed convergence for Arnoldi's method



Blue dots: Eigenvalues of A. Red dots: Ritz values in the fourth iteration.

2. Prescribed convergence for Arnoldi's method



Blue dots: Eigenvalues of A. Red dots: Ritz values in the one but last iteration.

2. Prescribed convergence for Arnoldi's method

- This negative result shows that no convergence result can be proved for the very popular Arnoldi method without special assumptions (see also [Embree - 2009] for the Arnoldi method with exact shifts).

2. Prescribed convergence for Arnoldi's method

- This negative result shows that no convergence result can be proved for the very popular Arnoldi method without special assumptions (see also [Embree - 2009] for the Arnoldi method with exact shifts).
- Obviously, any Ritz value ρ lies in the field of values of A:

$$H_k y = \rho y \quad \Rightarrow \quad y^* H_k y = \rho \quad \Rightarrow \quad y^* V_k^* A V_k y = \rho \quad (\|y\| = 1)$$

2. Prescribed convergence for Arnoldi's method

- This negative result shows that no convergence result can be proved for the very popular Arnoldi method without special assumptions (see also [Embree - 2009] for the Arnoldi method with exact shifts).
- Obviously, any Ritz value ρ lies in the field of values of A:

$$H_k y = \rho y \quad \Rightarrow \quad y^* H_k y = \rho \quad \Rightarrow \quad y^* V_k^* A V_k y = \rho \quad (\|y\| = 1)$$

- The result therefore also shows how to construct matrices with a field of values containing $n(n+1)/2$ prescribed complex points

2. Prescribed convergence for Arnoldi's method

- This negative result shows that no convergence result can be proved for the very popular Arnoldi method without special assumptions (see also [Embree - 2009] for the Arnoldi method with exact shifts).
- Obviously, any Ritz value ρ lies in the field of values of A:

$$H_k y = \rho y \quad \Rightarrow \quad y^* H_k y = \rho \quad \Rightarrow \quad y^* V_k^* A V_k y = \rho \quad (\|y\| = 1)$$

- The result therefore also shows how to construct matrices with a field of values containing $n(n+1)/2$ prescribed complex points
- Thus in Crouzeix's conjecture

$$\|p(A)\| \leq 2 \max_{z \in W(A)} |p(z)|,$$

the right-hand side is at least as large as

$$2 \max_{\rho \text{ is a Ritz value}} |p(\rho)|.$$

2. Prescribed convergence for Arnoldi's method

- This negative result shows that no convergence result can be proved for the very popular Arnoldi method without special assumptions (see also [Embree - 2009] for the Arnoldi method with exact shifts).
- Obviously, any Ritz value ρ lies in the field of values of A:

$$H_k y = \rho y \quad \Rightarrow \quad y^* H_k y = \rho \quad \Rightarrow \quad y^* V_k^* A V_k y = \rho \quad (\|y\| = 1)$$

- The result therefore also shows how to construct matrices with a field of values containing $n(n+1)/2$ prescribed complex points
- Thus in Crouzeix's conjecture

$$\|p(A)\| \leq 2 \max_{z \in W(A)} |p(z)|,$$

the right-hand side is at least as large as

$$2 \max_{\rho \text{ is a Ritz value}} |p(\rho)|.$$

- The following tries to gain insight in the conjecture for particular polynomials arising in Krylov subspace methods.

3. Prescribed convergence for Arnoldi *and* GMRES

First, let us consider the GMRES polynomial.

3. Prescribed convergence for Arnoldi *and* GMRES

First, let us consider the GMRES polynomial. To solve a linear system

$$Ax = b, \quad \|b\| = 1,$$

starting with initial guess $x_0 = 0$, GMRES iterates x_k minimize the residual vector $r_k = b - Ax_k$:,

$$\|r_k\| = \|b - Ax_k\| = \min \|b - As\| \quad \text{over all} \quad s \in \mathcal{K}_k(A, b).$$

3. Prescribed convergence for Arnoldi *and* GMRES

First, let us consider the GMRES polynomial. To solve a linear system

$$Ax = b, \quad \|b\| = 1,$$

starting with initial guess $x_0 = 0$, GMRES iterates x_k minimize the residual vector $r_k = b - Ax_k$:,

$$\|r_k\| = \|b - Ax_k\| = \min \|b - As\| \quad \text{over all} \quad s \in \mathcal{K}_k(A, b).$$

The kth residual norm can be written as

$$\|r_k\| = \|p_k^G(A)b\| = \min_{\pi \in \Pi_k^0} \|\pi(A)b\|,$$

where Π_k^0 is the set of polynomials of degree at most k with the value one in the origin.

3. Prescribed convergence for Arnoldi *and* GMRES

First, let us consider the GMRES polynomial. To solve a linear system

$$Ax = b, \quad \|b\| = 1,$$

starting with initial guess $x_0 = 0$, GMRES iterates x_k minimize the residual vector $r_k = b - Ax_k$:,

$$\|r_k\| = \|b - Ax_k\| = \min \|b - As\| \quad \text{over all} \quad s \in \mathcal{K}_k(A, b).$$

The kth residual norm can be written as

$$\|r_k\| = \|p_k^G(A)b\| = \min_{\pi \in \Pi_k^0} \|\pi(A)b\|,$$

where Π_k^0 is the set of polynomials of degree at most k with the value one in the origin. The minimizing polynomial p_k^G is called kth GMRES polynomial.

3. Prescribed convergence for Arnoldi *and* GMRES

First, let us consider the GMRES polynomial. To solve a linear system

$$Ax = b, \quad \|b\| = 1,$$

starting with initial guess $x_0 = 0$, GMRES iterates x_k minimize the residual vector $r_k = b - Ax_k$:,

$$\|r_k\| = \|b - Ax_k\| = \min \|b - As\| \quad \text{over all} \quad s \in \mathcal{K}_k(A, b).$$

The kth residual norm can be written as

$$\|r_k\| = \|p_k^G(A)b\| = \min_{\pi \in \Pi_k^0} \|\pi(A)b\|,$$

where Π_k^0 is the set of polynomials of degree at most k with the value one in the origin. The minimizing polynomial p_k^G is called kth GMRES polynomial.

For the kth GMRES polynomial, Crouzeix's conjecture is

$$\|p_k^G(A)\| \leq 2 \max_{z \in W(A)} |p_k^G(z)|.$$

3. Prescribed convergence for Arnoldi *and* GMRES

Obviously,

$$\|r_k\| = \|p_k^G(A)b\| \leq \|p_k^G(A)\| \leq 2 \max_{z \in W(A)} |p_k^G(z)|.$$

3. Prescribed convergence for Arnoldi *and* GMRES

Obviously,

$$\|r_k\| = \|p_k^G(A)b\| \leq \|p_k^G(A)\| \leq 2 \max_{z \in W(A)} |p_k^G(z)|.$$

Can we say anything about the relation between GMRES residual norms and $|p_k^G(z)|$ on the field of values ?

3. Prescribed convergence for Arnoldi *and* GMRES

Obviously,

$$\|r_k\| = \|p_k^G(A)b\| \leq \|p_k^G(A)\| \leq 2 \max_{z \in W(A)} |p_k^G(z)|.$$

Can we say anything about the relation between GMRES residual norms and $|p_k^G(z)|$ on the field of values ?

We can say something about the relation between GMRES residual norms and at least the Ritz values:

3. Prescribed convergence for Arnoldi *and* GMRES

Obviously,

$$\|r_k\| = \|p_k^G(A)b\| \le \|p_k^G(A)\| \le 2 \max_{z \in W(A)} |p_k^G(z)|.$$

Can we say anything about the relation between GMRES residual norms and $|p_k^G(z)|$ on the field of values ?

We can say something about the relation between GMRES residual norms and at least the Ritz values: In general, there need not be any relation, they can be independent from eachother.

3. Prescribed convergence for Arnoldi *and* GMRES

Obviously,

$$\|r_k\| = \|p_k^G(A)b\| \leq \|p_k^G(A)\| \leq 2 \max_{z \in W(A)} |p_k^G(z)|.$$

Can we say anything about the relation between GMRES residual norms and $|p_k^G(z)|$ on the field of values ?

We can say something about the relation between GMRES residual norms and at least the Ritz values: In general, there need not be any relation, they can be independent from eachother.

Let us try to explain why:

3. Prescribed convergence for Arnoldi *and* GMRES

Obviously,

$$\|r_k\| = \|p_k^G(A)b\| \le \|p_k^G(A)\| \le 2 \max_{z \in W(A)} |p_k^G(z)|.$$

Can we say anything about the relation between GMRES residual norms and $|p_k^G(z)|$ on the field of values ?

We can say something about the relation between GMRES residual norms and at least the Ritz values: In general, there need not be any relation, they can be independent from eachother.

Let us try to explain why: Writing x_k in the Arnoldi basis,

$$x_k = V_k y_k \in \mathcal{K}_k(A, r_0),$$

and using the Arnoldi decomposition $AV_k = V_{k+1}\tilde{H}_k$, we see that

$$\begin{aligned} \|b - Ax_k\| &= \|b - AV_k y_k\| = \|V_{k+1}e_1 - AV_k y_k\| \\ &= \|V_{k+1}(e_1 - \tilde{H}_k y_k)\| = \min_{y \in \mathbb{C}^k} \|e_1 - \tilde{H}_k y\|. \end{aligned}$$

3. Prescribed convergence for Arnoldi *and* GMRES

Thus the residual norms generated by the GMRES method are fully determined by the Hessenberg matrix $\tilde{H}_k$.

3. Prescribed convergence for Arnoldi *and* GMRES

Thus the residual norms generated by the GMRES method are fully determined by the Hessenberg matrix $\tilde{H}_k$.

- We have seen that the subdiagonal entries of $\tilde{H}_k$ can be chosen arbitrarily, for any prescribed Ritz values in the kth iteration.

3. Prescribed convergence for Arnoldi *and* GMRES

Thus the residual norms generated by the GMRES method are fully determined by the Hessenberg matrix $\tilde{H}_k$.

- We have seen that the subdiagonal entries of $\tilde{H}_k$ can be chosen arbitrarily, for any prescribed Ritz values in the kth iteration.

- Hence there is a chance we can modify the behavior of GMRES while maintaining the prescribed Ritz values.

3. Prescribed convergence for Arnoldi *and* GMRES

Thus the residual norms generated by the GMRES method are fully determined by the Hessenberg matrix $\tilde{H}_k$.

- We have seen that the subdiagonal entries of $\tilde{H}_k$ can be chosen arbitrarily, for any prescribed Ritz values in the kth iteration.
- Hence there is a chance we can modify the behavior of GMRES while maintaining the prescribed Ritz values.

Example from earlier: Consider the prescribed 'diverging' Ritz values

$$\begin{aligned}\mathcal{R} = \{ \quad & 1, \\ & (0,2), \\ & (-1,1,3), \\ & (-2,0,2,4), \\ & (1,1,1,1,1)\},\end{aligned}$$

and the prescribed subdiagonal entries of the generated Hessenberg matrix

$$\sigma_1 = 2^{-1}, \quad \sigma_2 = 2^{-2}, \quad \sigma_3 = 2^{-3}, \quad \sigma_4 = 2^{-4}.$$

3. Prescribed convergence for Arnoldi *and* GMRES

The corresponding GMRES convergence curve is

$$\|r^{(0)}\| = 1, \quad \|r^{(1)}\| = \sqrt{\frac{1}{5}}, \quad \|r^{(2)}\| = \sqrt{\frac{1}{5}}, \quad \|r^{(3)}\| = 0.0052, \quad \|r^{(4)}\| = 0.0052.$$

Question: Can we force any GMRES convergence speed with arbitrary Ritz values by modifying the subdiagonal entries?

Not *any*, because there is a relation between GMRES stagnation and zero Ritz values: A singular Hessenberg matrix corresponds to stagnation in the parallel GMRES process, see [Brown - 1991]. In our example we have

$$\begin{aligned}
\rho_1^{(1)} = 1, \quad \|r^{(1)}\| &= \frac{1}{\sqrt{5}} \\
(\rho_1^{(2)}, \rho_2^{(2)}) = (0, 2), \quad \|r^{(2)}\| &= \frac{1}{\sqrt{5}} \\
(\rho_1^{(3)}, \rho_2^{(3)}, \rho_3^{(3)}) = (-1, 1, 3), \quad \|r^{(3)}\| &= 0.0052 \\
(\rho_1^{(4)}, \rho_2^{(4)}, \rho_3^{(4)}, \rho_4^{(4)}) = (-2, 0, 2, 4), \quad \|r^{(4)}\| &= 0.0052.
\end{aligned}$$

3. Prescribed convergence for Arnoldi *and* GMRES

However, this is the *only* restriction Ritz values put on GMRES residual norms:

Theorem 3 [DT, Meurant - 2012]. Consider a set of tuples of complex numbers

$$\begin{aligned}\mathcal{R} = \{ \quad & \rho_1^{(1)}, \\ & (\rho_1^{(2)}, \rho_2^{(2)}), \\ & \vdots \\ & (\rho_1^{(n-1)}, \ldots, \rho_{n-1}^{(n-1)}), \\ & (\lambda_1, \ldots\ldots\ldots, \lambda_n)\},\end{aligned}$$

such that $(\lambda_1, \ldots, \lambda_n)$ contains no zero number and n positive numbers

$$1 \geq f(1) \geq \cdots \geq f(n-1) > 0,$$

such that the k-tuple $(\rho_1^{(k)}, \ldots, \rho_k^{(k)})$ contains a zero number if and only if

$$f(k-1) = f(k).$$

3. Prescribed convergence for Arnoldi *and* GMRES

Let A be a square matrix of size n and let b be a nonzero n-dimensional vector. The following assertions are equivalent:

1. The GMRES method applied to A and right-hand side b with zero initial guess yields residuals $r^{(k)},\ k = 0, \ldots, n-1$ such that

$$\|r^{(k)}\| = f(k), \quad k = 0, \ldots, n-1,$$

A has eigenvalues

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

and

$$\rho_1^{(k)}, \ldots, \rho_k^{(k)}$$

are the Ritz values generated at the kth iteration for $k = 1, \ldots, n-1$.

3. Prescribed convergence for Arnoldi *and* GMRES

2. The matrix A and right hand side b are of the form

$$A = VU^{-1}C^{(n)}UV^*, \qquad b = Ve_1,$$

where V is a unitary matrix,

$$U = \begin{bmatrix} & g^T & \\ 0 & T & \end{bmatrix}$$

where the first row g^T of U is

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

3. Prescribed convergence for Arnoldi *and* GMRES

2. The matrix A and right hand side b are of the form

$$A = VU^{-1}C^{(n)}UV^*, \qquad b = Ve_1,$$

where V is a unitary matrix,

$$U = \begin{bmatrix} & g^T & \\ 0 & & T \end{bmatrix}$$

where the first row g^T of U is

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

and the remaining submatrix T of has entries satisfying

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

3. Prescribed convergence for Arnoldi *and* GMRES

2. The matrix A and right hand side b are of the form

$$A = VU^{-1}C^{(n)}UV^*, \qquad b = Ve_1,$$

where V is a unitary matrix,

$$U = \begin{bmatrix} & g^T & \\ 0 & T & \end{bmatrix}$$

where the first row g^T of U is

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

and the remaining submatrix T of has entries satisfying

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

Note we exhausted all freedom modulo unitary transformation.

3. Prescribed convergence for Arnoldi *and* GMRES

Example: Standardly converging Ritz values and 'nearly stagnating' GMRES:

$$\mathcal{R} = \{ \quad 5, \\ (1,5)\,, \\ (1,4,5)\,, \\ (1,3,4,5)\,, \\ (1,2,3,4,5)\}\,,$$

$$\begin{aligned} \|r^{(0)}\| &= 1, \quad \|r^{(1)}\| = 0.9, \quad \|r^{(2)}\| = 0.8, \\ \|r^{(3)}\| &= 0.7, \quad \|r^{(4)}\| = 0.6, \quad \|r^{(5)}\| = 0 \quad \text{gives} \end{aligned}$$

3. Prescribed convergence for Arnoldi *and* GMRES

Example: Standardly converging Ritz values and 'nearly stagnating' GMRES:

$$\mathcal{R} = \{ \quad 5, \\ (1,5)\,, \\ (1,4,5)\,, \\ (1,3,4,5)\,, \\ (1,2,3,4,5)\}\,,$$

$$\begin{aligned} \|r^{(0)}\| &= 1, \quad \|r^{(1)}\| = 0.9, \quad \|r^{(2)}\| = 0.8, \\ \|r^{(3)}\| &= 0.7, \quad \|r^{(4)}\| = 0.6, \quad \|r^{(5)}\| = 0 \quad \text{gives} \end{aligned}$$

$$A = V \begin{bmatrix} 5 & 0 & 0 & 0 & 0 \\ 10.3237 & 1 & 0 & 0 & 0 \\ & 0.8458 & 4 & 0 & 0 \\ & & 3.312 & 3 & 0 \\ & & & 2.4169 & 2 \end{bmatrix} V^*, \qquad b = Ve_1.$$

3. Prescribed convergence for Arnoldi *and* GMRES

Again, this is not a highly non-normal example:

$$\|A\|\|A^{-1}\| = 28.9498,$$

and the eigenvector basis W of A has condition number

$$\|W\|\|W^{-1}\| = 57.735.$$

The residual norms $\|A(V_k y) - \rho(V_k y)\| = h_{k+1,k}|e_k^T y|$ for the Ritz pairs are

$$\begin{gathered} 10.3237, \\ (0.8458, 0.7886), \\ (0.8987, 3.312, 2.0509), \\ (0.9906, 2.4169, 2.3137, 1.7303). \end{gathered}$$

respectively, i.e. they give misleading information.

3. Prescribed convergence for Arnoldi *and* GMRES

Summarizing, any GMRES residual norms are possible with any Ritz values in all iterations.

3. Prescribed convergence for Arnoldi *and* GMRES

Summarizing, any GMRES residual norms are possible with any Ritz values in all iterations.

We have proved two more analogue results:

- Any GMRES residual norms are possible with any *harmonic* Ritz values in all iterations.

3. Prescribed convergence for Arnoldi *and* GMRES

Summarizing, any GMRES residual norms are possible with any Ritz values in all iterations.

We have proved two more analogue results:

- Any GMRES residual norms are possible with any *harmonic* Ritz values in all iterations. This is perhaps even more surprising, because the harmonic Ritz values are the roots of GMRES polynomials.

3. Prescribed convergence for Arnoldi *and* GMRES

Summarizing, any GMRES residual norms are possible with any Ritz values in all iterations.

We have proved two more analogue results:

- Any GMRES residual norms are possible with any *harmonic* Ritz values in all iterations. This is perhaps even more surprising, because the harmonic Ritz values are the roots of GMRES polynomials. However, harmonic Ritz values are not in general in the field of values of A, so this has no implications for Crouzeix's conjecture for the GMRES polynomial.

3. Prescribed convergence for Arnoldi *and* GMRES

Summarizing, any GMRES residual norms are possible with any Ritz values in all iterations.

We have proved two more analogue results:

- Any GMRES residual norms are possible with any *harmonic* Ritz values in all iterations. This is perhaps even more surprising, because the harmonic Ritz values are the roots of GMRES polynomials. However, harmonic Ritz values are not in general in the field of values of A, so this has no implications for Crouzeix's conjecture for the GMRES polynomial.
- Any FOM residual norms are possible with any Ritz values in all iterations.

3. Prescribed convergence for Arnoldi *and* GMRES

Summarizing, any GMRES residual norms are possible with any Ritz values in all iterations.

We have proved two more analogue results:

- Any GMRES residual norms are possible with any *harmonic* Ritz values in all iterations. This is perhaps even more surprising, because the harmonic Ritz values are the roots of GMRES polynomials. However, harmonic Ritz values are not in general in the field of values of A, so this has no implications for Crouzeix's conjecture for the GMRES polynomial.
- Any FOM residual norms are possible with any Ritz values in all iterations.

The FOM method differs from the GMRES method in that the residual norm is not minimized, but the kth FOM residual vector is characterized through

$$r_k^F \perp \mathcal{K}_k(A, b).$$

3. Prescribed convergence for Arnoldi *and* GMRES

The corresponding residual norms are related through to formula

$$\frac{1}{\|r_k^F\|} = \sqrt{\frac{1}{\|r_k^G\|^2} - \frac{1}{\|r_{k-1}^G\|^2}}.$$

Note that FOM residual norms need not be non-increasing and are not defined if the corresponding GMRES iterate stagnates.

3. Prescribed convergence for Arnoldi *and* GMRES

The corresponding residual norms are related through to formula

$$\frac{1}{\|r_k^F\|} = \sqrt{\frac{1}{\|r_k^G\|^2} - \frac{1}{\|r_{k-1}^G\|^2}}.$$

Note that FOM residual norms need not be non-increasing and are not defined if the corresponding GMRES iterate stagnates.

What seems interesting to me in the context of the Crouzeix's conjecture is that the Ritz values are the roots of the FOM polynomials:

3. Prescribed convergence for Arnoldi *and* GMRES

The corresponding residual norms are related through to formula

$$\frac{1}{\|r_k^F\|} = \sqrt{\frac{1}{\|r_k^G\|^2} - \frac{1}{\|r_{k-1}^G\|^2}}.$$

Note that FOM residual norms need not be non-increasing and are not defined if the corresponding GMRES iterate stagnates.

What seems interesting to me in the context of the Crouzeix's conjecture is that the Ritz values are the roots of the FOM polynomials:

FOM polynomials might lead to a way to test if the conjecture can be disproved. For the kth FOM polynomial p_k^F we have,

$$0 = 2 \max_{\rho \text{ is a Ritz value}} |p_k^F(\rho)| < \|r_k^F\| = \|p_k^F(A)b\| \leq \|p_k^F(A)\|.$$

3. Prescribed convergence for Arnoldi *and* GMRES

The corresponding residual norms are related through to formula

$$\frac{1}{\|r_k^F\|} = \sqrt{\frac{1}{\|r_k^G\|^2} - \frac{1}{\|r_{k-1}^G\|^2}}.$$

Note that FOM residual norms need not be non-increasing and are not defined if the corresponding GMRES iterate stagnates.

What seems interesting to me in the context of the Crouzeix's conjecture is that the Ritz values are the roots of the FOM polynomials:

FOM polynomials might lead to a way to test if the conjecture can be disproved. For the kth FOM polynomial p_k^F we have,

$$0 = 2 \max_{\rho \text{ is a Ritz value}} |p_k^F(\rho)| < \|r_k^F\| = \|p_k^F(A)b\| \leq \|p_k^F(A)\|.$$

Note the residual norms can be chosen arbitrarely large...

3. Prescribed convergence for Arnoldi *and* GMRES

The corresponding residual norms are related through to formula

$$\frac{1}{\|r_k^F\|} = \sqrt{\frac{1}{\|r_k^G\|^2} - \frac{1}{\|r_{k-1}^G\|^2}}.$$

Note that FOM residual norms need not be non-increasing and are not defined if the corresponding GMRES iterate stagnates.

What seems interesting to me in the context of the Crouzeix's conjecture is that the Ritz values are the roots of the FOM polynomials:

FOM polynomials might lead to a way to test if the conjecture can be disproved. For the kth FOM polynomial p_k^F we have,

$$0 = 2 \max_{\rho \text{ is a Ritz value}} |p_k^F(\rho)| < \|r_k^F\| = \|p_k^F(A)b\| \leq \|p_k^F(A)\|.$$

Note the residual norms can be chosen arbitrarely large...

The construction to prescribe Ritz values and FOM residual norms is the following:

3. Prescribed convergence for Arnoldi *and* GMRES

The matrix A and right hand side b are of the form

$$A = VU^{-1}C^{(n)}UV^*, \qquad b = Ve_1,$$

where V is a unitary matrix,

$$U = \begin{bmatrix} & g^T & \\ 0 & & T \end{bmatrix}$$

where 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

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

and the remaining submatrix T of has entries satisfying

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

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.