On optimal short recurrences for generating orthogonal Krylov subspace bases

Jörg Liesen

based on joint work with

Zdeněk Strakoš and Petr Tichý (Czech Academy of Sciences),

Vance Faber (BD Biosciences, Seattle),

Beresford Parlett (Berkeley), and Paul Saylor (Illinois)

Overview

1. Introduction: Krylov subspace methods
2. Optimal short recurrences
3. Characterization and examples

Introduction: Krylov subspace methods (1)

- Methods that are based on *projection onto the Krylov subspaces*

$$\mathcal{K}_n(A, v_1) \equiv \text{span}\{v_1, Av_1, \ldots, A^{n-1}v_1\}, \quad n = 1, 2, \ldots,$$

  where A is a given square matrix and v_1 is the initial vector.

- Must generate bases of $\mathcal{K}_n(A, v_1)$, $n = 1, 2, \ldots$.
- Trivial choice: $v_1, Av_1, \ldots, A^{n-1}v_1$.
  This is computationally infeasible (recall the Power Method).

Introduction: Krylov subspace methods (1)

- Methods that are based on *projection onto the Krylov subspaces*

$$\mathcal{K}_n(A, v_1) \equiv \text{span}\{v_1, Av_1, \dots, A^{n-1}v_1\}, \quad n = 1, 2, \dots,$$

  where A is a given square matrix and v_1 is the initial vector.

- Must generate bases of $\mathcal{K}_n(A, v_1)$, $n = 1, 2, \dots$.
- Trivial choice: $v_1, Av_1, \dots, A^{n-1}v_1$.
  This 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.*
- First such method for $Ax = b$:
  Conjugate gradient (CG) method of Hestenes and Stiefel (1952).

Introduction: Krylov subspace methods (2)

In case the matrix A is symmetric and positive definite, the following formulas are used in the conjugate gradient method.

(3:1a) $$p_o = r_o = k - Ax_o \qquad (x_o \text{ arbitrary})$$

(3:1b) $$a_i = \frac{|r_i|^2}{(p_i, Ap_i)}$$

(3:1c) $$x_{i+1} = x_i + a_i p_i$$

(3:1d) $$r_{i+1} = r_i - a_i A p_i$$

(3:1e) $$b_i = \frac{|r_{i+1}|^2}{|r_i|^2}$$

(3:1f) $$p_{i+1} = r_{i+1} + b_i p_i .$$

The classical CG method of Hestenes and Stiefel

(US National Bureau of Standards Preprint No. 1659, March 10, 1952)

The residual vectors $r_0, r_1, \ldots, r_{n-1}$ are generated by a short recurrence and form an orthogonal basis of $\mathcal{K}_n(A, r_0)$.

Introduction: Krylov subspace methods (3)

- CG is for symmetric positive definite A.
- (Paige and Saunders, 1975):
  Short recurrence & orthogonal basis methods for symmetric A.

SIAM J. NUMER. ANAL.
Vol. 12, No. 4, September 1975

SOLUTION OF SPARSE INDEFINITE SYSTEMS OF LINEAR EQUATIONS*

C. C. PAIGE† AND M. A. SAUNDERS‡

Abstract. The method of conjugate gradients for solving systems of linear equations with a symmetric positive definite matrix A is given as a logical development of the Lanczos algorithm for tridiagonalizing A. This approach suggests numerical algorithms for solving such systems when A is symmetric but indefinite. These methods have advantages when A is large and sparse.

Introduction: Krylov subspace methods (4)

- By the end of the 1970s it was unknown if such methods existed also for general unsymmetric A.
- *Gene Golub posed this fundamental question* at Gatlinburg VIII (now Householder VIII) held in Oxford from July 5 to 11, 1981:

SIGNUM
NEWSLETTER

A Quarterly Publication of the ACM Special Interest Group on Numerical Mathematics

December 1981
Volume 16, Number 4

It is not uncommon for speakers to pose an open problem to their audience during their talk. What is unprecedented is for such a proposition to be backed by some stake money. Perhaps a new era in numerical analysis sponsorship began at Gatlinburg VIII when three of the participants offered prize money for the solution of problems which they posed during the meeting. I list the challenges below so that you can make a bid for glory (or cash) by solving them.

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 does this mean?

Introduction: Krylov subspace methods (5)

- We want to solve $Ax = b$ iteratively, starting from x_0.
- Step $n = 1, 2, \ldots$: $x_n = x_{n-1} + \alpha_{n-1} p_{n-1}$,
  direction vector p_{n-1}, scalar α_{n-1} (both to be determined).
- Krylov subspace method:
  $\mathrm{span}\{p_0, \ldots, p_{n-1}\} = \mathcal{K}_n(A, v_1) \quad (v_1 = r_0 = b - Ax_0)$.
- CG-like descent method:
  Error is minimized in some given inner product norm, $\| \cdot \|_B = \langle \cdot, \cdot \rangle_B^{1/2}$.

Introduction: Krylov subspace methods (5)

- We want to solve $Ax = b$ iteratively, starting from x_0.
- Step $n = 1, 2, \ldots$: $x_n = x_{n-1} + \alpha_{n-1} p_{n-1}$,
  direction vector p_{n-1}, scalar α_{n-1} (both to be determined).
- Krylov subspace method:
  $\mathrm{span}\{p_0, \ldots, p_{n-1}\} = \mathcal{K}_n(A, v_1) \quad (v_1 = r_0 = b - Ax_0)$.
- CG-like descent method:
  Error is minimized in some given inner product norm, $\|\cdot\|_B = \langle \cdot, \cdot \rangle_B^{1/2}$.
- $\|x - x_n\|_B$ is minimal iff $x - x_n \perp_B \mathrm{span}\{p_0, \ldots, p_{n-1}\}$.
- By construction, this is satisfied iff

$$\alpha_{n-1} = \frac{\langle x - x_{n-1}, p_{n-1} \rangle_B}{\langle p_{n-1}, p_{n-1} \rangle_B} \quad \text{and} \quad \langle p_{n-1}, p_j \rangle_B = 0, \;\; j = 0, \ldots, n-2,$$

i.e. $p_0, \ldots, p_{n-1}$ *must be a* B*-orthogonal basis of* $\mathcal{K}_n(A, v_1)$.

Introduction: Krylov subspace methods (6)

- Faber and Manteuffel answered Golub's question in 1984:
  For a general matrix A there exists *no* CG-like descent method (related results by Voevodin and Tyrtyshnikov in the early 1980s).

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 $B^*=-B$.

What are the details of this result?

Optimal short recurrences (1)

Notation:

- Matrix $A \in \mathbb{C}^{N \times N}$, nonsingular.
- Matrix $B \in \mathbb{C}^{N \times N}$, Hermitian positive definite (HPD), defining the B-inner product, $\langle x, y \rangle_B \equiv y^* B x$.
- Initial vector $v_1 \in \mathbb{C}^N$.
- $d = d(A, v_1)$, the grade of v_1 with respect to A,

$$\mathcal{K}_1(A, v_1) \subset \ldots \subset \mathcal{K}_d(A, v_1) = \mathcal{K}_{d+1}(A, v_1) = \ldots = \mathcal{K}_N(A, v_1)\,.$$

Optimal short recurrences (1)

Notation:

- Matrix $A \in \mathbb{C}^{N \times N}$, nonsingular.
- Matrix $B \in \mathbb{C}^{N \times N}$, Hermitian positive definite (HPD), defining the B-inner product, $\langle x, y \rangle_B \equiv y^* B x$.
- Initial vector $v_1 \in \mathbb{C}^N$.
- $d = d(A, v_1)$, the grade of v_1 with respect to A,

$$\mathcal{K}_1(A, v_1) \subset \ldots \subset \mathcal{K}_d(A, v_1) = \mathcal{K}_{d+1}(A, v_1) = \ldots = \mathcal{K}_N(A, v_1)\,.$$

Our goal:
Generate a B-orthogonal basis $v_1, \ldots, v_d$ of $\mathcal{K}_d(A, v_1)$.

1. $\text{span}\{v_1, \ldots, v_n\} = \mathcal{K}_n(A, v_1), \quad \text{for} \quad n = 1, \ldots, d,$
2. $\langle v_j, v_k \rangle_B = 0, \quad \text{for} \quad j \neq k, \quad j, k = 1, \ldots, d.$

Optimal short recurrences (2)

- Standard way for generating the B-orthogonal basis: Arnoldi's method.

$$\begin{aligned} v_{n+1} &= Av_n - \sum_{m=1}^{n} h_{m,n} v_m \,, \quad n = 1, \ldots, d-1 \,, \\ h_{m,n} &= \frac{\langle Av_n, v_m \rangle_B}{\langle v_m, v_m \rangle_B} \,, \quad d = \dim \mathcal{K}_N(A, v_1) \,. \end{aligned}$$

(No normalization for convenience.)

Optimal short recurrences (2)

- Standard way for generating the B-orthogonal basis: Arnoldi's method.

$$
\begin{aligned}
v_{n+1} &= Av_n - \sum_{m=1}^{n} h_{m,n} v_m \,, \quad n = 1, \ldots, d-1 \,, \\
h_{m,n} &= \frac{\langle Av_n, v_m \rangle_B}{\langle v_m, v_m \rangle_B} \,, \quad d = \dim \mathcal{K}_N(A, v_1) \,.
\end{aligned}
$$

(No normalization for convenience.)

- Rewritten in matrix notation: $AV_{d-1} = V_d H_{d,d-1}$, where

matrix of size $d \times (d-1)$

$$
V_d \equiv [v_1, \ldots, v_d], \qquad H_{d,d-1} \equiv \begin{bmatrix} h_{1,1} & \cdots & h_{1,d-1} \\ 1 & \ddots & \vdots \\ & \ddots & h_{d-1,d-1} \\ & & 1 \end{bmatrix}
$$

$$
V_d^* B V_d \ \text{ is diagonal} \,, \quad d = \dim \mathcal{K}_N(A, v_1) \,.
$$

Optimal short recurrences (3)

- The *full recurrence* in Arnoldi's method,

$$v_{n+1} = Av_n - \sum_{m=1}^{n} h_{m,n} v_m \,, \quad n = 1, \ldots, d-1 \,,$$

is an *optimal* $(s+2)$*-term recurrence* when

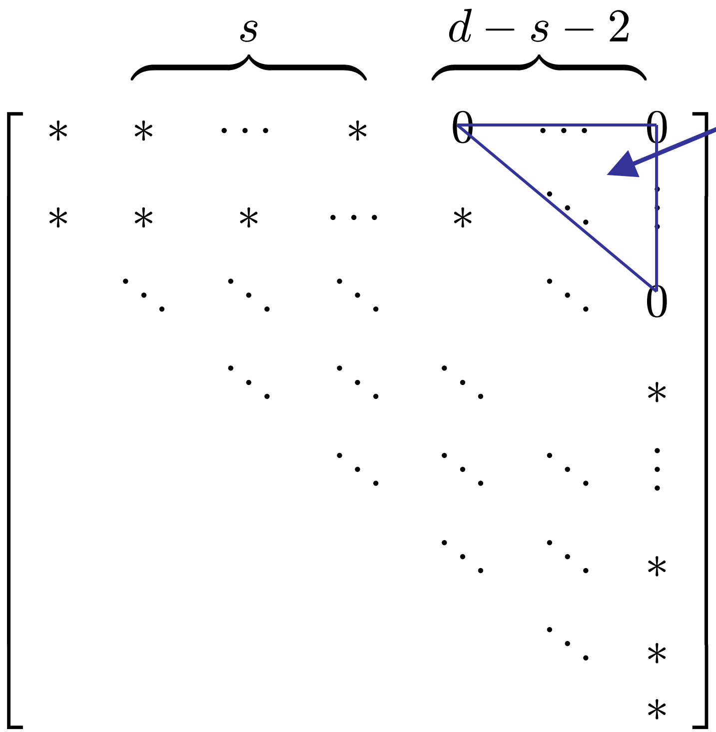
$$v_{n+1} = Av_n - \sum_{m=n-s}^{n} h_{m,n} v_m \,, \quad n = 1, \ldots, d-1 \,.$$

- For $s = 1$: Optimal 3-term recurrence,

$$v_{n+1} \; = \; Av_n - h_{n,n} v_n - h_{n-1,n} v_{n-1}$$

- Why *optimal*?
  1. Only one multiplication with A is performed.
  2. Only the previous $s+1$ vectors are required.

Optimal short recurrences (4)

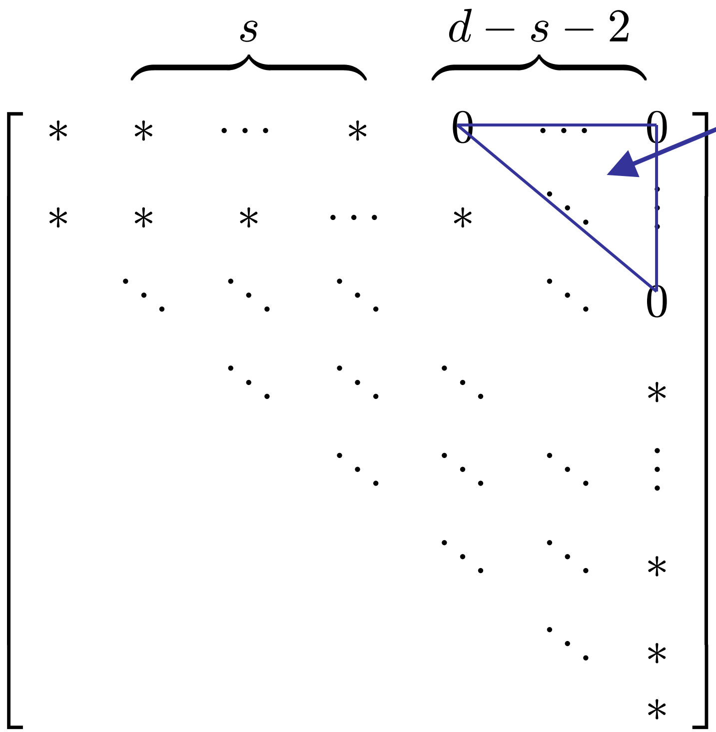


largest upper triangle that is zero

Optimal $(s+2)$-term recurrence:
$H_{d,d-1}$ *is* $(s+2)$*-band Hessenberg*

(e.g. 3-band Hessenberg = tridiagonal)

Optimal short recurrences (4)



Optimal $(s+2)$-term recurrence:
$H_{d,d-1}$ *is* $(s+2)$*-band Hessenberg*

(e.g. 3-band Hessenberg = tridiagonal)

Definition (L. and Strakoš, 2007)

Given A, B as above and a nonnegative integer s with $s+2 \leq d_{\min}(A)$.
($d_{\min}(A)$ = degree of A's minimal polynomial.)

Then A *admits for* B *an optimal* $(s+2)$*-term recurrence*, if

- for any v_1 the matrix $H_{d,d-1}$ is at most $(s+2)$-band Hessenberg, and
- for at least one v_1 the matrix $H_{d,d-1}$ is $(s+2)$-band Hessenberg.

Optimal short recurrences (5)

- $H_{d,d-1}$ is at most $(s+2)$-band Hessenberg if

$$h_{m,n} = \frac{\langle Av_n, v_m \rangle_B}{\langle v_m, v_m \rangle_B} = 0 \,, \quad \text{for} \quad n > m + s \,, \quad n = 1, \ldots, d-1 \,.$$

- Therefore: $h_{m,n} = 0 \iff 0 = \langle Av_n, v_m \rangle_B = \langle v_n, A^+ v_m \rangle_B$, where $A^+ \equiv B^{-1} A^* B$ is the *B-adjoint of A*.

Optimal short recurrences (5)

- $H_{d,d-1}$ is at most $(s+2)$-band Hessenberg if

$$h_{m,n} = \frac{\langle Av_n, v_m \rangle_B}{\langle v_m, v_m \rangle_B} = 0\,, \quad \text{for} \quad n > m+s\,, \quad n = 1, \dots, d-1\,.$$

- Therefore: $h_{m,n} = 0 \iff 0 = \langle Av_n, v_m \rangle_B = \langle v_n, A^+ v_m \rangle_B$, where $A^+ \equiv B^{-1} A^* B$ is the B*-adjoint of* A.

- If $A^+ = p_s(A)$ for a polynomial of degree s, then $A^+ v_m \in \mathcal{K}_{m+s}(A, v_1)$.
- Then for $n > m+s$: $v_n \perp_B A^+ v_m \Leftrightarrow h_{m,n} = 0$.
- Hence: If $A^+ = p_s(A)$, then $H_{d,d-1}$ is at most $(s+2)$-band Hessenberg.
- The condition $A^+ = p_s(A)$ is *essential* in this context.

Optimal short recurrences (6)

Definition

If $A^+ = p_s(A)$, where p_s is a polynomial of the smallest possible degree s, then A is called B-normal(s).

Theorem (Faber and Manteuffel, 1984)
For A, B as above, and a nonnegative integer s with $s + 2 < d_{\min}(A)$:
A admits for the given B an optimal $(s + 2)$-term recurrence if and only if A is B-normal(s).

- This is a (generalized) characterization of normality.
- 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).
- In (Faber, L., Tichý, 2007) we give *two new proofs*,
  both using only "elementary" tools.

Optimal short recurrences (7)

In (Faber, L., Tichý, 2007) we explain why necessity is so difficult to prove:

- Assumption: For any v_1,

$$(1) \qquad A\,[v_1, \ldots, v_{d-1}] \;=\; [v_1, \ldots, v_{d-1}, v_d]\, H_{d,d-1},$$

where $H_{d,d-1}$ is $(s+2)$-band Hessenberg. To prove: A is B-normal(s).

Optimal short recurrences (7)

In (Faber, L., Tichý, 2007) we explain why necessity is so difficult to prove:

- Assumption: For any v_1,

$$(1) \qquad A\,[v_1,\ldots,v_{d-1}] \;=\; [v_1,\ldots,v_{d-1},v_d]\,H_{d,d-1},$$

  where $H_{d,d-1}$ is $(s+2)$-band Hessenberg. To prove: A is B-normal(s).

- By construction, $Av_d \in \mathcal{K}_d(A,v_1)$, i.e. $\quad Av_d \;=\; \sum_{j=1}^{d} h_{j,d} v_j$.

  Adding this in (1) gives

$$(2) \qquad A\,[v_1,\ldots,v_{d-1},v_d] \;=\; [v_1,\ldots,v_{d-1},v_d]\,H_{d,d}.$$

  But we don't know whether $H_{d,d}$ is $(s+2)$-band Hessenberg.

Optimal short recurrences (7)

In (Faber, L., Tichý, 2007) we explain why necessity is so difficult to prove:

- Assumption: For any v_1,

$$(1) \qquad A\,[v_1, \ldots, v_{d-1}] \;=\; [v_1, \ldots, v_{d-1}, v_d]\, H_{d,d-1},$$

  where $H_{d,d-1}$ is $(s+2)$-band Hessenberg. To prove: A is B-normal(s).

- By construction, $Av_d \in \mathcal{K}_d(A, v_1)$, i.e. $\quad Av_d \;=\; \sum_{j=1}^{d} h_{j,d} v_j$.

  Adding this in (1) gives

$$(2) \qquad A\,[v_1, \ldots, v_{d-1}, v_d] \;=\; [v_1, \ldots, v_{d-1}, v_d]\, H_{d,d}.$$

  But we don't know whether $H_{d,d}$ is $(s+2)$-band Hessenberg.

- $H_{d,d}$ is the matrix representation of $A : \mathcal{K}_d(A, v_1) \to \mathcal{K}_d(A, v_1)$ with respect to the orthonormal basis $v_1, \ldots, v_d$.
- Problem: *Prove something about the linear operator A, without complete knowledge of the structure of its matrix representation.*

Characterization and examples (1)

- In practice A is given and we ask:
  Does there exist an HPD B such that A is B-normal(s) with small s?
- Standard example: $A = A^*$ and $B = I$, then $A^+ = p_1(A)$ for $p_1(z) = z$.

Characterization and examples (1)

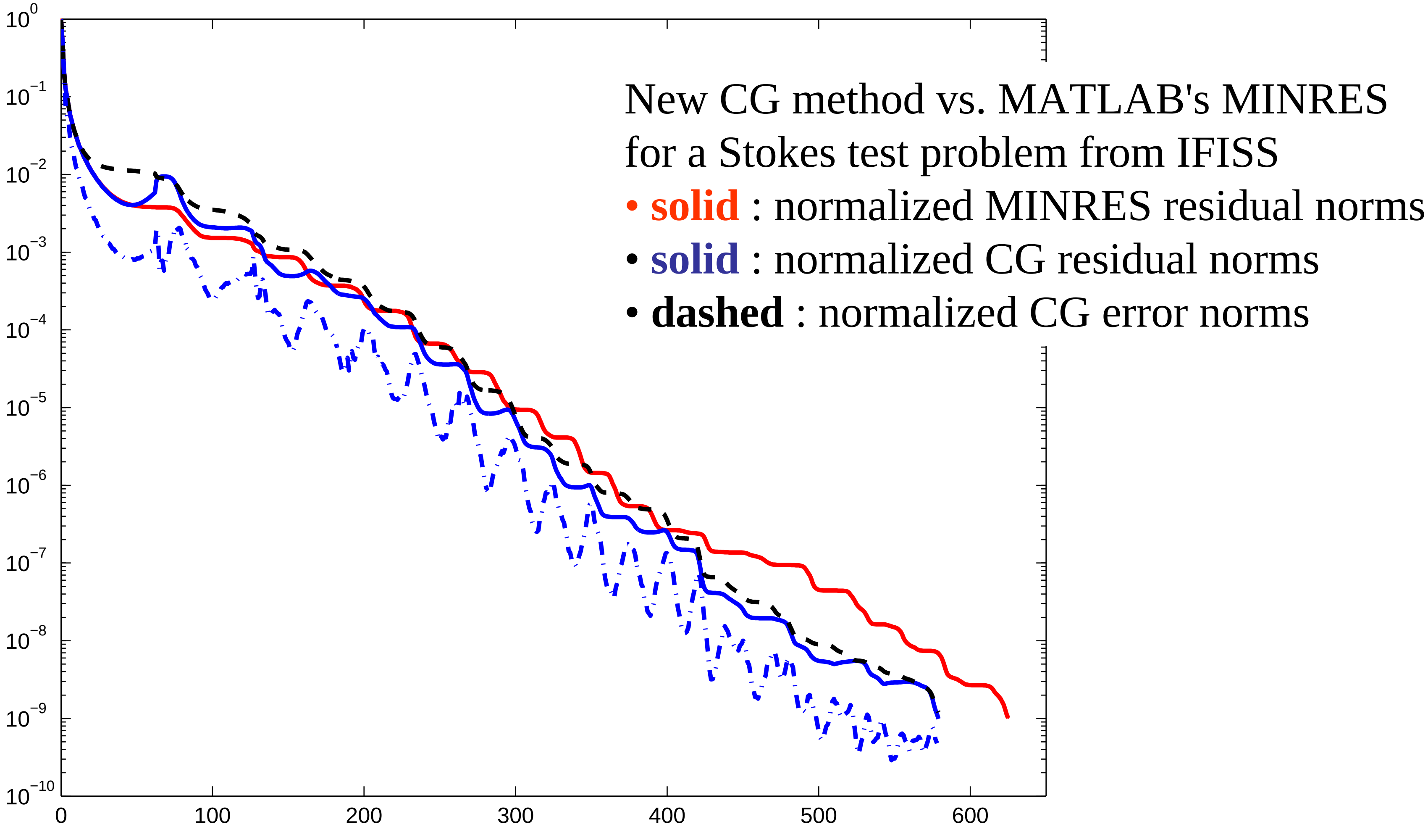
- In practice A is given and we ask:
  Does there exist an HPD B such that A is B-normal(s) with small s?
- Standard example: $A = A^*$ and $B = I$, then $A^+ = p_1(A)$ for $p_1(z) = z$.
- More interesting example: Saddle point matrices.

$$A = \begin{bmatrix} A_1 & A_2^T \\ -A_2 & A_3 \end{bmatrix}, \quad B = B(\gamma) = \begin{bmatrix} A_1 - \gamma I_m & A_2^T \\ A_2 & \gamma I_k - A_3 \end{bmatrix},$$

  where $A_1 = A_1^T > 0$, $A_3 = A_3^T \geq 0$, and $A_2 \in \mathbb{R}^{k \times m}$ has full rank k.
- A is a standard saddle point matrix with negated second block row.
- Transformation: symmetric indefinite into nonsymmetric positive real.
- **Theorem** (L. and Parlett, 2007)
  $B = B^T > 0$ holds if $\|A_2\|^2 < (\lambda_{\min}(A_1) - \gamma)\,(\gamma - \lambda_{\max}(A_3))$,
  and in this case A is B-normal(1).
  (Generalization of (Fischer et al., 1998); (Benzi and Simoncini, 2006).)

Characterization and examples (2)

- In (L. and Parlett, 2007) a corresponding 3-term CG method is constructed and analyzed. This method appears to be competitive with MINRES:



Characterization and examples (3)

General characterization:

Theorem. (L. and Strakoš, 2007)
A is B-normal(s) if and only if

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

Characterization and examples (3)

General characterization:

Theorem. (L. and Strakoš, 2007)
A is B-normal(s) if and only if

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

- Every $A = W\Lambda W^{-1}$ is B-normal(s) for *some* HPD B and *some* s, but no optimal short recurrence in the nondiagonalizable case.
- s is the smallest degree of a polynomial p_s for which $p_s(\Lambda) = \Lambda^*$.

Characterization and examples (4)

Theorem. (Faber and Manteuffel, 1984; Khavinson and Świątek, 2003)

1. $s = 1$ if and only if the eigenvalues of A lie on a line in $\mathbb{C}$ (are collinear).
2. If the eigenvalues of A are *not collinear*,
   the shortest optimal recurrence A may admit for any HPD B
   has length at least $d_{\min}(A)/3 + 4$.

$\Longrightarrow$ Except for a few unimportant cases, the length of the optimal recurrence is either 3 or $d_{\min}(A) - 1$.

$\Longrightarrow$ Overabundant supply of Krylov subspace methods for general A.

Characterization and examples (4)

Theorem. (Faber and Manteuffel, 1984; Khavinson and Świątek, 2003)

1. $s = 1$ if and only if the eigenvalues of A lie on a line in $\mathbb{C}$ (are collinear).
2. If the eigenvalues of A are *not collinear*,
   the shortest optimal recurrence A may admit for any HPD B
   has length at least $d_{\min}(A)/3 + 4$.

$\Longrightarrow$ Except for a few unimportant cases, the length of the optimal recurrence is either 3 or $d_{\min}(A) - 1$.

$\Longrightarrow$ Overabundant supply of Krylov subspace methods for general A.

Options for $Ax = b$ in case the eigenvalues of A are not collinear:

- Orthogonality but full recurrence (GMRES).
- Short recurrence but no orthogonality (BiCG, QMR, etc.).
- "Preconditioners" P so that PA is B-normal(1) for some B, e.g. (Concus and Golub, 1978) and (Widlund, 1978).
- *Special types of recurrences (Isometric Arnoldi).*

Characterization and examples (5)

- (Gragg, 1982) discovered the *isometric Arnoldi process*:
  Orthogonal Krylov subspace bases for unitary A can be generated by a (non-optimal) 3-term recurrence of the form

$$v_{n+1} \; = \; Av_n - \beta_{n,1} A v_{n-1} - \beta_{n,2} v_n$$

  (stable implementation is in form of two coupled 2-term recurrences).

- This algorithm is used for solving unitary eigenvalue problems and linear systems with shifted unitary matrices.

Characterization and examples (5)

- (Gragg, 1982) discovered the *isometric Arnoldi process*:
  Orthogonal Krylov subspace bases for unitary A can be generated by a (non-optimal) 3-term recurrence of the form

$$v_{n+1} = Av_n - \beta_{n,1} A v_{n-1} - \beta_{n,2} v_n$$

  (stable implementation is in form of two coupled 2-term recurrences).

- This algorithm is used for solving unitary eigenvalue problems and linear systems with shifted unitary matrices.

- Generalized in (Barth and Manteuffel, 2000):
  If $A^+ = r(A)$ for $r = p/q$, where p and q have degrees s and t,
  a B-orthogonal Krylov subspace can be generated by a (non-optimal) recurrence of length at most $t + s + 2$.

- In case A is unitary: $A^* = A^{-1}$, hence $p(z) = 1$, $q(z) = z$.

Characterization and examples (5)

- Are thre any other matrices A whose adjoint A^+ (for some B) is a low degree rational function in A?
- **Theorem** (L., 2007)
  There is an HPD B such that $A^+ = r(A)$ with *small* $\deg r \equiv \max\{\deg p,\ \deg q\}$ if and only if either $d_{\min}(A)$ is small,
  or A is diagonalizable with collinear or concyclic eigenvalues.

More precisely:
For diagonalizable A with $n \geq 4$ distinct eigenvalues that are neither collinear nor concyclic,

$$d_r(A) > \frac{n}{5}, \qquad d_p(A) > \frac{n}{3}, \qquad \text{and} \qquad 1 \leq \frac{d_p(A)}{d_r(A)} < 5.$$

Concluding remarks

- Completely reworked the theory of short recurrences for generating orthogonal Krylov subspace bases; new, mathematically rigorous definitions of all important concepts have been given.
- In particular, here we discussed:

  new proofs of the fundamental theorem of Faber and Manteuffel,
  a new 3-term CG method for saddle point matrices,
  the existence of alternative (isometric Arnoldi style) recurrences.

- Visit `http://www.math.tu-berlin.de/~liesen` for the related papers:

J.L. and P. Saylor, *Orthogonal Hessenberg reduction and orthogonal Krylov subspace bases*, SINUM 42 (2005).

J.L. and Z. Strakoš, *On optimal short recurrences for generating orthogonal Krylov subspace bases*, to appear in SIREV.

J.L., *When is the adjoint of a matrix a low degree rational function in the matrix?*, to appear in SIMAX.

V. Faber, J.L. and P. Tichý, *The Faber-Manteuffel Theorem for linear operators*, submitted.

J.L. and B. N. Parlett, *On nonsymmetric saddle point matrices that allow conjugate gradient iterations*, submitted.