The Faber-Manteuffel Theorem and its Consequences

Petr Tichý

joint work with

Vance Faber, Jörg Liesen

Czech Academy of Sciences

July 21, 2011
ICIAM 2011, Vancouver, BC, Canada

Optimal Krylov subspace methods

and low memory requirements?

- Consider a system of linear algebraic equations

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

$\mathbf{A} \in \mathbb{R}^{n \times n}$ is nonsingular, $b \in \mathbb{R}^n$.

- Given x_0, find an *optimal*

$$x_j \in x_0 + \mathcal{K}_j(\mathbf{A}, r_0)$$

so that the error is minimized in a given vector norm.

- What are necessary and sufficient conditions on $\mathbf{A}$ so that the optimal x_j can be computed using short recurrences? (only a constant number of vectors is needed)

Examples of optimal Krylov subspace methods

with short recurrences

CG [Hestenes, Stiefel 1952], MINRES, SYMMLQ [Paige, Saunders 1975]

- *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,

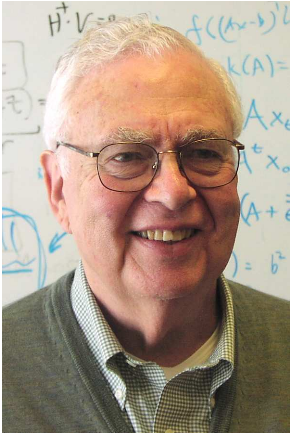
  $\| x - x_j \|$ in SYMMLQ - here $x_j \in x_0 + \mathbf{A}\mathcal{K}_j(\mathbf{A}, r_0)$.

- Generate orthogonal (or $\mathbf{A}$-orthogonal) Krylov subspace basis using a three-term recurrence,

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

- An important assumption: $\mathbf{A}$ is symmetric (MINRES, SYMMLQ) and positive 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 Symposium) held in Oxford in 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, $\| \cdot \|_{\mathbf{B}} = \langle \cdot, \cdot \rangle_{\mathbf{B}}^{1/2}$.
- 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 $\mathcal{K}_{j+1}(\mathbf{A}, r_0)$.

Optimal Krylov subspace method with short recurrences

The question about

> the existence of an optimal Krylov subspace method with short recurrences

can be reduced to the question:

> For which $\mathbf{A}$ is it possible to generate a $\mathbf{B}$-orthogonal basis of the Krylov subspace using short recurrences?

(for each initial starting vector)

Faber, 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

1. The Faber-Manteuffel theorem
2. Ideas of a new proof
3. Consequences
4. Other types of recurrences

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 multiplication 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. $\mathrm{span}\{v_1, \ldots, v_j\} = \mathcal{K}_j(\mathbf{A}, v),$ for $j = 1, \ldots, d$,
  2. $\langle v_i, v_j \rangle = 0,$ for $i \neq j$, $i, j = 1, \ldots, d$.

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

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

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

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: $$v_{j+1} = \mathbf{A}\,v_j - \sum_{\mathbf{i=j-s}}^{j} h_{i,j}v_i\,.$$

Formulation of the problem

Optimal short recurrences (Definition - Liesen, 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.

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

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

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, Manteuffel 1984], [Liesen, Strakoš 2008]

Given nonsingular $\mathbf{A}$ and nonnegative s, $s + 2 < d_{\min}(\mathbf{A})$.

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

if and only if

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

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

A new proof of the Faber-Manteuffel theorem

- Motivated by the paper [Liesen, 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."

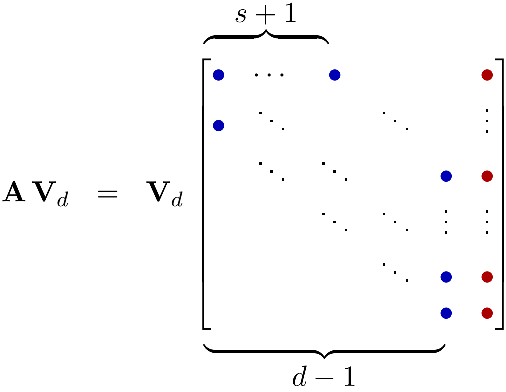
- In [Faber, Liesen, T. 2008] we give two new proofs of the Faber-Manteuffel theorem that use more elementary tools.

Extension of $\mathbf{A}\,\mathbf{V}_{d-1} = \mathbf{V}_d\,\mathbf{H}_{d,d-1}$

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

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

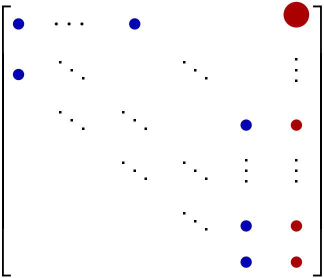
$$\mathbf{A}v_d = \sum_{i=1}^{d} h_{i,d}\, v_i\,.$$



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

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 $\mathbf{G}$ such that $\mathbf{G}^*\mathbf{H}\mathbf{G}$ is unreduced upper Hessenberg, but $\mathbf{G}^*\mathbf{H}\mathbf{G}$ is not $(s+2)$-band (up to the last column).

Faber-Manteuffel Theorem – 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, Manteuffel 1984],

  [Khavinson, Świątek 2003]

  [Liesen, Strakoš 2008]

  1. $s = 1$ if and only if the eigenvalues of $\mathbf{A}$ lie on a line in $\mathbb{C}$.
  2. For $s > 1$, $\mathbf{A}$ has at most $3s - 2$ different eigenvalues.
- All classes of “interesting” matrices are known.

When is $\mathbf{A}$ orthogonally reducible
to $(s+2)$-band Hessenberg form?

The matrix representation of the Arnoldi algorithm can be extended by one column to

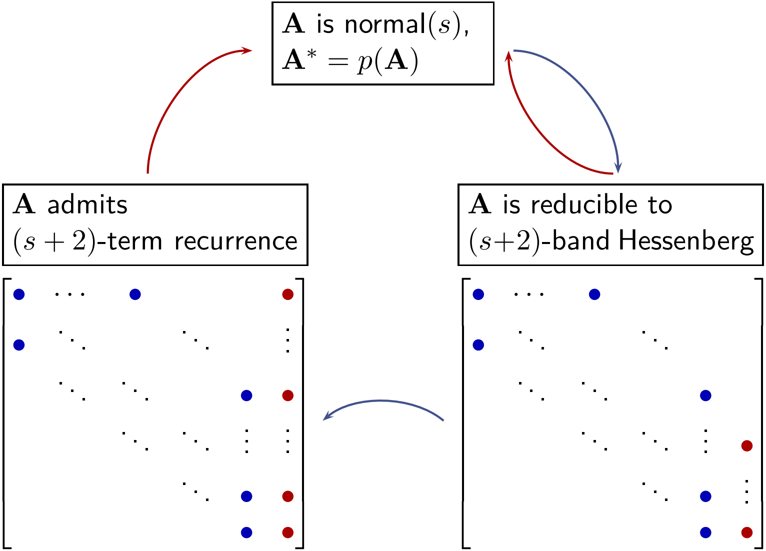
$$\mathbf{A}\,\mathbf{V}_d \;=\; \mathbf{V}_d\,\mathbf{H}_d$$

where $\mathbf{H}_d \in \mathbb{C}^{d\times d}$ is unreduced upper Hessenberg matrix.

We say that $\mathbf{A}$ is orthogonally reducible to $(s+2)$-band Hessenberg form if $\mathbf{H}_d$ is $(s+2)$-band Hessenberg matrix for each starting vector v_1.

What are necessary and sufficient conditions on $\mathbf{A}$ to be orthogonally reducible to $(s+2)$-band Hessenberg form?

When is $\mathbf{A}$ orthogonally reducible
to $(s+2)$-band Hessenberg form?



When is $\mathbf{A}$ orthogonally reducible

to $(s+2)$-band Hessenberg form?

Theorem [Liesen, Strakoš 2008]

Let s be a nonnegative integer, $s+2 < d_{\min}(\mathbf{A})$. Then the following three assertions are equivalent:

1. $\mathbf{A}$ admits an optimal $(s+2)$-term recurrence.
2. $\mathbf{A}$ is normal(s).
3. $\mathbf{A}$ is orthogonally reducible to $(s+2)$-band Hessenberg form.

- $1 \iff 2$: [Faber, Manteuffel 1984].
- $2 \iff 3$: a simple proof in [Faber, Liesen, T. 2009].
- The subtle difference between 1. and 3. $\to$ source of confusions [Voevodin, Tyrtyshnikov 1981], [Liesen, Saylor 2005].

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

Faber-Manteuffel theorem

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 exists 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, Manteuffel 1984], [Liesen, 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).

- If $\mathbf{A}$ is diagonalizable and the eigenvalues are collinear, then there exists an HPD $\mathbf{B}$ such that $\mathbf{A}$ is $\mathbf{B}$-normal(1).
  [Liesen, Strakoš 2008] $\rightarrow$ complete parametrization of all $\mathbf{B}$'s.
- Find a preconditioner $\mathbf{P}$ so that $\mathbf{PA}$ is $\mathbf{B}$-normal(1) for some $\mathbf{B}$, e.g. [Concus, Golub 1976], [Widlund 1978], [Eisenstat 1983], [Bramble, Pasciak 1988], [Stoll, Wathen 2008].
- Saddle point matrix:

$$\mathbf{A} = \begin{bmatrix} A_1 & A_2^T \\ -A_2 & A_3 \end{bmatrix}, \qquad \mathbf{B}_\gamma = \begin{bmatrix} A_1 - \gamma I_m & A_2^T \\ A_2 & \gamma I_k - A_3 \end{bmatrix}$$

where $A_1 = A_1^T > 0$, $A_3 = A_3^T \geq 0$, A_2 full rank.

This matrix satisfies $\mathbf{B}_\gamma^{-1} \mathbf{A}^T \mathbf{B}_\gamma = \mathbf{A}$.

How to choose γ such that $\mathbf{B}_\gamma$ is positive definite?
[Fischer et al. 1998], [Benzi, Simoncini 2006], [Liesen, Parlett 2007].

Other types of recurrences

The existence of an optimal Krylov subspace method with short recurrences

For which $\mathbf{A}$ is it possible to generate an orthogonal basis of the Krylov subspace using short recurrences?

- We can use a different kind of recurrences than Arnoldi-like.
- For (shifted) unitary matrices: Isometric Arnoldi process [Gragg 1982; Jagels, Reichel 1994].
- Generalized by [Barth, Manteuffel 2000] to (ℓ, m)-recursion.
  A sufficient condition: $\mathbf{A}^*$ is a low degree rational func. of $\mathbf{A}$.
  Practical use: matrices with concyclic eigenvalues [Liesen 2007].
- [Barth, Manteuffel 2000]: Short multiple recursion for $\mathbf{A}$ such that $\Delta \equiv \mathbf{A}^* q_m(\mathbf{A}) - p_\ell(\mathbf{A})$ has low rank.
- [Beckermann, Reichel 2008]: GMRES-like algorithm with short recurrences for $\mathbf{A}$ such that $\Delta \equiv \mathbf{A}^* - \mathbf{A}$ is of low rank.
  Application: Path following methods.

Conclusions

- We characterized matrices for which it is possible to generate an orthogonal basis of Krylov subspaces via short recurrences.
- We presented ideas of a new proof of the Faber-Manteuffel theorem and studied its consequences.
- Practical case: If the eigenvalues of $\mathbf{A}$ are collinear or concyclic, then there exists an HPD matrix $\mathbf{B}$ such that $\mathbf{A}$ admits short recurrences for generating a $\mathbf{B}$-orthogonal basis.
- Examples: Find a preconditioner $\mathbf{P}$ so that short recurrences exist for $\mathbf{PA}$, saddle point matrices.

An interesting case to study:

- Short multiple recursion for $\mathbf{A}$ such that $\mathbf{A}^* q_m(\mathbf{A}) - p_\ell(\mathbf{A})$ has low rank. Practical cases? Algorithmic realizations?

Related papers

- V. Faber and T. Manteuffel, [Necessary and sufficient conditions for the existence of a conjugate gradient method, SIAM J. Numer. Anal., 21 (1984), pp. 352–362.]
- T. Barth and T. Manteuffel, [Multiple recursion conjugate gradient algorithms. I. Sufficient conditions, SIAM J. Matrix Anal. Appl., 21 (2000), pp. 768–796.]
- J. Liesen and Z. Strakoš, [On optimal short recurrences for generating orthogonal Krylov subspace bases, SIAM Review, 50, 2008, pp. 485-503].
- 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].
- V. Faber, J. Liesen and P. Tichý, [The Faber-Manteuffel Theorem for Linear Operators, SIAM J. Numer. Anal., 46 (2008), pp. 1323-1337.]
- V. Faber, J. Liesen, and P. Tichý, [On orthogonal reduction to Hessenberg form with small bandwidth, Numer. Algorithms, 51 (2009), pp. 133–142.]

Thank you for your attention!