On short recurrences for generating orthogonal Krylov subspace bases

Petr Tichý

joint work with

Vance Faber, Jörg Liesen, Zdeněk Strakoš

Institute of Computer Science AS CR

March 19, 2009, Podbanské, Slovakia
Algoritmy 2009

Outline

Outline

Krylov subspace methods

Basis

Methods based on projection onto the Krylov subspaces

$$\mathcal{K}_j(\mathbf{A}, v) \equiv \mathrm{span}(v, \mathbf{A}v, \ldots, \mathbf{A}^{j-1}v) \qquad j = 1, 2, \ldots .$$

$\mathbf{A} \in \mathbb{R}^{n \times n}$, $v \in \mathbb{R}^n$.

Methods based on projection onto the Krylov subspaces

$$\mathcal{K}_j(\mathbf{A}, v) \equiv \mathrm{span}(v, \mathbf{A}v, \ldots, \mathbf{A}^{j-1}v) \qquad j = 1, 2, \ldots .$$

$\mathbf{A} \in \mathbb{R}^{n \times n}$, $v \in \mathbb{R}^n$.

Each method must generate a basis of $\mathcal{K}_j(\mathbf{A}, v)$.

- The trivial choice $v, \mathbf{A}v, \ldots, \mathbf{A}^{j-1}v$ is computationally infeasible (recall the Power Method).
- For numerical stability: Well conditioned basis.
- For computational efficiency: Short recurrence.

Methods based on projection onto the Krylov subspaces

$$\mathcal{K}_j(\mathbf{A}, v) \equiv \mathrm{span}(v, \mathbf{A}v, \dots, \mathbf{A}^{j-1}v) \qquad j = 1, 2, \dots .$$

$\mathbf{A} \in \mathbb{R}^{n \times n}$, $v \in \mathbb{R}^n$.

Each method must generate a basis of $\mathcal{K}_j(\mathbf{A}, v)$.

- The trivial choice $v, \mathbf{A}v, \dots, \mathbf{A}^{j-1}v$ is computationally infeasible (recall the Power Method).
- For numerical stability: Well conditioned basis.
- For computational efficiency: Short recurrence.
- Best of both worlds:
  Orthogonal basis computed by short recurrence.

Optimal Krylov subspace methods

with short recurrences

CG (1952), MINRES, SYMMLQ (1975)

- based on three-term recurrences

$$r_{j+1} = \gamma_j \mathbf{A} r_j - \alpha_j r_j - \beta_j r_{j-1}\,,$$

Optimal Krylov subspace methods

with short recurrences

CG (1952), MINRES, SYMMLQ (1975)

- based on three-term recurrences

$$r_{j+1} = \gamma_j \mathbf{A} r_j - \alpha_j r_j - \beta_j r_{j-1} \,,$$

- generate orthogonal (or $\mathbf{A}$-orthogonal) Krylov subspace basis,

Optimal Krylov subspace methods

with short recurrences

CG (1952), MINRES, SYMMLQ (1975)

- based on three-term recurrences

$$r_{j+1} = \gamma_j \mathbf{A} r_j - \alpha_j r_j - \beta_j r_{j-1} \,,$$

- generate orthogonal (or $\mathbf{A}$-orthogonal) Krylov subspace basis,

- *optimal* in the sense that they minimize some error norm:

  $\|x - x_j\|_{\mathbf{A}}$ in CG,
  $\|x - x_j\|_{\mathbf{A}^T\mathbf{A}} = \|r_j\|$ in MINRES,
  $\|x - x_j\|$ in SYMMLQ -here $x_j \in x_0 + \mathbf{A}\mathcal{K}_j(\mathbf{A}, r_0)$.

Optimal Krylov subspace methods
with short recurrences

CG (1952), MINRES, SYMMLQ (1975)

- based on three-term recurrences

$$r_{j+1} = \gamma_j \mathbf{A} r_j - \alpha_j r_j - \beta_j r_{j-1}\,,$$

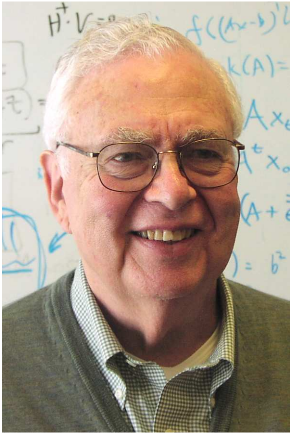
- generate orthogonal (or $\mathbf{A}$-orthogonal) Krylov subspace basis,

- *optimal* in the sense that they minimize some error norm:

  $\|x - x_j\|_{\mathbf{A}}$ in CG,
  $\|x - x_j\|_{\mathbf{A}^T\mathbf{A}} = \|r_j\|$ in MINRES,
  $\|x - x_j\|$ in SYMMLQ -here $x_j \in x_0 + \mathbf{A}\mathcal{K}_j(\mathbf{A}, r_0)$.

- An important assumption on $\mathbf{A}$:
  $\mathbf{A}$ is symmetric (MINRES, SYMMLQ) & pos. definite (CG).

Gene Golub



G. H. Golub, 1932–2007

- By the end of the 1970s it was unknown if such methods existed also for general unsymmetric $\mathbf{A}$.
- Gatlinburg VIII (now Householder VIII) held in Oxford from July 5 to 11, 1981.
- "A prize of $500 has been offered by Gene Golub for the construction of a 3-term conjugate gradient like descent method for non-symmetric real matrices or a proof that there can be no such method".

What kind of method Golub had in mind

- We want to solve $\mathbf{A}x = b$ using CG-like descent method: error is minimized in some given inner product norm, $\| \cdot \|_{\mathbf{B}} = \langle \cdot, \cdot \rangle_{\mathbf{B}}^{1/2}$.

What kind of method Golub had in mind

- We want to solve $\mathbf{A}x = b$ using CG-like descent method: error is minimized in some given inner product norm, $\| \cdot \|_{\mathbf{B}} = \langle \cdot, \cdot \rangle_{\mathbf{B}}^{1/2}$.
- Starting from x_0, compute

$$x_{j+1} = x_j + \alpha_j p_j \,, \qquad j = 0, 1, \dots \,,$$

p_j is a direction vector, α_j is a scalar (to be determined),

$$\mathrm{span}\{p_0, \dots, p_j\} = \mathcal{K}_{j+1}(\mathbf{A}, r_0), \qquad r_0 = b - \mathbf{A}x_0 \,.$$

What kind of method Golub had in mind

- We want to solve $\mathbf{A}x = b$ using CG-like descent method: error is minimized in some given inner product norm, $\| \cdot \|_{\mathbf{B}} = \langle \cdot, \cdot \rangle_{\mathbf{B}}^{1/2}$.
- Starting from x_0, compute

$$x_{j+1} = x_j + \alpha_j p_j \,, \qquad j = 0, 1, \dots \,,$$

p_j is a direction vector, α_j is a scalar (to be determined),

$$\mathrm{span}\{p_0, \dots, p_j\} = \mathcal{K}_{j+1}(\mathbf{A}, r_0), \qquad r_0 = b - \mathbf{A}x_0 \,.$$

- $\|x - x_{j+1}\|_{\mathbf{B}}$ is minimal iff

$$\alpha_j = \frac{\langle x - x_j, p_j \rangle_{\mathbf{B}}}{\langle p_j, p_j \rangle_{\mathbf{B}}} \quad \text{and} \quad \langle p_j, p_i \rangle_{\mathbf{B}} = 0 \,.$$

What kind of method Golub had in mind

- We want to solve $\mathbf{A}x = b$ using CG-like descent method: error is minimized in some given inner product norm, $\| \cdot \|_{\mathbf{B}} = \langle \cdot, \cdot \rangle_{\mathbf{B}}^{1/2}$.
- Starting from x_0, compute

$$x_{j+1} = x_j + \alpha_j p_j \,, \qquad j = 0, 1, \dots \,,$$

p_j is a direction vector, α_j is a scalar (to be determined),

$$\mathrm{span}\{p_0, \dots, p_j\} = \mathcal{K}_{j+1}(\mathbf{A}, r_0), \qquad r_0 = b - \mathbf{A}x_0 \,.$$

- $\|x - x_{j+1}\|_{\mathbf{B}}$ is minimal iff

$$\alpha_j = \frac{\langle x - x_j, p_j \rangle_{\mathbf{B}}}{\langle p_j, p_j \rangle_{\mathbf{B}}} \quad \text{and} \quad \langle p_j, p_i \rangle_{\mathbf{B}} = 0 \,.$$

- $p_0, \dots, p_j$ has to be a $\mathbf{B}$-orthogonal basis of $\mathcal{K}_{j+1}(\mathbf{A}, r_0)$.

Faber and Manteuffel, 1984

SIAM J. NUMER. ANAL.
Vol. 21, No. 2, April 1984

011

NECESSARY AND SUFFICIENT CONDITIONS FOR THE EXISTENCE OF A CONJUGATE GRADIENT METHOD*

VANCE FABER† AND THOMAS MANTEUFFEL†

Abstract. We characterize the class $CG(s)$ of matrices A for which the linear system $A\mathbf{x}=\mathbf{b}$ can be solved by an s-term conjugate gradient method. We show that, except for a few anomalies, the class $CG(s)$ consists of matrices A for which conjugate gradient methods are already known. These matrices are the Hermitian matrices, $A^*=A$, and the matrices of the form $A=e^{i\theta}(dI+B)$, with $B^*=-B$.

- Faber and Manteuffel gave the answer in 1984:
  For a general matrix $\mathbf{A}$ there exists *no* short recurrence for generating orthogonal Krylov subspace bases.
- **What are the details of this statement**?

Outline

Formulation of the problem

$\mathbf{B}$-inner product, Input and Notation

Without loss of generality, $\mathbf{B} = \mathbf{I}$. Otherwise change the basis:

$$\langle x, y \rangle_{\mathbf{B}} = \langle \mathbf{B}^{1/2} x, \mathbf{B}^{1/2} y \rangle, \quad \hat{\mathbf{A}} \equiv \mathbf{B}^{1/2} \mathbf{A} \mathbf{B}^{-1/2}, \quad \hat{v} \equiv \mathbf{B}^{1/2} v \,.$$

Formulation of the problem

$\mathbf{B}$-inner product, Input and Notation

Without loss of generality, $\mathbf{B} = \mathbf{I}$. Otherwise change the basis:

$$\langle x, y \rangle_{\mathbf{B}} = \langle \mathbf{B}^{1/2} x, \mathbf{B}^{1/2} y \rangle, \quad \hat{\mathbf{A}} \equiv \mathbf{B}^{1/2} \mathbf{A} \mathbf{B}^{-1/2}, \quad \hat{v} \equiv \mathbf{B}^{1/2} v \, .$$

Input data:

- $\mathbf{A} \in \mathbb{C}^{n \times n}$, a nonsingular matrix.
- $v \in \mathbb{C}^{n}$, an initial vector.

Without loss of generality, $\mathbf{B} = \mathbf{I}$. Otherwise change the basis:

$$\langle x, y \rangle_{\mathbf{B}} = \langle \mathbf{B}^{1/2}x, \mathbf{B}^{1/2}y \rangle, \quad \hat{\mathbf{A}} \equiv \mathbf{B}^{1/2}\mathbf{A}\mathbf{B}^{-1/2}, \quad \hat{v} \equiv \mathbf{B}^{1/2}v\,.$$

Input data:

- $\mathbf{A} \in \mathbb{C}^{n \times n}$, a nonsingular matrix.
- $v \in \mathbb{C}^{n}$, an initial vector.

Notation:

- $d_{\min}(\mathbf{A})$ ... the degree of the minimal polynomial of $\mathbf{A}$.
- $d = d(\mathbf{A}, v)$ ... the grade of v with respect to $\mathbf{A}$,
  the smallest d s.t. $\mathcal{K}_d(\mathbf{A}, v)$ is invariant under mult. with $\mathbf{A}$.

Formulation of the problem

Our Goal

- Generate a basis $v_1, \ldots, v_d$ of $\mathcal{K}_d(\mathbf{A}, v)$ s.t.
  1. $\mathrm{span}\{v_1, \ldots, v_j\} = \mathcal{K}_j(A, v),$ for $j = 1, \ldots, d$,
  2. $\langle v_i, v_j \rangle = 0,$ for $i \neq j, \quad i, j = 1, \ldots, d$.

- Generate a basis $v_1, \ldots, v_d$ of $\mathcal{K}_d(\mathbf{A}, v)$ s.t.
  1. $\mathrm{span}\{v_1, \ldots, v_j\} = \mathcal{K}_j(A, v),$ for $j = 1, \ldots, d$,
  2. $\langle v_i, v_j \rangle = 0,$ for $i \neq j,$ $i, j = 1, \ldots, d$.

Arnoldi's algorithm:
Standard way for generating the orthogonal basis
(no normalization for convenience): $v_1 \equiv v$,

$$v_{j+1} = \mathbf{A} v_j - \sum_{i=1}^{j} h_{i,j} \, v_i \,, \qquad h_{i,j} = \frac{\langle \mathbf{A} v_j, v_i \rangle}{\langle v_i, v_i \rangle} \,,$$

$j = 0, \ldots, d-1$.

Formulation of the problem

Arnoldi's algorithm - matrix formulation

In matrix notation:

$$v_1 = v\,,$$

$$\mathbf{A}\underbrace{[v_1,\dots,v_{d-1}]}_{\equiv \mathbf{V}_{d-1}} = \underbrace{[v_1,\dots,v_d]}_{\equiv \mathbf{V}_d}\underbrace{\begin{bmatrix} h_{1,1} & \cdots & h_{1,d-1} \\ 1 & \ddots & \vdots \\ & \ddots & h_{d-1,d-1} \\ & & 1 \end{bmatrix}}_{\equiv \mathbf{H}_{d,d-1}},$$

$$\mathbf{V}_d^*\mathbf{V}_d \text{ is diagonal}\,, \quad d = \dim \mathcal{K}_n(\mathbf{A}, v)\,.$$

Formulation of the problem

Arnoldi's algorithm - matrix formulation

In matrix notation:

$$v_1 = v\,,$$

$$\mathbf{A}\,\underbrace{[v_1,\ldots,v_{d-1}]}_{\equiv\,\mathbf{V}_{d-1}} = \underbrace{[v_1,\ldots,v_d]}_{\equiv\,\mathbf{V}_d}\underbrace{\begin{bmatrix} h_{1,1} & \cdots & h_{1,d-1} \\ 1 & \ddots & \vdots \\ & \ddots & h_{d-1,d-1} \\ & & 1 \end{bmatrix}}_{\equiv\,\mathbf{H}_{d,d-1}},$$

$$\mathbf{V}_d^*\mathbf{V}_d \text{ is diagonal}\,, \quad d = \dim \mathcal{K}_n(\mathbf{A}, v)\,.$$

$(s+2)$-term recurrence: $\quad v_{j+1} = \mathbf{A}\,v_j - \sum_{\mathbf{i=j-s}}^{j} h_{i,j} v_i\,.$

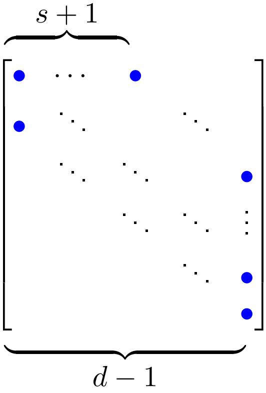
Formulation of the problem

Optimal short recurrences (Definition - Liesen and Strakoš, 2008)

$\mathbf{A}$ admits an optimal $(s+2)$-term recurrence, if

- for any v, $\mathbf{H}_{d,d-1}$ is at most $(s+2)$-band Hessenberg, and
- for at least one v, $\mathbf{H}_{d,d-1}$ is $(s+2)$-band Hessenberg.

$$\mathbf{A}\,\mathbf{V}_{d-1} \;=\; \mathbf{V}_d$$



$\mathbf{A}$ admits an optimal $(s+2)$-term recurrence, if

- for any v, $\mathbf{H}_{d,d-1}$ is at most $(s+2)$-band Hessenberg, and
- for at least one v, $\mathbf{H}_{d,d-1}$ is $(s+2)$-band Hessenberg.

$$\mathbf{A}\,\mathbf{V}_{d-1} \;=\; \mathbf{V}_d \; \underbrace{\begin{bmatrix} \overbrace{\bullet \quad \cdots \quad \bullet}^{s+1} & & \\ \bullet \quad \ddots & \ddots & \\ \ddots \quad \ddots & & \bullet \\ & \ddots \quad \ddots & \vdots \\ & \ddots & \bullet \\ & & \bullet \end{bmatrix}}_{d-1}$$

Sufficient and necessary conditions on $\mathbf{A}$?

Outline

The Faber-Manteuffel theorem

Definition. If $\mathbf{A}^* = p_s(\mathbf{A})$, where p_s is a polynomial of the smallest possible degree s, $\mathbf{A}$ is called normal(s).

The Faber-Manteuffel theorem

Definition. If $\mathbf{A}^* = p_s(\mathbf{A})$, where p_s is a polynomial of the smallest possible degree s, $\mathbf{A}$ is called normal(s).

Theorem. [Faber and Manteuffel, 1984], [Liesen and Strakoš, 2008]

Let $\mathbf{A}$ be a nonsingular matrix with minimal polynomial degree $d_{\min}(\mathbf{A})$. Let s be a nonnegative integer, $s + 2 < d_{\min}(\mathbf{A})$:

$\mathbf{A}$ admits an optimal $(s+2)$-term recurrence

if and only if

$\mathbf{A}$ is normal(s).

The Faber-Manteuffel theorem

Definition. If $\mathbf{A}^* = p_s(\mathbf{A})$, where p_s is a polynomial of the smallest possible degree s, $\mathbf{A}$ is called normal(s).

Theorem. [Faber and Manteuffel, 1984], [Liesen and Strakoš, 2008]

Let $\mathbf{A}$ be a nonsingular matrix with minimal polynomial degree $d_{\min}(\mathbf{A})$. Let s be a nonnegative integer, $s + 2 < d_{\min}(\mathbf{A})$:

$\mathbf{A}$ admits an optimal $(s + 2)$-term recurrence

if and only if

$\mathbf{A}$ is normal(s).

- Sufficiency is rather straightforward, necessity *is not*. Key words from the proof of necessity in (Faber and Manteuffel, 1984) include: "continuous function" (analysis), "closed set of smaller dimension" (topology), "wedge product" (multilinear algebra).

Outline

V. Faber, J. Liesen and P. Tichý, 2008

The Faber-Manteuffel Theorem for Linear Operators

- Motivated by the paper [J. Liesen and Z. Strakoš, 2008] which contains a completely reworked theory of short recurrences for generating orthogonal Krylov subspace bases.

  "It is unknown if a simpler proof of the necessity part can be found. In view of the fundamental nature of the Faber-Manteuffel Theorem, such proof would be a welcome addition to the existing literature. It would lead to a better understanding of the theorem by enlightening some (possibly unexpected) relationships, and it would also be more suitable for classroom teaching."

V. Faber, J. Liesen and P. Tichý, 2008

The Faber-Manteuffel Theorem for Linear Operators

- Motivated by the paper [J. Liesen and Z. Strakoš, 2008] which contains a completely reworked theory of short recurrences for generating orthogonal Krylov subspace bases.

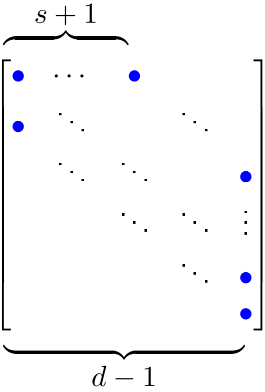
  "It is unknown if a simpler proof of the necessity part can be found. In view of the fundamental nature of the Faber-Manteuffel Theorem, such proof would be a welcome addition to the existing literature. It would lead to a better understanding of the theorem by enlightening some (possibly unexpected) relationships, and it would also be more suitable for classroom teaching."

- We give two new proofs of the Faber-Manteuffel theorem that use more elementary tools,
- first proof - improved version of the Faber-Manteuffel proof,
- second proof - completely new proof based on orthogonal transformations of upper Hessenberg matrices.

Idea of the second proof

Optimal $(s+2)$-term recurrence

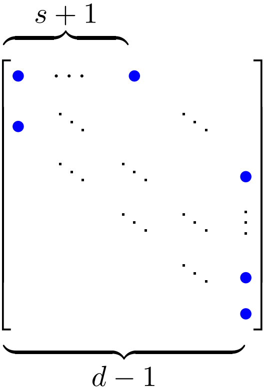
$$\mathbf{A}\,\mathbf{V}_{d-1} \;=\; \mathbf{V}_d$$



Idea of the second proof

Optimal $(s+2)$-term recurrence

$$\mathbf{A}\,\mathbf{V}_{d-1} \;=\; \mathbf{V}_d$$



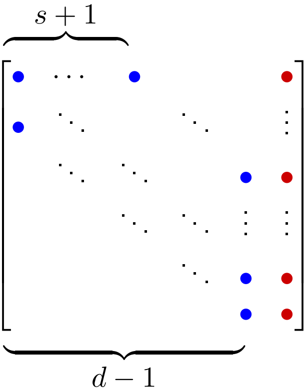
Since $\mathcal{K}_d(\mathbf{A}, v)$ is invariant, $\mathbf{A}v_d \in \mathcal{K}_d(\mathbf{A}, v)$ and

$$\mathbf{A}v_d = \sum_{i=1}^{d} h_{id}\, v_i.$$

Idea of the second proof

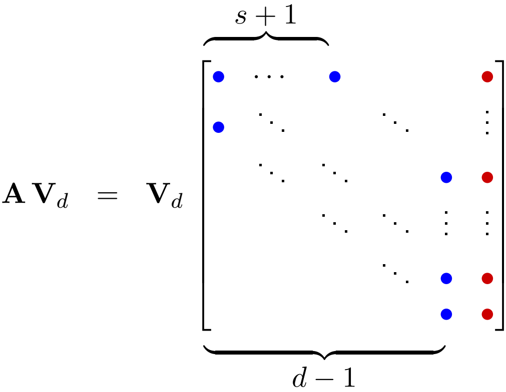
Matrix representation of $\mathbf{A}$ in $\mathbf{V}_d$

$$\mathbf{A}\,\mathbf{V}_d \;=\; \mathbf{V}_d$$



Idea of the second proof

Matrix representation of $\mathbf{A}$ in $\mathbf{V}_d$



Prove something about the linear operator $\mathbf{A}$, without complete knowledge of the structure of its matrix representation.

Idea of the second proof

V. Faber, J. Liesen and P. Tichý, 2008

(for simplicity, we omit indices by $\mathbf{V}_d$ and $\mathbf{H}_{d,d}$)

Let $\mathbf{A}$ admit an optimal $(s+2)$-term recurrence

$$\mathbf{A}\,\mathbf{V} \;=\; \mathbf{V}\,\mathbf{H}, \quad \mathbf{V}^*\mathbf{V} = \mathbf{I}\,.$$

Up to the last column, $\mathbf{H}$ is $(s+2)$-band Hessenberg.

Idea of the second proof

V. Faber, J. Liesen and P. Tichý, 2008

(for simplicity, we omit indices by $\mathbf{V}_d$ and $\mathbf{H}_{d,d}$)

Let $\mathbf{A}$ admit an optimal $(s+2)$-term recurrence

$$\mathbf{A}\,\mathbf{V} \;=\; \mathbf{V}\,\mathbf{H}, \quad \mathbf{V}^*\mathbf{V} = \mathbf{I}\,.$$

Up to the last column, $\mathbf{H}$ is $(s+2)$-band Hessenberg.
Let $\mathbf{G}$ be a $d \times d$ unitary matrix, $\mathbf{G}^*\mathbf{G} = \mathbf{I}$. Then

$$\mathbf{A}\,\underbrace{(\mathbf{V}\mathbf{G})}_{\mathbf{W}} \;=\; \underbrace{(\mathbf{V}\mathbf{G})}_{\mathbf{W}}\,\underbrace{(\mathbf{G}^*\mathbf{H}\mathbf{G})}_{\widetilde{\mathbf{H}}}\,.$$

$\mathbf{W}$ is unitary.

Idea of the second proof

V. Faber, J. Liesen and P. Tichý, 2008

(for simplicity, we omit indices by $\mathbf{V}_d$ and $\mathbf{H}_{d,d}$)

Let $\mathbf{A}$ admit an optimal $(s+2)$-term recurrence

$$\mathbf{A}\,\mathbf{V} \;=\; \mathbf{V}\,\mathbf{H}, \quad \mathbf{V}^*\mathbf{V} = \mathbf{I}\,.$$

Up to the last column, $\mathbf{H}$ is $(s+2)$-band Hessenberg.
Let $\mathbf{G}$ be a $d \times d$ unitary matrix, $\mathbf{G}^*\mathbf{G} = \mathbf{I}$. Then

$$\mathbf{A}\,\underbrace{(\mathbf{VG})}_{\mathbf{W}} \;=\; \underbrace{(\mathbf{VG})}_{\mathbf{W}}\,\underbrace{(\mathbf{G}^*\mathbf{HG})}_{\widetilde{\mathbf{H}}}\,.$$

$\mathbf{W}$ is unitary. If $\mathbf{G}$ is chosen such that $\widetilde{\mathbf{H}}$ is again unreduced upper Hessenberg matrix, then

$$\mathbf{A}\,\mathbf{W} \;=\; \mathbf{W}\,\tilde{\mathbf{H}}\,.$$

represents the result of Arnoldi's algorithm applied to $\mathbf{A}$ and w_1.
Up to the last column, $\widetilde{\mathbf{H}}$ has to be $(s+2)$-band Hessenberg.

Idea of the second proof

V. Faber, J. Liesen and P. Tichý, 2008

Proof by contradiction. Let $\mathbf{A}$ admit an optimal $(s+2)$-term recurrence and $\mathbf{A}$ not be normal(s).

Then there exists a starting vector v such that $h_{1,d} \neq 0$.

$$\mathbf{A}\ \mathbf{V} \quad = \quad \mathbf{V} \begin{bmatrix} \bullet & \cdots & \bullet & & & \bullet \\ \bullet & \ddots & & \ddots & & \vdots \\ & \ddots & \ddots & & \bullet & \bullet \\ & & \ddots & \ddots & \vdots & \vdots \\ & & & \ddots & \bullet & \bullet \\ & & & & \bullet & \bullet \end{bmatrix}$$

Idea of the second proof

V. Faber, J. Liesen and P. Tichý, 2008

Proof by contradiction. Let $\mathbf{A}$ admit an optimal $(s+2)$-term recurrence and $\mathbf{A}$ not be normal(s).

Then there exists a starting vector v such that $h_{1,d} \neq 0$.

$$\mathbf{A}\,(\mathbf{VG}) \;=\; (\mathbf{VG})\,\mathbf{G}^{*} \begin{bmatrix} \bullet & \cdots & \bullet & & & \bullet \\ \bullet & \ddots & & \ddots & & \vdots \\ & \ddots & \ddots & & \bullet & \bullet \\ & & \ddots & \ddots & \vdots & \vdots \\ & & & \ddots & \bullet & \bullet \\ & & & & \bullet & \bullet \end{bmatrix} \mathbf{G}$$

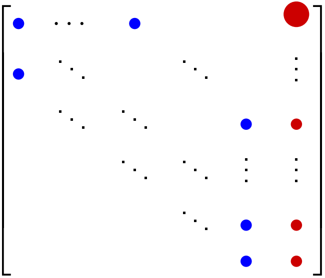
Idea of the second proof

V. Faber, J. Liesen and P. Tichý, 2008

Proof by contradiction. Let $\mathbf{A}$ admit an optimal $(s+2)$-term recurrence and $\mathbf{A}$ not be normal(s).

Then there exists a starting vector v such that $h_{1,d} \neq 0$.

$$\mathbf{A}\,(\mathbf{V}\mathbf{G}) \;=\; (\mathbf{V}\mathbf{G})\,\mathbf{G}^*$$



$$\mathbf{G}$$

Find unitary $\mathbf{G}$ (a product of Givens rotations) such that $\widetilde{\mathbf{H}}$ is unreduced upper Hessenberg, but $\widetilde{\mathbf{H}}$ is not $(s+2)$-band (up to the last column) - **contradiction**.

Summary

Generating an orthogonal basis of $\mathcal{K}_d(\mathbf{A}, v)$ via Arnoldi-type recurrence

Arnoldi-type recurrence
$(s+2)$-term

$$\Updownarrow$$

$\mathbf{A}$ is normal(s)
$\mathbf{A}^* = p(\mathbf{A})$

- When is $\mathbf{A}$ normal(s)?

Summary

Generating an orthogonal basis of $\mathcal{K}_d(\mathbf{A}, v)$ via Arnoldi-type recurrence

Arnoldi-type recurrence
$(s+2)$-term

$\Updownarrow$

$\mathbf{A}$ is normal(s)
$\mathbf{A}^* = p(\mathbf{A})$

- When is $\mathbf{A}$ normal(s)?
- $\mathbf{A}$ is normal and
  [Faber and Manteuffel, 1984],
  [Khavinson and Świątek, 2003]
  [Liesen and Strakoš, 2008]
  1. $s = 1$ if and only if the eigenvalues of $\mathbf{A}$ lie on a line in $\mathbb{C}$.
  2. If the eigenvalues of $\mathbf{A}$ are *not* on a line, then $d_{\min}(\mathbf{A}) \leq 3s - 2$.

Summary

Generating an orthogonal basis of $\mathcal{K}_d(\mathbf{A}, v)$ via Arnoldi-type recurrence

Arnoldi-type recurrence
$(s+2)$-term

$\Updownarrow$

$\mathbf{A}$ is normal(s)
$\mathbf{A}^* = p(\mathbf{A})$

$\updownarrow$

the only interesting case is $s = 1$,
collinear eigenvalues

- When is $\mathbf{A}$ normal(s)?
- $\mathbf{A}$ is normal and
  [Faber and Manteuffel, 1984],
  [Khavinson and Świątek, 2003]
  [Liesen and Strakoš, 2008]
  1. $s = 1$ if and only if the eigenvalues of $\mathbf{A}$ lie on a line in $\mathbb{C}$.
  2. If the eigenvalues of $\mathbf{A}$ are *not* on a line, then $d_{\min}(\mathbf{A}) \leq 3s - 2$.

Summary

Generating an orthogonal basis of $\mathcal{K}_d(\mathbf{A}, v)$ via Arnoldi-type recurrence

Arnoldi-type recurrence $(s+2)$-term

$\Updownarrow$

$\mathbf{A}$ is normal(s) $\mathbf{A}^* = p(\mathbf{A})$

$\updownarrow$

the only interesting case is $s = 1$, collinear eigenvalues

- When is $\mathbf{A}$ normal(s)?
- $\mathbf{A}$ is normal and
  [Faber and Manteuffel, 1984],
  [Khavinson and Świątek, 2003]
  [Liesen and Strakoš, 2008]
  1. $s = 1$ if and only if the eigenvalues of $\mathbf{A}$ lie on a line in $\mathbb{C}$.
  2. If the eigenvalues of $\mathbf{A}$ are *not* on a line, then $d_{\min}(\mathbf{A}) \leq 3s - 2$.
- All classes of "interesting" matrices are known.

Outline

Unitary matrices

Example

- Consider a unitary matrix $\mathbf{A}$ with different eigenvalues.

  $\mathbf{A}$ is normal $\implies$ $\mathbf{A}^*$ is a polynomial in $\mathbf{A}$

$$\mathbf{A}^* = p(\mathbf{A})\,.$$

Unitary matrices

Example

- Consider a unitary matrix $\mathbf{A}$ with different eigenvalues.

  $\mathbf{A}$ is normal $\Longrightarrow$ $\mathbf{A}^*$ is a polynomial in $\mathbf{A}$

$$\mathbf{A}^* = p(\mathbf{A})\,.$$

- The smallest degree of such polynomial is $n-1$ (n is the size of the matrix), i.e. $\mathbf{A}$ is normal$(n-1)$ [Liesen, 2007].

Unitary matrices

Example

- Consider a unitary matrix $\mathbf{A}$ with different eigenvalues.

  $\mathbf{A}$ is normal $\Longrightarrow$ $\mathbf{A}^*$ is a polynomial in $\mathbf{A}$

  $$\mathbf{A}^* = p(\mathbf{A})\,.$$

- The smallest degree of such polynomial is $n-1$ (n is the size of the matrix), i.e. $\mathbf{A}$ is normal$(n-1)$ [Liesen, 2007].

- Using Faber-Manteuffel theorem: generating orthogonal Krylov subspace bases for unitary matrices via the Arnoldi process would require a full recurrence.

Unitary matrices

Isometric Arnoldi process

- Gragg (1982) discovered the *isometric Arnoldi process*: Orthogonal Krylov subspace bases for unitary $\mathbf{A}$ can be generated by a 3-term recurrence of the form

$$v_{j+1} = \beta_{j,j}\mathbf{A}v_j - \beta_{j-1,j}\mathbf{A}v_{j-1} - \sigma_{j,j}v_{j-1}$$

  (stable implementation - two coupled 2-term recurrences).

- Used for solving unitary eigenvalue problems and linear systems with shifted unitary matrices [Jagels and Reichel, 1994].

- This short recurrence is not of the "Arnoldi-type".

Generalization: (ℓ, m)-recursion

Barth and Manteuffel, 2000

Generate an orthogonal basis via the (ℓ, m)-recursion of the form

$$v_{j+1} = \sum_{i=j-m}^{j} \beta_{i,j}\, \mathbf{A}\, v_i \;-\; \sum_{i=j-\ell}^{j} \sigma_{i,j}\, v_i \,, \tag{1}$$

- $(\ell, m) = (0, 1)$ if $\mathbf{A}$ is unitary,
  $(\ell, m) = (1, 1)$ if $\mathbf{A}$ is shifted unitary.

Generalization: (ℓ, m)-recursion

Barth and Manteuffel, 2000

Generate an orthogonal basis via the (ℓ, m)-recursion of the form

$$(1) \qquad v_{j+1} = \sum_{i=j-m}^{j} \beta_{i,j}\, \mathbf{A}\, v_i \;-\; \sum_{i=j-\ell}^{j} \sigma_{i,j}\, v_i \,,$$

- $(\ell, m) = (0, 1)$ if $\mathbf{A}$ is unitary,
  $(\ell, m) = (1, 1)$ if $\mathbf{A}$ is shifted unitary.

- A sufficient condition [Barth and Manteuffel, 2000]:
  $\mathbf{A}^*$ is a rational function in $\mathbf{A}$,

  $$\mathbf{A}^* = r(\mathbf{A})\,,$$

  where $r = p/q$, p and q have degrees ℓ and m.

  Example: Unitary matrices, $\mathbf{A}^* = \mathbf{A}^{-1}$, i.e. $r = 1/z$.

  Matrices $\mathbf{A}$ such that $\mathbf{A}^* = r(\mathbf{A})$ are called normal(ℓ, m).

Degree of a rational function, degrees of normality

normal degree of $\mathbf{A}$, McMillan degree of $\mathbf{A}$

Definition. McMillan degree of a rational function $r = p/q$ where p and q are relatively prime is defined as

$$\deg r = \max\{\deg p, \deg q\}.$$

Degree of a rational function, degrees of normality

normal degree of $\mathbf{A}$, McMillan degree of $\mathbf{A}$

Definition. McMillan degree of a rational function $r = p/q$ where p and q are relatively prime is defined as

$$\deg r = \max\{\deg p, \deg q\}.$$

Definition. Let $\mathbf{A}$ be a diagonalizable matrix.

- $d_p(\mathbf{A})$ ... normal degree of $\mathbf{A}$
  the smallest degree of a polynomial p that satisfies

$$p(\lambda) = \overline{\lambda} \quad \text{for all eigenvalues} \quad \lambda \quad \text{of} \quad \mathbf{A}\,.$$

Degree of a rational function, degrees of normality

normal degree of $\mathbf{A}$, McMillan degree of $\mathbf{A}$

Definition. McMillan degree of a rational function $r = p/q$ where p and q are relatively prime is defined as

$$\deg r = \max\{\deg p, \deg q\}.$$

Definition. Let $\mathbf{A}$ be a diagonalizable matrix.

- $d_p(\mathbf{A})$ ... normal degree of $\mathbf{A}$
  the smallest degree of a polynomial p that satisfies

  $$p(\lambda) = \overline{\lambda} \text{ for all eigenvalues } \lambda \text{ of } \mathbf{A}.$$

- $d_r(\mathbf{A})$ ... McMillan degree of $\mathbf{A}$
  the smallest McMillan degree of a rational function r that satisfies

  $$r(\lambda) = \overline{\lambda} \text{ for all eigenvalues } \lambda \text{ of } \mathbf{A}.$$

When is $\mathbf{A}$ a low degree rational function in $\mathbf{A}$?

Collinear or concyclic eigenvalues

Application of results from rational interpolation theory:

Theorem. [Liesen, 2007] Let $\mathbf{A}$ be a diagonalizable matrix with $k \geq 4$ distinct eigenvalues.

- If the eigenvalues are collinear, then $d_r(\mathbf{A}) = d_p(\mathbf{A}) = 1$.
- If the eigenvalues are concyclic, then $d_r(\mathbf{A}) = 1$, $d_p(\mathbf{A}) = k - 1$.
- In all other cases $d_r(\mathbf{A}) > \frac{k}{5}$, $d_p(\mathbf{A}) > \frac{k}{3}$.

Summary

Generating an orthogonal basis of $\mathcal{K}_k(\mathbf{A}, v)$ via short recurrences

Arnoldi-type recurrence $(s+2)$-term	Barth-Manteuffel (ℓ, m)-recursion
$\Updownarrow$	$\Uparrow$
$\mathbf{A}$ is normal(s) $\mathbf{A}^* = p(\mathbf{A})$	$\mathbf{A}$ is normal(ℓ, m) $\mathbf{A}^* = r(\mathbf{A})$
$\updownarrow$	$\updownarrow$
the only interesting case is $s = 1$, collinear eigenvalues	the only interesting cases are $(0, 1)$ or $(1, 1)$ concyclic eigenvalues

Outline

The role of the matrix $\mathbf{B}$

Generating a $\mathbf{B}$-orthogonal basis

Let $\mathbf{B} \in \mathbb{C}^{n\times n}$ be a Hermitian positive definite (HPD), defining the $\mathbf{B}$-inner product, $\langle x, y\rangle_{\mathbf{B}} \equiv y^*\mathbf{B}x$.

The role of the matrix $\mathbf{B}$

Generating a $\mathbf{B}$-orthogonal basis

Let $\mathbf{B} \in \mathbb{C}^{n \times n}$ be a Hermitian positive definite (HPD), defining the $\mathbf{B}$-inner product, $\langle x, y \rangle_{\mathbf{B}} \equiv y^* \mathbf{B} x$.

$\mathbf{B}$-normal(s) matrices: there exist a polynomial p_s of the smallest possible degree s such that

$$\mathbf{A}^+ \equiv \mathbf{B}^{-1} \mathbf{A}^* \mathbf{B} = p_s(\mathbf{A}),$$

where $\mathbf{A}^+$ the $\mathbf{B}$-adjoint of $\mathbf{A}$.

The role of the matrix $\mathbf{B}$

Generating a $\mathbf{B}$-orthogonal basis

Let $\mathbf{B} \in \mathbb{C}^{n\times n}$ be a Hermitian positive definite (HPD), defining the $\mathbf{B}$-inner product, $\langle x, y\rangle_{\mathbf{B}} \equiv y^*\mathbf{B}x$.

$\mathbf{B}$-normal(s) matrices: there exist a polynomial p_s of the smallest possible degree s such that

$$\mathbf{A}^+ \equiv \mathbf{B}^{-1}\mathbf{A}^*\mathbf{B} = p_s(\mathbf{A}),$$

where $\mathbf{A}^+$ the $\mathbf{B}$-adjoint of $\mathbf{A}$.

Theorem. [Faber and Manteuffel, 1984], [Liesen and Strakoš, 2008]
For $\mathbf{A}$, $\mathbf{B}$ as above, and an integer $s \geq 0$ with $s + 2 < d_{\min}(\mathbf{A})$:

$\mathbf{A}$ admits for the given $\mathbf{B}$ an optimal $(s+2)$-term recurrence if and only if $\mathbf{A}$ is $\mathbf{B}$-normal(s).

The role of the matrix $\mathbf{B}$

Characterization of $\mathbf{B}$-normal(s) matrices

Theorem. [Liesen and Strakoš, 2008]

$\mathbf{A}$ is $\mathbf{B}$-normal(s) if and only if

1. $\mathbf{A}$ is diagonalizable ($\mathbf{A} = \mathbf{W}\mathbf{\Lambda}\mathbf{W}^{-1}$), and
2. $\mathbf{B} = (\mathbf{W}\mathbf{D}\mathbf{W}^*)^{-1}$, where $\mathbf{D}$ is HPD and block diagonal with blocks corresponding to those of $\mathbf{\Lambda}$, and
3. $\mathbf{\Lambda}^* = p_s(\mathbf{\Lambda})$ for a polynomial p_s of (smallest possible) degree s.

Theorem. [Liesen and Strakoš, 2008]

$\mathbf{A}$ is $\mathbf{B}$-normal(s) if and only if

1. $\mathbf{A}$ is diagonalizable ($\mathbf{A} = \mathbf{W}\mathbf{\Lambda}\mathbf{W}^{-1}$), and
2. $\mathbf{B} = (\mathbf{W}\mathbf{D}\mathbf{W}^{*})^{-1}$, where $\mathbf{D}$ is HPD and block diagonal with blocks corresponding to those of $\mathbf{\Lambda}$, and
3. $\mathbf{\Lambda}^{*} = p_s(\mathbf{\Lambda})$ for a polynomial p_s of (smallest possible) degree s.

The only interesting case: $\mathbf{B}$-normal(1) matrices

- If $\mathbf{A}$ is diagonalizable and the eigenvalues are collinear, then there exists $\mathbf{B}$ such that $\mathbf{A}$ is $\mathbf{B}$-normal(1).
- Find a preconditioner $\mathbf{P}$ so that $\mathbf{PA}$ is $\mathbf{B}$-normal(1) for some $\mathbf{B}$, e.g. [Concus and Golub, 1978], [Widlund, 1978].

Conclusions

We characterized matrices for which it is possible to generate an orthogonal basis of Krylov subspaces using short recurrences (normal(s), normal(ℓ, m)).

Conclusions

We characterized matrices for which it is possible to generate an orthogonal basis of Krylov subspaces using short recurrences (normal(s), normal(ℓ, m)).

We presented ideas of a new proof of the Faber-Manteuffel theorem.

Conclusions

We characterized matrices for which it is possible to generate an orthogonal basis of Krylov subspaces using short recurrences (normal(s), normal(ℓ, m)).

We presented ideas of a new proof of the Faber-Manteuffel theorem.

Practical cases:

- $\mathbf{A}$ is normal and the eigenvalues are collinear or concyclic.

Conclusions

We characterized matrices for which it is possible to generate an orthogonal basis of Krylov subspaces using short recurrences (normal(s), normal(ℓ, m)).

We presented ideas of a new proof of the Faber-Manteuffel theorem.

Practical cases:

- $\mathbf{A}$ is normal and the eigenvalues are collinear or concyclic.
- If eigenvalues of $\mathbf{A}$ are collinear or concyclic, then there exists a HPD matrix $\mathbf{B}$ such that $\mathbf{A}$ admits short recurrences for generating a $\mathbf{B}$-orthogonal basis.

Conclusions

We characterized matrices for which it is possible to generate an orthogonal basis of Krylov subspaces using short recurrences (normal(s), normal(ℓ, m)).

We presented ideas of a new proof of the Faber-Manteuffel theorem.

Practical cases:

- $\mathbf{A}$ is normal and the eigenvalues are collinear or concyclic.
- If eigenvalues of $\mathbf{A}$ are collinear or concyclic, then there exists a HPD matrix $\mathbf{B}$ such that $\mathbf{A}$ admits short recurrences for generating a $\mathbf{B}$-orthogonal basis.
- Find a preconditioner $\mathbf{P}$ so that $\mathbf{PA}$ is $\mathbf{B}$-normal(1) ($\mathbf{B}$-normal$(0,1)$, $\mathbf{B}$-normal$(1,1)$) for some $\mathbf{B}$.

Related papers

- J. Liesen and Z. Strakoš, [On optimal short recurrences for generating orthogonal Krylov subspace bases, SIAM Review, 50, 2008, pp. 485-503].
  The completely reworked theory of short recurrences for generating orthogonal Krylov subspace bases.
- V. Faber, J. Liesen and P. Tichý, [The Faber-Manteuffel Theorem for Linear Operators, SIAM Journal on Numerical Analysis, Volume 46, 2008, pp. 1323-1337.]
  New proofs of the fundamental theorem of Faber and Manteuffel.
- J. Liesen, [When is the adjoint of a matrix a low degree rational function in the matrix? SIAM J. Matrix Anal. Appl., 2007 , 29 , 1171-1180].
  A nice application of results from rational approximation theory.

More details

More details can be found at

http://www.cs.cas.cz/~tichy
http://www.math.tu-berlin.de/~liesen
http://www.cs.cas.cz/~strakos

More details

More details can be found at

http://www.cs.cas.cz/~tichy
http://www.math.tu-berlin.de/~liesen
http://www.cs.cas.cz/~strakos

Thank you for your attention!