Polynomial Based Iteration Methods for Large Linear Systems

Bernd Fischer
Institute of Mathematics
(Medical) University of Lübeck
http://www.math.uni-luebeck.de

Kompaktkurs Stuttgart – 27/02/2003

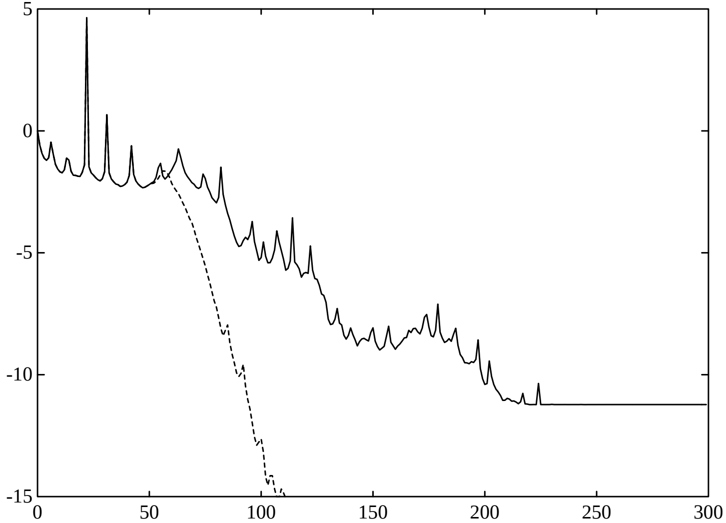
Program (Symmetric Systems)

- A positive definite
  - Conjugate Gradient (CG)
  - Conjugate Residual (CR)
- A indefinite
  - MINimal RESidual (MINRES)
  - SYMMetric LQ (SYMMLQ)

Program (Nonsymmetric Systems)

- Fast Convergence
  - Generalized Minimal RESidual (GMRES)
- Efficient Implementation
  - BI Conjugate Gradient (Bi-CG)
- Both
  - Bi-CG STABilized (Bi-CGSTAB)
  - Quasi Minimal Residual (QMR)

Stable Implementation (Symmetric)

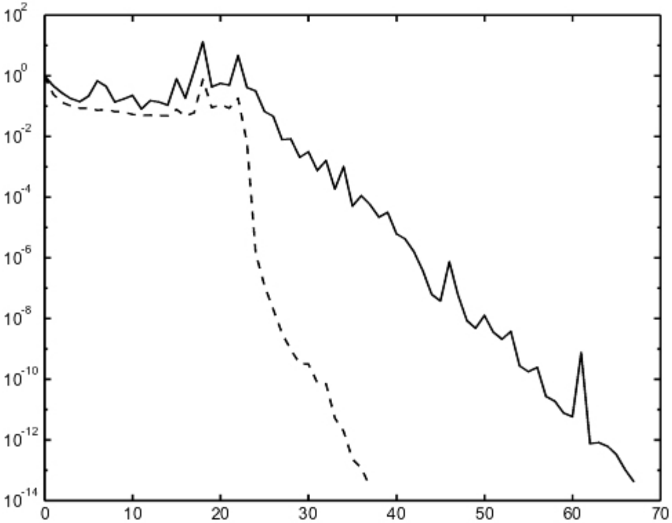


Implementations

- CG classical (–)
- CG stable (- -)

$$\log_{10}\left(\frac{\|f - Ax_n\|_2}{\|r_0\|_2}\right) \quad \text{against} \quad n$$

Stable Implementation (Nonsymmetric)



$$\log_{10}\left(\frac{\|f - Ax_n\|_2}{\|r_0\|_2}\right) \quad \text{against} \quad n$$

Schemes

- BiCG (–)
- BiCGSTAB (- -)

Some Literature (1/2)

- **Owe Axelsson**: Iterative solution methods, Cambridge University Press, 1994
- **Richard Barett** . . . **Henk van der Vorst**: Templates for the solution of linear systems: building blocks for iterative methods, SIAM, 1994
- **Claude Brezinski**: Projection methods for systems of equations, Elsevier, 1997
- **Roland Freund, Gene Golub, Noel Nachtigal**: Iterative solution of linear systems, Acta Numerica 1, pp. 57-100, 1991
- **Anne Greenbaum**: Iterative methods for solving linear systems, SIAM, 1997
- **Wolfgang Hackbusch**: Iterative solution of large sparse systems of equations, Springer, 1994

Some Literature (2/2)

- **C.T. Kelley**: Iterative methods for linear and nonlinear equations, SIAM, 1995
- **A. Meister**: Numerik linearer Gleichungssysteme, Vieweg, 1999
- **G. Meurant**: Computer solution of large linear systems, North Holland, 1999
- **Yousef Saad**: Iterative methods for sparse linear systems, PWS Publishing, Boston, 1996
- **H.A. van der Vorst**: Iterative methods for large linear systems, http://www.math.ruu.nl/people/vorst/
- **Rüdiger Weiß**: Parameter-free iterative solvers, Akademie, 1996
- **Bernd Fischer**: Polynomial based iteration methods for symmetric linear systems, Wiley-Teubner, 1996

Introduction

The System

$$Ax = f, \quad A \in \mathbb{R}^{N \times N},$$

where A is in general (non)**symmetric** and **large**.

Large systems have structure:

- $A = [a_{jk}]$ is **sparse**, i.e., $a_{jk} = 0$ for most j, k
- A is dense and somehow **structured**

Consequently: matrix-vector products $A \cdot v$ (and $A^T \cdot w$) can be computed cheaply.

Typically: $\mathcal{O}(N)$ or $\mathcal{O}(N \log N)$ work, instead of $\mathcal{O}(N^2)$

Notation

System	$Ax = f$
Solution	x_*
Initial guess	x_0
Error	$\varepsilon_n = x_* - x_n$
Residual	$r_n = f - Ax_n (= A\varepsilon_n)$

Polynomial Based Iteration Method (Definition 1/2)

Iteration polynomial

$$x_n = x_0 + q_{n-1}(A)r_0$$

Residual polynomial

$$\begin{aligned} r_n &= f - Ax_n \\ &= f - A(x_0 + q_{n-1}(A)r_0) \\ &= (I - Aq_{n-1}(A))r_0 \\ &= p_n(A)r_0, \end{aligned}$$

where

$$p_n(t) = 1 - tq_{n-1}(t).$$

Polynomial Based Iteration Method (Definition 2/2)

Polynomial based method

$$
\begin{aligned}
x_n &= x_0 + q_{n-1}(A) r_0, \quad q_{n-1} \in \Pi_{n-1} \\
r_n &= p_n(A) r_0, \quad p_n \in \Pi_n, \ p_n(0) = 1
\end{aligned}
$$

Krylov Subspace

$$K_n(A, r_0) := \operatorname{span}\{r_0, Ar_0, A^2 r_0, \dots, A^{n-1} r_0\}$$

Consequently

$$x_n \in x_0 + K_n(A, r_0), \quad r_n \in K_{n+1}(A, r_0)$$

Polynomial Based Iteration Method (Design 1/2)

Scheme

$$r_n = p_n(A) r_0, \quad p_n \in \Pi_n,\ p_n(0) = 1$$

Goal: drive r_n to 0 as fast as possible with little storage requirement

Design principles: x_n defined by a

- **Minimal residual property** (fast convergence)
  CR, MINRES, GMRES

$$\|r_n\| = \|p_n(A) r_0\| = \min\{\|p(A) r_0\| :\ p \in \Pi_n,\ p(0) = 1\}$$

Polynomial Based Iteration Method (Design 2/2)

- **Galerkin condition** (efficient implementation)
  CG, SYMMLQ, BiCG

$$\widehat{\langle r_n, r_j \rangle} = 0, \quad j < n \quad \Leftrightarrow \quad \langle p_n, p_j \rangle = 0, \quad j < n$$

- **Both**
  - Symmetric system
    CR, CG, MINRES, SYMMLQ
  - Nonsymmetric system
    $\approx$ BiCGSTAB, $\approx$ QMR

Orthogonal residual polynomials

Residual Polynomials (Orthogonality)

Orthogonal residual polynomials

$$\langle p_j, p_k \rangle = \begin{cases} = 0 & \text{for } j \neq k \\ > 0 & \text{for } j = k \end{cases}, \qquad \deg p_j = j$$

Fourier-expansion

$$\begin{aligned} tp_{j-1}(t) &= \sum_{k=0}^{j} \frac{\langle tp_{j-1}, p_k \rangle}{\langle p_k, p_k \rangle} p_k(t) \\ &= \sum_{k=0}^{j} \frac{\langle p_{j-1}, tp_k \rangle}{\langle p_k, p_k \rangle} p_k(t) \\ &= \sum_{k=j-2}^{j} \frac{\langle tp_{j-1}, p_k \rangle}{\langle p_k, p_k \rangle} p_k(t) \end{aligned}$$

Three-term recurrence relation

$$p_{-1}(t) := 0, \quad p_0(t) = p_0,$$
$$\gamma_j p_j(t) = (t - \alpha_j) p_{j-1}(t) - \beta_j p_{j-2}(t), \quad j \geq 1$$

$$\gamma_j = \frac{\langle t p_{j-1}, p_j \rangle}{\langle p_j, p_j \rangle}, \quad \alpha_j = \frac{\langle t p_{j-1}, p_{j-1} \rangle}{\langle p_{j-1}, p_{j-1} \rangle}, \quad \beta_j = \frac{\langle t p_{j-1}, p_{j-2} \rangle}{\langle p_{j-2}, p_{j-2} \rangle}$$

Interpolatory constraint

$$p_j(0) = p_{j-1}(0) = p_{j-2}(0) = 1 \quad \Rightarrow \quad \gamma_j = -(\alpha_j + \beta_j)$$

Stieltjes procedure

Set

$$p_{-1}(t) := 0, \quad p_0(t) = 1,$$

For $j = 1, 2, \ldots$ compute

$$\alpha_j = \frac{\langle t p_{j-1}, p_{j-1} \rangle}{\langle p_{j-1}, p_{j-1} \rangle}, \quad \beta_j = \frac{\langle t p_{j-1}, p_{j-2} \rangle}{\langle p_{j-2}, p_{j-2} \rangle},$$

and

$$-(\alpha_j + \beta_j) p_j(t) = (t - \alpha_j) p_{j-1}(t) - \beta_j p_{j-2}(t).$$

End

Polynomial Based Iteration (Implementation 1/2)

$$r_n = p_n(A) r_0,$$

where

$$\gamma_n p_n(t) = (t + (\gamma_n + \beta_n)) p_{n-1}(t) - \beta_n p_{n-2}(t).$$

Consequently

$$\begin{aligned}
\gamma_n r_n &= (A + (\gamma_n + \beta_n) E_n) r_{n-1} - \beta_n r_{n-2} \\
&= \gamma_n r_{n-1} + A r_{n-1} + \beta_n ((f - A x_{n-1}) - (f - A x_{n-2})) \\
&= \gamma_n r_{n-1} + A(r_{n-1} + \beta_n (x_{n-2} - x_{n-1})).
\end{aligned}$$

This leads to

$$r_n - r_{n-1} = \frac{1}{\gamma_n} A(r_{n-1} + \beta_n (x_{n-2} - x_{n-1})) =: -\frac{1}{\gamma_n} A w_{n-1},$$

where

$$w_{n-1} = \beta_n (x_{n-1} - x_{n-2}) - r_{n-1} = \frac{\beta_n}{\gamma_{n-1}} w_{n-2} - r_{n-1}.$$

Polynomial Based Iteration (Implementation 2/2)

Note

$$r_n - r_{n-1} = (f - Ax_n) - (f - Ax_{n-1}) = -A(x_n - x_{n-1}).$$

Update formulae

$$\begin{aligned} x_n &= x_{n-1} + \frac{1}{\gamma_n} w_{n-1}, \\ r_n &= r_{n-1} - \frac{1}{\gamma_n} A w_{n-1}. \end{aligned}$$

Prototype Algorithm

Given the orthogonal residual polynomial

$$\gamma_n p_n(t) = (t + (\gamma_n + \beta_n))p_{n-1}(t) - \beta_n p_{n-2}(t),$$

the next algorithm implements the polynomial based method $r_n = p_n(A)r_0$.

Set

$$w_{-1} := 0, \quad r_0 = f - Ax_0.$$

For $n = 1, 2, \ldots$ compute

$$\begin{aligned} w_{n-1} &= \frac{\beta_n}{\gamma_{n-1}} w_{n-2} - r_{n-1}, \\ x_n &= x_{n-1} + \frac{1}{\gamma_n} w_{n-1}, \\ r_n &= r_{n-1} - \frac{1}{\gamma_n} A w_{n-1}. \end{aligned}$$

Conjugate Residual (CR)

Minimal Residual Property (general 1/2)

Scheme

$$r_n = p_n(A) r_0, \quad p_n \in \Pi_n,\ p_n(0) = 1$$

Minimal residual property

$$\|r_n\|_2 = \|p_n(A) r_0\|_2 = \min\{\|p(A) r_0\|_2 :\ p \in \Pi_n,\ p(0) = 1\}$$

(CR, MINRES, GMRES)

Minimal Residual Property (general 2/2)

Note:

$$\|r_n\|_2^2 = r_n^T r_n = r_0^T p_n(A) p_n(A) r_0 =: \langle p_n, p_n \rangle_{MR}$$

Inner product (w.r.t. A and r_0)

$$\langle p, q \rangle_{MR} := r_0^T p(A) q(A) r_0.$$

Minimal residual property (equivalent formulation)

Find a polynomial $p_n \in \Pi_n,\ p_n(0) = 1$, such that

$$\langle p_n, p_n \rangle_{MR} \leq \langle p, p \rangle_{MR} \quad \text{for all} \quad p \in \Pi_n,\ p(0) = 1.$$

Minimal Residual Property (explicit solution)

Orthonormal polynomials

$$\langle \psi_j, \psi_k \rangle_{MR} = \delta_{j,k}$$

Kernel polynomial

$$p_n^{MR}(t) = \frac{\sum_{j=0}^{n} \psi_j(0)\psi_j(t)}{\sum_{j=0}^{n} \psi_j(0)^2}$$

Extremal property

$$\langle p_n^{MR}, p_n^{MR} \rangle_{MR} \leq \langle p, p \rangle_{MR} \quad \text{for all} \quad p \in \Pi_n,\ p(0) = 1.$$

Kernel Polynomials (properties)

Reproducing property (defining property)

$$\langle p_n^{MR}, q_n \rangle_{MR} = q_n(0), \quad \text{for all} \quad q_n \in \Pi_n.$$

Special case: $q_n(t) = t q_{n-1}(t)$

$$\langle p_n^{MR}, t q_{n-1} \rangle_{MR} = 0, \quad \text{for all} \quad q_{n-1} \in \Pi_{n-1}.$$

Orthogonality

Kernel polynomials p_n^{MR} are orthogonal with respect to the modified inner product

$$\langle p, tq \rangle_{MR} = r_0^T p(A) A q(A) r_0$$

$\hookrightarrow$ three-term recurrence relation

Conjugate Residual Method

Polynomial based method w.r.t Kernel polynomials

$$r_n^{MR} = p_n^{MR}(A) r_0$$

Extremal property

$$\|r_n^{MR}\|_2 \le \|p_n(A) r_0\|_2, \quad \text{for all} \quad p_n \in \Pi_n,\ p_n(0) = 1.$$

Conjugate residual

$$\langle p_k^{MR}, t p_j^{MR} \rangle_{MR} = (r_k^{MR})^T A r_j^{MR} = 0 \quad \text{for} \quad k \neq j$$

Three-term recurrence coefficients

$$\alpha_j^{MR} = \frac{\langle t p_{j-1}^{MR}, t p_{j-1}^{MR} \rangle_{MR}}{\langle p_{j-1}^{MR}, t p_{j-1}^{MR} \rangle_{MR}} = \frac{(A r_{j-1}^{MR})^T A r_{j-1}^{MR}}{(r_{j-1}^{MR})^T A r_{j-1}^{MR}}$$

Prototype Algorithm

Given the orthogonal residual polynomial

$$\gamma_n p_n(t) = (t + (\gamma_n + \beta_n))p_{n-1}(t) - \beta_n p_{n-2}(t),$$

the next algorithm implements the polynomial based method $r_n = p_n(A)r_0$.

Set

$$w_{-1} := 0, \quad r_0 = f - Ax_0.$$

For $n = 1, 2, \ldots$ compute

$$\begin{aligned} w_{n-1} &= \frac{\beta_n}{\gamma_{n-1}} w_{n-2} - r_{n-1}, \\ x_n &= x_{n-1} + \frac{1}{\gamma_n} w_{n-1}, \\ r_n &= r_{n-1} - \frac{1}{\gamma_n} A w_{n-1}. \end{aligned}$$

Conjugate Residual Method (implementation, Stiefel 57)

Set

$$w_{-1}^{MR} = 0, \quad x_0^{MR} = x_0, \quad r_0^{MR} = r_0.$$

For $n = 1, 2, \ldots,$ until convergence

$$\nu_n^{MR} = \frac{\beta_n^{MR}}{\gamma_{n-1}^{MR}} = \frac{(r_{n-1}^{MR})^T A r_{n-1}^{MR}}{(r_{n-2}^{MR})^T A r_{n-2}^{MR}}$$

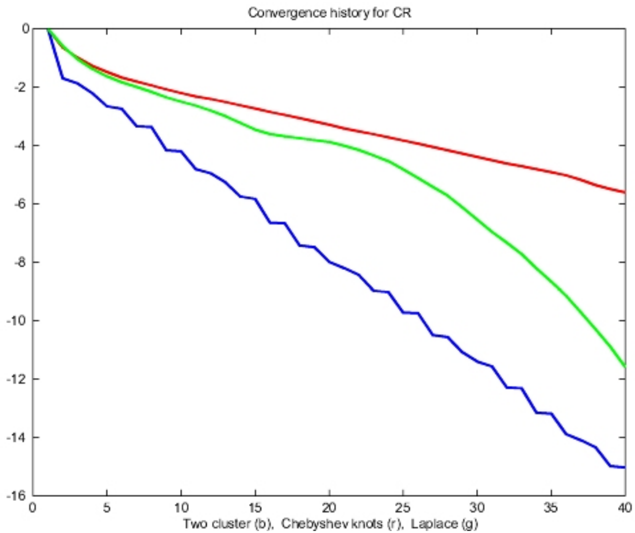
$$w_{n-1}^{MR} = \nu_n^{MR} w_{n-2}^{MR} - r_{n-1}^{MR}, \quad w_0^{MR} = r_0^{MR}$$

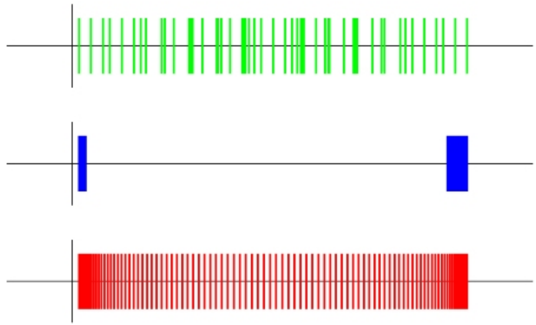
$$\eta_n^{MR} = \frac{1}{\gamma_n^{MR}} = \frac{(r_{n-1}^{MR})^T A r_{n-1}^{MR}}{(A w_{n-1}^{MR})^T A w_{n-1}^{MR}}$$

$$x_n^{MR} = x_{n-1}^{MR} + \eta_n^{MR} w_{n-1}^{MR}$$

$$r_n^{MR} = r_{n-1}^{MR} - \eta_n^{MR} A w_{n-1}^{MR}$$

Conjugate Residual Method (convergence)





Residual Polynomials (convergence 1/2)

Eigenvalues and orthonormal eigenvectors $(A = A^T)$

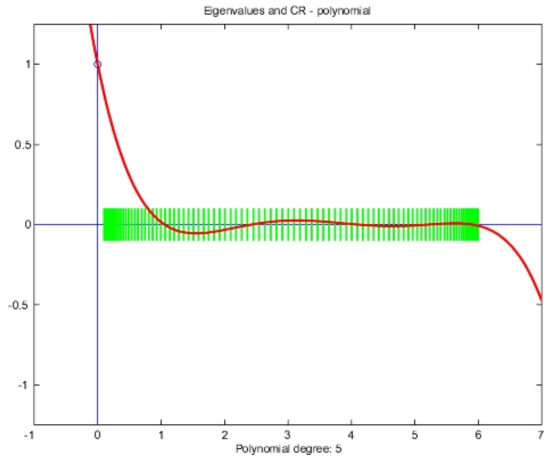
$$Av_j = \lambda_j v_j, \quad v_j^T v_k = \delta_{j,k}$$

Initial residual

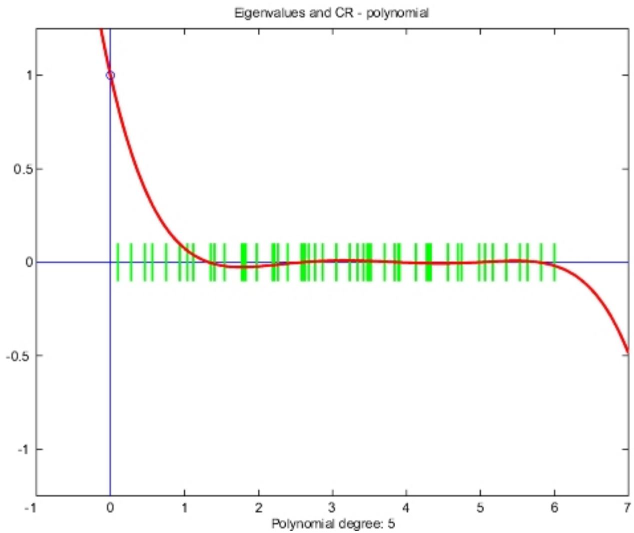
$$r_0 = \sum_{j=1}^{L} \sigma_j v_j, \quad \lambda_1 < \lambda_2 < \cdots \lambda_L, \quad \sigma_j \neq 0$$

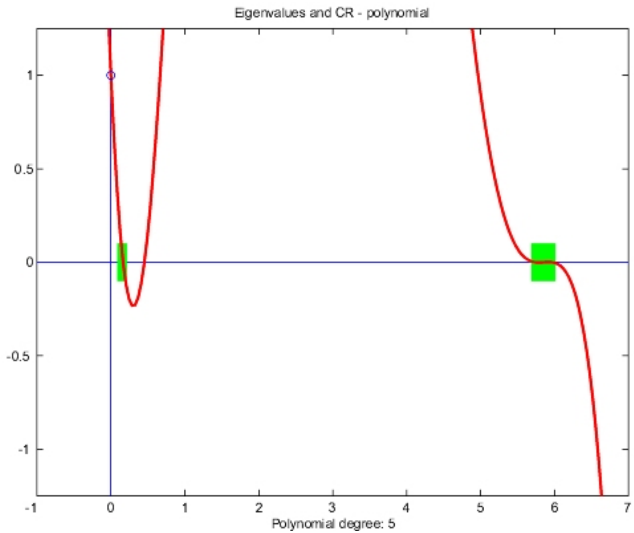
Norm

$$\|r_n^{MR}\|_2^2 = \|p_n^{MR}(A) r_0\|_2^2 = \sum_{j=1}^{L} \sigma_j^2 (p_n^{MR}(\lambda_j))^2$$



Residual Polynomials (convergence 2/2)





Conjugate Gradient (CG)

Conjugate Gradient Method (energy norm)

Recall:

$$r_n = p_n(A) r_0, \quad p_n \in \Pi_n, \quad p_n(0) = 1$$

$$r_n = f - Ax_n = A(x_* - x_n) = A\varepsilon_n \quad \Rightarrow \quad \varepsilon_n = p_n(A)\varepsilon_0$$

Note:

$$\begin{aligned} g(x_n) &= \frac{1}{2} x_n^T A x_n - x_n^T f; \; g'(x_n) = Ax_n - f, \\ &= \frac{1}{2}(x_* - x_n)^T A (x_* - x_n) - \frac{1}{2} x_*^T A x_* \\ &= \frac{1}{2}\|x_* - x_n\|_A^2 - \frac{1}{2} x_*^T A x_* \\ &= \frac{1}{2}\|\varepsilon_n\|_A^2 - \frac{1}{2} x_*^T A x_* \end{aligned}$$

Conjugate Gradient Method (definition)

Minimize A-norm of the error (energy-norm)

$$\|\varepsilon_n\|_A = \|p_n(A)\varepsilon_0\|_A = \min\left\{\|p(A)\varepsilon_0\|_A : \ p \in \Pi_n, \ p(0) = 1\right\}$$

Inner product (w.r.t. A and r_0)

$$\|\varepsilon_n\|_A^2 = \varepsilon_0^T p_n(A) A p_n(A)\varepsilon_0 \ =: \langle p_n, p_n\rangle_{GAL}$$

Find a polynomial $p_n \in \Pi_n, \ p_n(0) = 1$, such that

$$\langle p_n, p_n\rangle_{GAL} \le \langle p, p\rangle_{GAL} \quad \text{for all} \quad p \in \Pi_n, \ p(0) = 1.$$

Conjugate Gradient Method (Kernel polynomials)

Solution: Kernel polynomials p_n^{GAL}, orthogonal w.r.t. $\langle \cdot, t\cdot \rangle_{GAL}$

$$\begin{aligned}
\langle p_j^{GAL}, t p_k^{GAL} \rangle_{GAL} &= (A\varepsilon_0)^T p_j^{GAL}(A) p_k^{GAL}(A) A\varepsilon_0 \\
&= r_0^T p_j^{GAL}(A) p_k^{GAL}(A) r_0 \\
&= (r_j^{GAL})^T r_k^{GAL} \\
&= \langle p_j^{GAL}, p_k^{GAL} \rangle_{MR} = 0, \quad j \neq k
\end{aligned}$$

The CG residual polynomials

$$p_n^{GAL}(t) = \frac{\psi_n(t)}{\psi_n(0)}$$

are rescaled orthonormal polynomials $\langle \psi_j, \psi_k \rangle_{MR} = \delta_{j,k}$

Prototype Algorithm

Given the orthogonal residual polynomial

$$\gamma_n p_n(t) = (t + (\gamma_n + \beta_n))p_{n-1}(t) - \beta_n p_{n-2}(t),$$

the next algorithm implements the polynomial based method $r_n = p_n(A)r_0$.

Set

$$w_{-1} := 0, \quad r_0 = f - Ax_0.$$

For $n = 1, 2, \ldots$ compute

$$\begin{aligned} w_{n-1} &= \frac{\beta_n}{\gamma_{n-1}} w_{n-2} - r_{n-1}, \\ x_n &= x_{n-1} + \frac{1}{\gamma_n} w_{n-1}, \\ r_n &= r_{n-1} - \frac{1}{\gamma_n} A w_{n-1}. \end{aligned}$$

Conjugate Gradient Algorithm (Hestenes/Stiefel 52)

Scheme

$$r_n^{GAL} = p_n^{GAL}(A) r_0$$

with (Galerkin condition)

$$0 = \langle p_j^{GAL}, p_k^{GAL} \rangle_{MR} = (p_j^{GAL}(A) r_0)^T p_k^{GAL}(A) r_0 = (r_j^{GAL})^T r_k^{GAL}$$

Implementation

$$\begin{aligned} w_{n-1} &= \nu_n w_{n-2} - r_{n-1}^{GAL}, \quad & \nu_n = \frac{(r_{n-1}^{GAL})^T r_{n-1}^{GAL}}{(r_{n-2}^{GAL})^T r_{n-2}^{GAL}} \\ x_n^{GAL} &= x_{n-1}^{GAL} + \eta_n w_{n-1}, \quad & \eta_n = -\frac{(r_{n-1}^{GAL})^T r_{n-1}^{GAL}}{w_{n-1}^T A w_{n-1}} \\ r_n^{GAL} &= r_{n-1}^{GAL} - \eta_n A w_{n-1} & \end{aligned}$$

CG versus CR

Inner product

$$\langle p, q \rangle_{MR} := (p(A)r_0)^T q(A)r_0$$

CG

- residuals are orthogonal

$$\begin{aligned} & (r_j^{GAL})^T r_k^{GAL} \\ = \; & (p_j^{GAL}(A)r_0)^T p_k^{GAL}(A)r_0 \\ = \; & \langle p_j^{GAL}, p_k^{GAL} \rangle_{MR} = 0. \end{aligned}$$

- minimizes the "A-norm" of the error

$$\|\varepsilon_n^{GAL}\|_A = \|p_n^{GAL}(A)\varepsilon_0\|_A.$$

CR

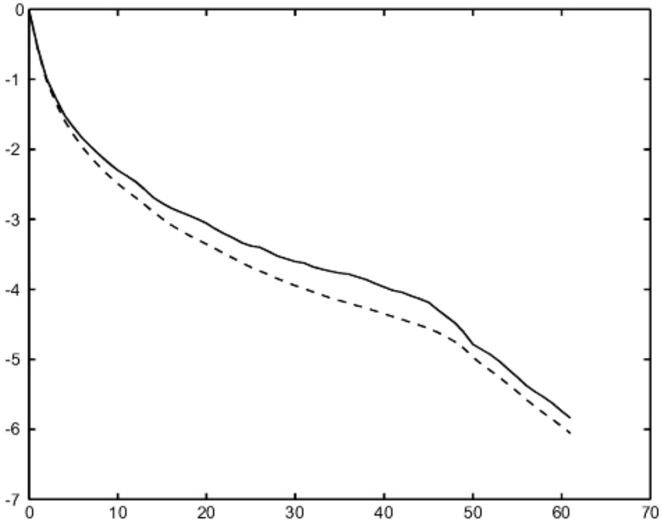
- residuals are A-conjugate

$$\begin{aligned} & (r_j^{MR})^T A r_k^{MR} \\ = \; & (p_j^{MR}(A)r_0)^T A p_k^{MR}(A)r_0 \\ = \; & \langle p_j^{MR}, t p_k^{MR} \rangle_{MR} = 0. \end{aligned}$$

- minimizes the 2-norm of the residual

$$\|r_n^{MR}\|_2 = \|p_n^{MR}(A)r_0\|_2.$$

CG versus CR (Residuals)



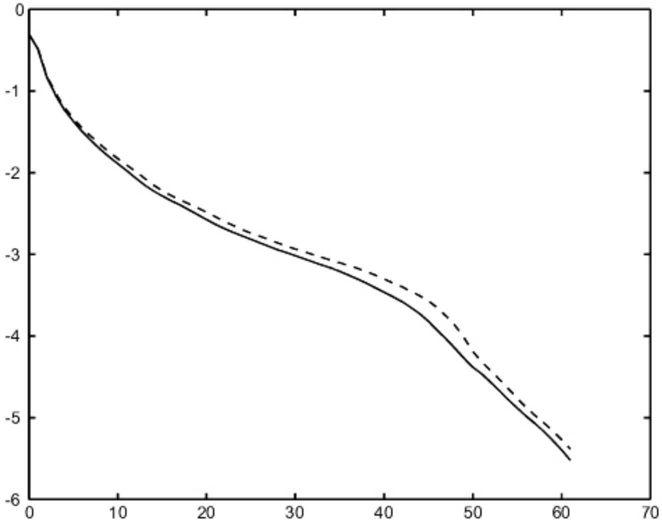
Laplace-matrix

$$A \in \mathbb{R}^{625 \times 625}$$

$$\text{cond}(A) \approx 270$$

$\log_{10}\left(\frac{\|f - Ax_n\|_2}{\|r_0\|_2}\right)$ against n, CG (–), CR (- -)

CG versus CR (Errors)



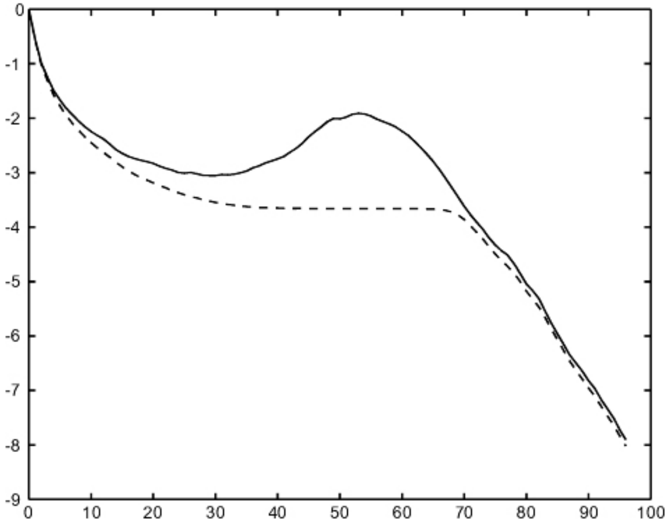
Laplace-matrix

$$A \in \mathbb{R}^{625 \times 625}$$

$$\text{cond}(A) \approx 270$$

$$\log_{10}\left(\frac{\|x_* - x_n\|_A}{\|\varepsilon_0\|_A}\right)$$ against n, CG (–), CR (- -)

CG versus CR (Residuals)



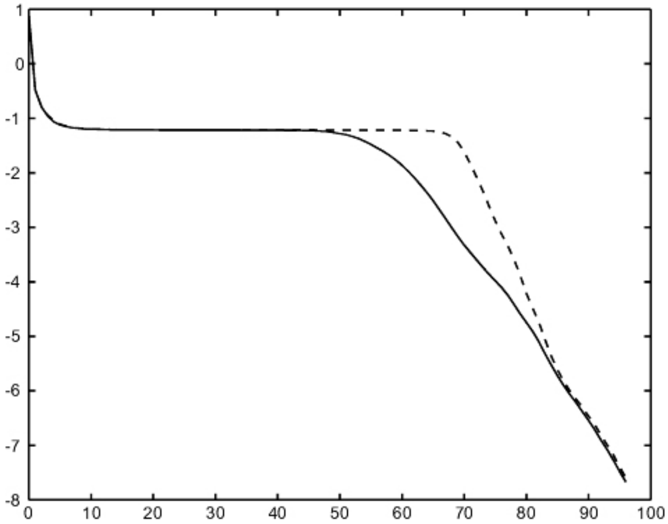
Shifted
Laplace-matrix

$$A \in \mathbb{R}^{625\times 625}$$

cond$(A) \approx 120.000$

$$\log_{10}\left(\frac{\|f - Ax_n\|_2}{\|r_0\|_2}\right) \quad \text{against} \quad n, \quad \text{CG (–), CR (- -)}$$

CG versus CR (Errors)



Shifted
Laplace-matrix

$$A \in \mathbb{R}^{625 \times 625}$$

cond$(A) \approx 120.000$

$$\log_{10}\left(\frac{\|x_* - x_n\|_A}{\|\varepsilon_0\|_A}\right)$$ against n, CG (–), CR (- -)

Indefinite systems

Breakdown of CG

Orthonormal polynomials

$$\langle \psi_j, \psi_k \rangle_{MR} = \delta_{j,k}.$$

Recall:

$$r_n^{GAL} = p_n^{GAL}(A) r_0, \quad \text{where} \quad p_n^{GAL}(t) = \frac{\psi_n(t)}{\psi_n(0)}.$$

Fact:

CG breaks down at step n iff $\psi_n(0) = 0$

(does **not** depend on the implementation)

Breakdown of CR

Recall:

$$r_n^{MR} = p_n^{MR}(A) r_0, \quad \text{where} \quad p_n^{MR}(t) = \frac{\sum_{j=0}^{n} \psi_j(0)\psi_j(t)}{\sum_{j=0}^{n} \psi_j(0)^2}.$$

$$\begin{aligned} \psi_n(0) = 0 \quad & \Rightarrow \quad p_n^{MR} = p_{n-1}^{MR}, \quad (r_n^{MR} = r_{n-1}^{MR}, \text{ plateau}) \\ & \Rightarrow \quad 0 = \langle p_n^{MR}, t p_{n-1}^{MR} \rangle_{MR} = \langle p_{n-1}^{MR}, t p_{n-1}^{MR} \rangle_{MR} \\ & \Rightarrow \quad \eta_n^{MR} = \infty \quad (\gamma_n^{MR} = 0). \end{aligned}$$

Fact:

CR breaks down at step n iff $\psi_n(0) = 0$

(**does** depend on the implementation based on the three-term recurrence relation, ($\psi_0(0) \neq 0$))

Breakdown of CG/CR (1/2)

Eigenvalues and eigenvectors of A

$$Av_j = \lambda_j v_j, \quad \lambda_1 < \lambda_2 < \cdots < \lambda_L, \quad v_i^T v_j = \delta_{i,j}$$

Initial residual

$$r_0 = \sum_{k=1}^{L} \sigma_k v_k$$

Orthonormal polynomials

$$\langle \psi_i, \psi_j \rangle_{MR} = (\psi_i(A) r_0)^T \psi_j(A) r_0 = \sum_{k=1}^{L} \sigma_k^2 \psi_i(\lambda_k) \psi_j(\lambda_k)$$

Breakdown of CG/CR (2/2)

Note

$$\psi_L(t) = c \prod_{k=1}^{L} (t - \lambda_k)$$

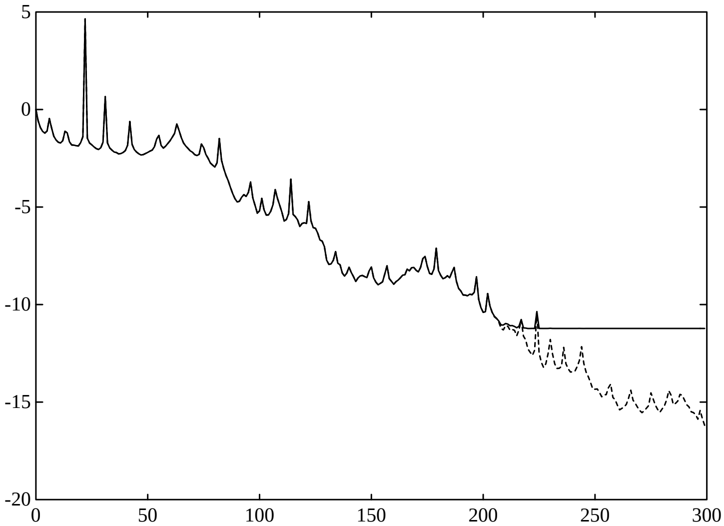
Interlacing property of **Ritz values**

$$\psi_n(\theta_j^{(n)}) = 0 \quad \Rightarrow \quad \begin{cases} \theta_j^{(n)} \in [\lambda_1, \lambda_L] \\ \lambda_k \in [\theta_j^{(n)}, \theta_{j+1}^{(n)}] \end{cases}$$

A positive definite $\Rightarrow$ $\psi_n(0) = 0$ impossible

A indefinite $\Rightarrow$ $\psi_n(0) = 0$ possible

True and Updated Residual (CG)



Shifted
"Van der Vorst matrix"

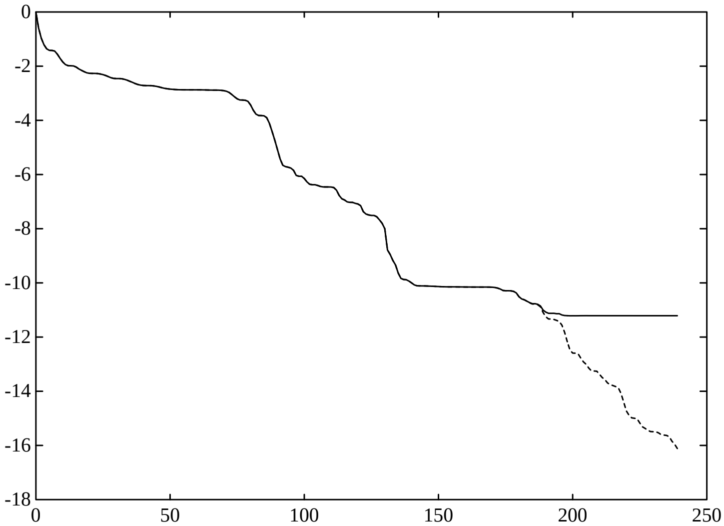
$$A = \text{diag}(-9, -7, \ldots, 189) - c$$

Shift $c \approx 0.97$ such that

$$\theta_3^{(22)} = 8.48\ldots * 10^{-7}$$

$$\log_{10}\left(\frac{\|r_n\|_2}{\|r_0\|_2}\right) \text{ (- -)}, \ \log_{10}\left(\frac{\|f - Ax_n\|_2}{\|r_0\|_2}\right) \text{ (–)}$$

True and Updated Residual (CR)



Shifted
"Van der Vorst matrix"

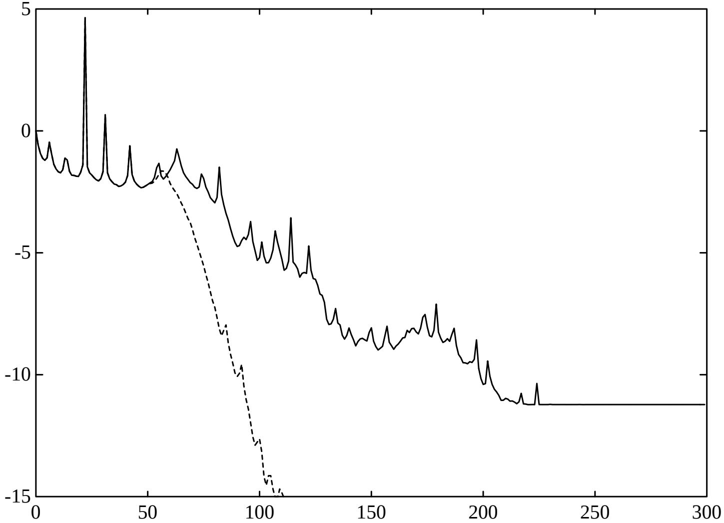
$$A = \text{diag}(-9, -7, \dots, 189) - c$$

Shift $c \approx 0.97$ such that

$$\theta_3^{(22)} = 8.48\ldots * 10^{-7}$$

$$\log_{10}\left(\frac{\|r_n\|_2}{\|r_0\|_2}\right) \text{ (- -)}, \ \log_{10}\left(\frac{\|f - Ax_n\|_2}{\|r_0\|_2}\right) \text{ (–)}$$

Stable Implementation (CG)



Shifted
"Van der Vorst matrix"

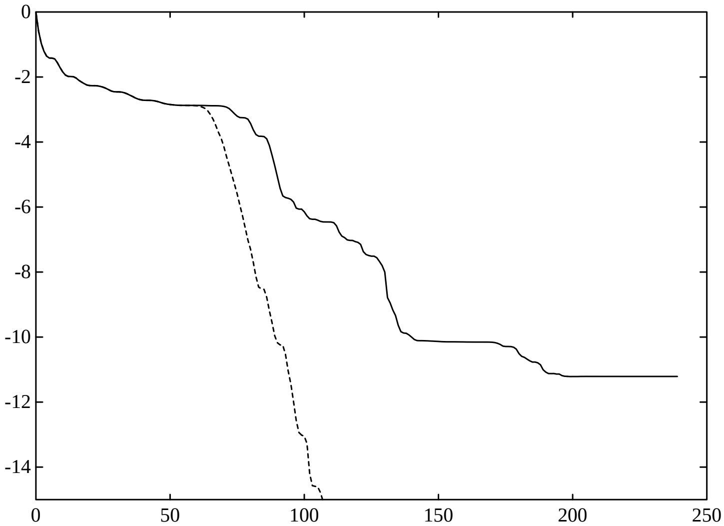
$$A = \text{diag}(-9, -7, \ldots, 189) - c$$

Implementation of CG based on

- three-term (–)
- MINRES (- -)

$$\log_{10}\left(\frac{\|f - Ax_n\|_2}{\|r_0\|_2}\right) \quad \text{against} \quad n$$

Stable Implementation (CR)



Shifted
"Van der Vorst matrix"

$$A = \text{diag}(-9, -7, \ldots, 189) - c$$

Implementation of CR based on

- three-term (–)
- MINRES (- -)

$$\log_{10}\left(\frac{\|f - Ax_n\|_2}{\|r_0\|_2}\right) \quad \text{against} \quad n$$

MINRES

MINRES (Idea)

$$\langle \psi_j, \psi_k \rangle_{MR} = \delta_{j,k}.$$

and

$$r_n^{MR} = p_n^{MR}(A) r_0, \quad p_n^{MR}(t) = \frac{\sum_{j=0}^n \psi_j(0)\psi_j(t)}{\sum_{j=0}^n \psi_j(0)^2}.$$

Alternative

$$p_n^{MR}(t) = 1 - t \sum_{j=0}^{n-1} y_j \psi_j(t) = 1 - t\Psi_{n-1}(t) y$$

with

$$\Psi_n(t) := [\psi_0(t),\ \psi_1(t), \ldots,\ \psi_n(t)],\ y := [y_0,\ y_1, \ldots,\ y_{n-1}]^T.$$

Jacobi Matrix

Expansion (three-term recurrence relation)

$$t\psi_{j-1}(t) = \sum_{k=j-1}^{j+1} h_{k,j}\,\psi_{k-1}(t), \quad \text{with} \quad h_{k,j} = \langle t\psi_{j-1}, \psi_{k-1}\rangle$$

Example

$$\begin{aligned} t\psi_0(t) &= \alpha_1\psi_0(t) + \beta_2\psi_1(t) \\ t\psi_1(t) &= \beta_2\psi_0(t) + \alpha_2\psi_1(t) + \beta_3\psi_2(t) \\ t\psi_2(t) &= 0 \qquad\qquad + \beta_3\psi_1(t) + \alpha_3\psi_2(t) + \alpha_4\psi_3(t) \end{aligned}$$

Matrix notation

$$t\Psi_2(t) = \Psi_3(t) \begin{pmatrix} \alpha_1 & \beta_2 & 0 \\ \beta_2 & \alpha_2 & \beta_3 \\ 0 & \beta_3 & \alpha_3 \\ 0 & 0 & \alpha_4 \end{pmatrix} = \Psi_3(t) H_3^E.$$

Least-squares problem

$$\begin{aligned} p_n(t) &= 1 - t\Psi_{n-1}(t)y \\ &= 1 - \Psi_n(t)H_n^E y \\ &= \Psi_n(t)(d_n - H_n^E y), \quad d_n := \psi_0^{-1} e_1, \end{aligned}$$

and

$$\langle p_n, p_n \rangle_{MR} = (d_n - H_n^E y)^T (d_n - H_n^E y).$$

Least-squares problem

$$\begin{aligned} \|r_n^{MR}\|_2^2 &= \|p_n^{MR}(A) r_0\|_2^2 \\ &= \langle p_n^{MR}, p_n^{MR} \rangle_{MR} \\ &= \|d_n - H_n^E y^{MR}\|_2^2. \end{aligned}$$

Computing Orthonormal Polynomials

Orthonormal polynomials

$$\begin{aligned}\beta_{j+1}\psi_j(t) &= (t-\alpha_j)\psi_{j-1}(t) - \beta_j\psi_{j-2}(t)\\ \alpha_j &= \langle t\psi_{j-1}, \psi_{j-1}\rangle, \quad \beta_j = \langle t\psi_{j-1}, \psi_{j-2}\rangle\end{aligned}$$

Stieltjes procedure

Set

$$\psi_{-1}(t) := 0, \quad \beta_1 = \sqrt{\langle 1, 1\rangle}, \quad \psi_0(t) = 1/\beta_1$$

Iterate: $j = 1, 2, \ldots, n$

$$\begin{aligned}\alpha_j &= \langle t\psi_{j-1}, \psi_{j-1}\rangle\\ \hat{\psi}_j(t) &= (t-\alpha_j)\psi_{j-1}(t) - \beta_j\psi_{j-2}(t)\\ \beta_{j+1} &= \|\hat{\psi}_j(t)\|, \quad \psi_j(t) = \hat{\psi}_j(t)/\beta_{j+1}\end{aligned}$$

Computing an Orthonormal Basis and H_n^E

Orthonormal polynomials

$$\delta_{j,k} = \langle \psi_j, \psi_k \rangle_{MR} = (\psi_j(A) r_0)^T \psi_k(A) r_0 = v_{j+1}^T v_{k+1}$$

provide orthonormal basis for $K_n(A, r_0) = \mathrm{span}\{v_1, v_2, \ldots, v_n\}$.

Stieltjes procedure translates into the **Lanczos process**

Set

$$v_0 := 0, \quad \beta_1 = \sqrt{\langle 1, 1 \rangle} = \|r_0\|_2, \quad v_1 = r_0/\beta_1$$

Iterate: $j = 1, 2, \ldots, n-1$

$$\begin{aligned} \alpha_j &= \langle t\psi_{j-1}, \psi_{j-1} \rangle_{MR} = v_j^T A v_j \\ \hat{v}_{j+1} &= \hat{\psi}_j(A) r_0 = (A - \alpha_j I_N) v_j - \beta_j v_{j-1} \\ \beta_{j+1} &= \|\hat{\psi}_j(t)\|_{MR} = \|\hat{v}_{j+1}\|_2, \quad v_{j+1} = \hat{v}_{j+1}/\beta_{j+1} \end{aligned}$$

Tridiagonalization

$$H_n = J_n = \left(\begin{array}{ccccc} \alpha_1 & \beta_2 & 0 & \cdots & 0 \\ \beta_2 & \ddots & \ddots & \ddots & \vdots \\ 0 & \ddots & \ddots & \ddots & 0 \\ \vdots & \ddots & \ddots & \ddots & \beta_n \\ 0 & \cdots & 0 & \beta_n & \alpha_n \\ \hline 0 & \cdots & \cdots & 0 & \alpha_{n+1} \end{array}\right) = J_n^E = H_n^E.$$

Tridiagonalization: from

$$t\Psi_{n-1}(t) = \Psi_{n-1}(t)H_n + \beta_{n+1}\,\psi_n(t)e_n^T = \Psi_n(t)H_n^E$$

it follows

$$AV_n = V_n H_n + \beta_{n+1} v_{n+1} e_n^T = V_{n+1} H_n^E$$

and

$$V_n^T A V_n = H_n.$$

MINRES (Implementation, Paige/Saunders 75)

Recall

$$p_j^{MR}(t) = 1 - t\Psi_{j-1}(t)y^{MR}$$

where

$$\|d_j - H_j^E y^{MR}\|_2 = \min\{\|d_j - H_j^E y\|_2 : \; y \in \mathbb{R}^j\}$$

$$\begin{aligned}
v_j &\leftarrow \text{Lanczos} \\
c_j, s_j, r_{k,j} &\leftarrow \text{Givens} \\
\eta_{j-1} &= -s_{j-1}\eta_{j-2} \\
w_j &= (v_j - r_{2j}w_{j-1} - r_{3j}w_{j-2})/r_{1j}, \\
x_j^{MR} &= x_{j-1}^{MR} + c_j\eta_{j-1}w_j, \\
r_j^{MR} &= r_{j-1}^{MR} - c_j\eta_{j-1}Aw_j
\end{aligned}$$

MINRES (properties)

- **No** breakdown possible
- **Stable implementation of CG**, if $c_j \neq 0$ then

$$x_j^{GAL} = x_j^{MR} + \frac{s_j^2}{c_j} \eta_{j-1} w_j$$

- **Peaks and plateaus** ($s_j^2 + c_j^2 = 1$)

$$\|r_j^{MR}\|_2 = |s_j| \|r_{j-1}^{MR}\|_2, \quad \|r_j^{GAL}\|_2 = \|r_j^{MR}\|_2 / |c_j|$$

- **more expensive**

Method	Aw	$v^T w$	$\alpha v + w$	storage
CG	1	2	6N	4
CR	1	2	8N	5
MINRES	1	2	12N	7

SYMMLQ

SYMMLQ (Definition 1/2)

Krylov Subspace

$$K_n(A, r_0) := \text{span}\{r_0, Ar_0, A^2r_0, \ldots, A^{n-1}r_0\}$$

CG approach

$$\begin{aligned} \|x_* - x_n^{GAL}\|_A &= \min\{\|x_* - x\|_A : x \in x_0 + K_n(A, r_0)\} \\ r_n^{GAL} &\perp K_n(A, r_0), \quad r_n^{GAL} = f - Ax_n^{GAL}. \end{aligned}$$

CR, MINRES approach

$$\begin{aligned} \|f - Ax_n^{MR}\|_2 &= \min\{\|f - Ax\|_2 : x \in x_0 + K_n(A, r_0)\} \\ r_n^{MR} &\perp AK_n(A, r_0), \quad r_n^{MR} = f - Ax_n^{MR}. \end{aligned}$$

SYMMLQ (Definition 2/2)

A ”non-feasible” approach

$$\|x_* - x_n\|_2 = \min\{\|x_* - x\|_2 : \; x \in x_0 + K_n(A, r_0)\}$$

An economic implementation requires the knowledge of the solution $x_* \ldots$

SYMMLQ approach

$$\begin{aligned} \|x_* - x_n^{ME}\|_2 &= \min\{\|x_* - x\|_2 : \; x \in x_0 + AK_n(A, r_0)\} \\ r_n^{ME} &\perp K_n(A, r_0), \quad r_n^{ME} = f - Ax_n^{ME}. \end{aligned}$$

SYMMLQ (Implementation, Paige/Saunders 75)

Here

$$p_j^{ME}(t) = 1 - t^2 \Psi_{j-1}(t) y^{ME}$$

where

$$(H_j^E)^T H_j^E y^{ME} = d_j.$$

$$\begin{aligned}
v_{j+1} &\leftarrow \text{Lanczos} \\
c_j, s_j, r_{k,j} &\leftarrow \text{Givens} \\
\eta_j &= -(r_{3j}\eta_{j-2} + r_{2j}\eta_{j-1})/r_{1j} \\
w_j &= -s_{j-1} w_{j-1} + c_{j-1} v_j, \\
x_j^{ME} &= x_{j-1}^{ME} + \eta_j (c_j w_j + s_j v_{j+1}), \\
r_j^{ME} &= r_{j-1}^{ME} - \eta_{j-1} A (c_j w_j + s_j v_{j+1})
\end{aligned}$$

SYMMLQ (properties)

- **No** breakdown possible

- **Stable implementation of CG**: if $c_j \neq 0$ then

$$x_j^{GAL} = x_j^{ME} + \frac{\eta_j}{c_j} w_j.$$

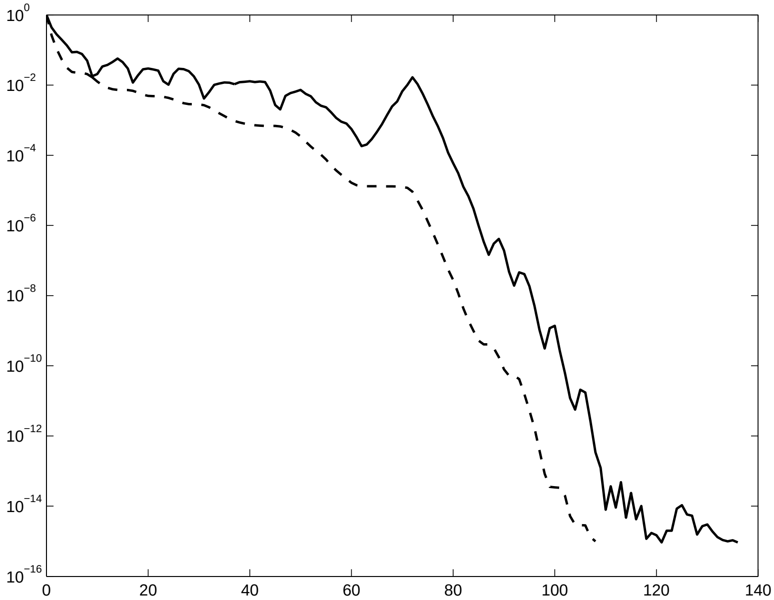
- **Stability**: More stable than MINRES (Sleijpen, van der Vorst, Modersitzki, 2000).

- **Speed**: May converge slower than MINRES, in particular for ill-conditioned systems.

- **Operation count**:

Method	Aw	$v^T w$	$\alpha v + w$	storage
MINRES	1	2	12N	7
SYMMLQ	1	2	12N	6

MINRES versus SYMMLQ



$$\log_{10}\left(\frac{\|f - Ax_n\|_2}{\|r_0\|_2}\right) \quad \text{against} \quad n$$

Shifted
"Van der Vorst matrix"

$$A = \mathrm{diag}(-9, -7, \ldots, 189) - c$$

- SYMMLQ (–)
- MINRES (- -)

Nonsymmetric Systems

Nonsymmetric case (general 1/2)

Scheme

$$\begin{aligned} x_n &= x_0 + q_{n-1}(A), \\ r_n &= p_n(A)r_0,\ p_n(0) = 1, \quad A \neq A^T. \end{aligned}$$

Ideal method

1. Fast convergence, e.g., minimal residual property

$$\|r_n\| = \min\{\|p(A)r_0\| :\ p \in \Pi_n,\ p(0) = 1\}.$$

2. Efficient implementation: work and storage requirement should not grow with n (short recurrences).

Nonsymmetric case (general 1/2)

Bad (or good) **news**

Polynomial based (Krylov subspace) methods with (1) **and** (2) exist only for

$$A = e^{i\theta}(B + \sigma I_N), \; B = B^T, \; \sigma \in \mathbb{C}$$

Voevodin 83, Faber/Manteuffel 84

Nonsymmetric Systems (Program)

- Fast Convergence
  - Generalized Minimal RESidual (GMRES)
- Efficient Implementation
  - BI Conjugate Gradient (Bi-CG)
- ”Both”
  - Bi-CG STABilized (Bi-CGSTAB)
  - Quasi Minimal Residual (QMR)
- ”Poor man’s version”
  - CG, CR applied to the normal equations

Test matrix (1/2)

Advection diffusion equation: vertical wind $w = (0, 1)$, no source $f = 0$, simple geometry Ω = unit square

$$\begin{aligned} -\nu \nabla^2 u + \frac{\partial u}{\partial y} &= 0 \quad \text{in} \quad \Omega \\ u &= g \quad \text{on} \quad \partial\Omega \end{aligned}$$

System matrix ($N^2 \times N^2$ block tridiagonal)

$$A = \begin{pmatrix} T_1 & T_2 & & & 0 \\ T_3 & T_1 & T_2 & & \\ & \ddots & \ddots & \ddots & \\ & & T_3 & T_1 & T_2 \\ 0 & & & T_3 & T_1 \end{pmatrix}$$

with T_j ($N \times N$) tridiagonal.

Test matrix (2/2)

Eigenvalues and eigenvectors

$$A v_{jk} = \sigma_{jk}\, v_{jk}, \quad j,k = 1,2,\ldots,N,$$

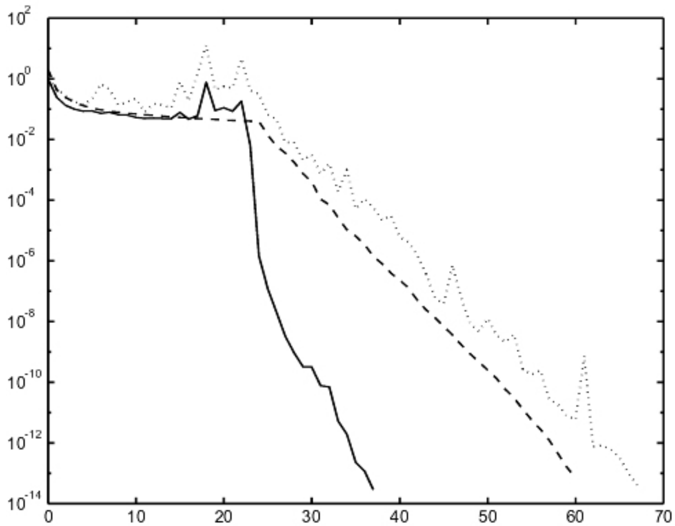
with

$$\begin{aligned} \sigma_{jk} &= \lambda_j + 2\sqrt{\mu_j \gamma_j}\cos\frac{k\pi}{N+1}, \\ v_{jk} &= \left(\sqrt{\frac{\gamma_j}{\mu_j}}\sin\frac{k\pi}{N+1}u_j^T, \ldots, \sqrt{\frac{\gamma_j}{\mu_j}}^N \sin\frac{N\pi}{N+1}u_j^T\right)^T \end{aligned}$$

and

$$T_1 u_j = \lambda_j u_j,\ T_2 u_j = \mu_j u_j,\ T_3 u_j = \gamma_j u_j.$$

Advection-diffusion (example)



Advection-diffusion matrix

$$A \in \mathbb{R}^{625 \times 625}$$

$$\log_{10}\left(\frac{\|f - Ax_n\|_2}{\|r_0\|_2}\right)$$ against n, BiCGStab (–), GMRES (- -), BiCG (...)

CG applied to the normal equations

Normal equations come in two versions

$$A^T A x = A^T f$$

or

$$A A^T y = f, \quad x = A^T y.$$

If we apply CG to the first version, we obtain for the kth iterate

$$\begin{aligned} \|x_k - x_*\|_{A^T A}^2 &= (x_k - x_*)^T A^T A (x_k - x_*) \\ &= \|A(x_k - x_*)\|_2^2 = \|r_k\|_2^2 \end{aligned}$$

in the Krylov subspace

$$x_0 + K_n(A^T r_0, A^T A).$$

This method is called **CGNR**, where the letter ”R” indicates that the residual is minimized (Hesteness, Stiefel, 53).

CG applied to the normal equations (Craig)

If we apply CG to the second version,

$$AA^T y = f, \quad x = A^T y.$$

we obtain for the kth iterate

$$\begin{aligned} \|y_k - y_*\|^2_{A^T A} &= (y_k - y_*)^T A^T A (y_k - y_*) \\ &= \|A(y_k - y_*)\|_2^2 = \|x_k - x_*\|_2^2 \end{aligned}$$

in the Krylov subspace

$$A^T x_0 + A^T K_n(r_0, A^T A).$$

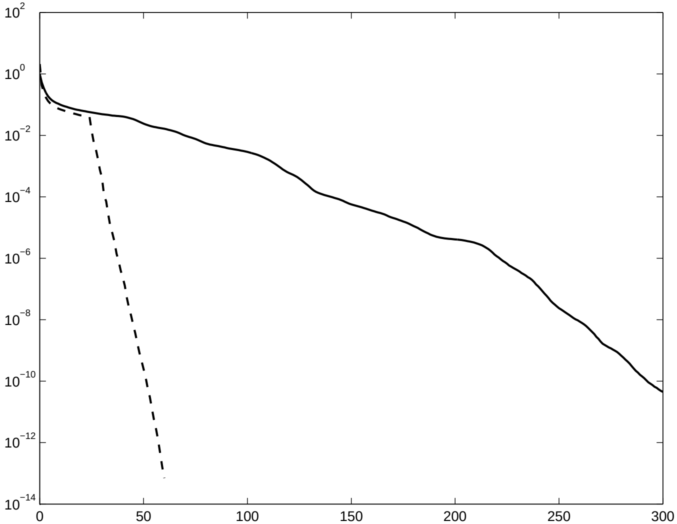
This method is called **CGNE**, where the letter ”E” indicates that the error is minimized (Craig, 55).

CG applied to the normal equations (Properties)

Properties:

- **Short** recurrences
- Optimality: regular convergence behaviour
- **two** matvecs one with the **transpose** matrix
- Condition number is squared: slow convergence

CGNR versus GMRES



Advection-diffusion matrix

$$A \in \mathbb{R}^{625\times 625}$$

$$\log_{10}\left(\frac{\|f - Ax_n\|_2}{\|r_0\|_2}\right)$$ against n, CGNR (–), GMRES (- -)

GMRES

Symmetric versus Nonsymmetric (1/2)

Recall

$$t\psi_{j-1}(t) = \sum_{k=1}^{j+1} h_{k,j}\, \psi_{k-1}(t), \quad \text{with} \quad h_{k,j} = \langle t\psi_{j-1}, \psi_{k-1} \rangle$$

Crucial observation:

- if $\langle t\psi_{j-1}, \psi_{k-1} \rangle = \langle \psi_{j-1}, t\psi_{k-1} \rangle$ then

$$h_{k,j} = 0, \quad \text{for} \quad k = 1, 2, \ldots, j-2$$

  - **Three-term recurrence relation**
  - H_n is **tridiagonal**

Symmetric versus Nonsymmetric (2/2)

- if $\langle t\psi_{j-1}, \psi_{k-1}\rangle \neq \langle \psi_{j-1}, t\psi_{k-1}\rangle$ then (in general)

$$h_{k,j} \neq 0, \quad \text{for} \quad k = 1, 2, \dots, j+1$$

  – **No** short recurrence

  – H_n is **full** upper Hessenberg

”Our inner product”

$$\langle p, q\rangle_{MR} := (p(A)r_0)^T q(A)r_0 = r_0^T p(A^T)q(A)r_0$$

Symmetric A implies **short** recurrences

$$\langle t\psi_{j-1}, \psi_{k-1}\rangle_{MR} = \langle \psi_{j-1}, t\psi_{k-1}\rangle_{MR} \quad \Leftrightarrow \quad A = A^T$$

Computing an Orthonormal Basis (Arnoldi 51)

Inner product, orthonormal polynomials, orthonormal vectors

$$\delta_{j,k} = \langle \psi_j, \psi_k \rangle_{MR} = (\psi_j(A) r_0)^T \psi_k(A) r_0 = v_{j+1}^T v_{k+1}$$

Orthonormal basis of $K_n(A, r_0)$ and entries of H_n^E

Set

$$v_1 = \psi_0(A) r_0 = r_0 / \|r_0\|_2$$

Iterate: $j = 1, 2 \ldots, n-1$

$$\begin{aligned} h_{k,j} &= \langle t\psi_{j-1}, \psi_{k-1} \rangle_{GAL} = (Av_j)^T v_k, \; k = 1, \ldots, j \\ \hat{v}_{j+1} &= \hat{\psi}_j(A) r_0 = (A - h_{j,j} I_N) v_j - \sum_{k=1}^{j-1} h_{k,j} v_k \\ h_{j+1,j} &= \|\hat{\psi}_j\|_{GAL} = \|\hat{v}_{j+1}\|_2, \quad v_{j+1} = \hat{v}_{j+1} / h_{j+1,j} \end{aligned}$$

GMRES (definition)

Scheme

$$r_n = p_n^{MR}(A) r_0$$

with minimal residual property

$$\|r_n\|_2 = \min\{\|p(A)r_0\| :\ p \in \Pi_n,\ p(0) = 1\}.$$

Implementation based on the same ideas as MINRES, that is

$$p_n^{MR}(t) = 1 - t\Psi_{n-1}(t) y^{MR} = \Psi_n(t)(d_n - H_n^E y^{MR})$$

where $d_n = \|r_0\|_2 e_1$ and

$$\|d_n - H_n^E y^{MR}\|_2 = \min\{\|d_n - H_n^E y\|_2 :\ y \in \mathbb{R}^n\}$$

Note: H_n^E has always full rank.

GMRES (implementation, Saad/Schultz 86)

Implementation (main ingredients)

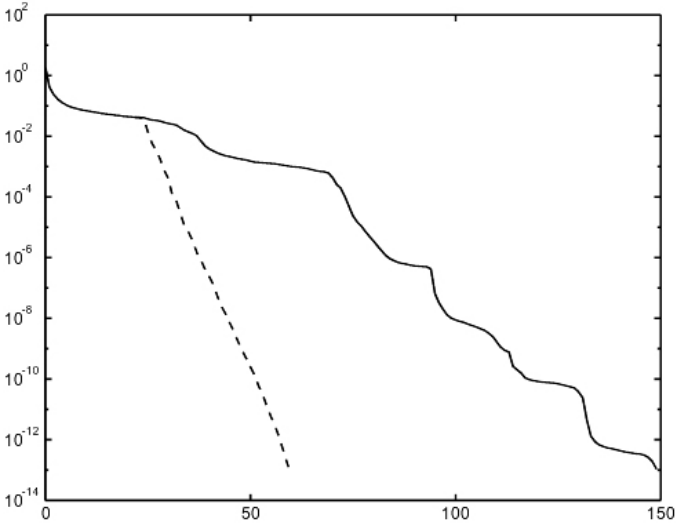
- Arnoldi process: v_j and H_j^E
- Least squares problem: y_j^{MR}
- Update:

$$
\begin{aligned}
r_j &= V_{j+1}(d_j - H_j^E y^{MR}) \\
x_j &= x_0 + AV_j y^{MR}
\end{aligned}
$$

Properties

- **No** breakdown possible
- Optimal w.r.t matvecs: m steps $\longrightarrow$ m matvecs
- **BUT**: computational costs m^2; memory space m $\hookrightarrow$ restarted GMRES(m)

GMRES versus Restarted GMRES



Advection-diffusion matrix

$$A \in \mathbb{R}^{625 \times 625}$$

$$\log_{10}\left(\frac{\|f - Ax_n\|_2}{\|r_0\|_2}\right)$$ against n, GMRES(20) (–), GMRES (- -)

BiCG

Bilinear Form (Formally Orthogonal Polynomials 1/2)

$$\langle p, q \rangle_{BI} := (p(A^T)s_0)^T q(A) r_0 = s_0^T p(A) q(A) r_0$$

(BiCG, BiCGSTAB, QMR)

Properties:

- **short** recurrences

$$\langle tp, q \rangle_{BI} = \langle p, tq \rangle_{BI}$$

- $\langle \cdot, \cdot \rangle_{BI}$ is not a positive definite inner product, there can exist polynomials $p \neq 0$ of any degree such that

$$\langle p, p \rangle_{BI} < 0 \quad \text{or} \quad \langle p, p \rangle_{BI} = 0 \quad (\Rightarrow \textbf{ breakdown})$$

Bilinear Form (Formally Orthogonal Polynomials 2/2)

- Orthogonality (of polynomials) translates into Biorthogonality (for residuals)

$$s_j^T r_k = (p_j(A^T)s_0)^T p_k(A) r_0 = \langle p_j, p_k \rangle_{BI}$$

  where $r_k = p_k(A)r_0$ and $s_j = p_j(A^T)s_0$

- For symmetric A and $s_0 = r_0$, we have

$$\langle p, q \rangle_{BI} = \langle p, q \rangle_{MR}$$

Computing Formally Orthogonal Polynomials

Formally orthogonal polynomials

$$\langle \varphi_j, \varphi_k \rangle_{BI} = (\varphi_j(A^T)s_0)^T \varphi_k(A) r_0 = \begin{cases} \neq 0 & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases}$$

Stieltjes procedure (monic version)

Set

$$\varphi_{-1}(t) = 0, \quad \varphi_0(t) = 1$$

Iterate: $j = 1, 2, \ldots, n$

$$\begin{aligned} \alpha_j &= \frac{\langle t\varphi_{j-1}, \varphi_{j-1} \rangle_{BI}}{\langle \varphi_{j-1}, \varphi_{j-1} \rangle_{BI}}, \ \beta_j = \frac{\langle \varphi_{j-1}, \varphi_{j-1} \rangle_{BI}}{\langle \varphi_{j-2}, \varphi_{j-2} \rangle_{BI}} \\ \varphi_j(t) &= (t - \alpha_j)\varphi_{j-1}(t) - \beta_j \varphi_{j-2}(t) \end{aligned}$$

Computing a Biorthogonal Basis (1/2)

Formally orthogonal polynomials

$$\langle \varphi_j, \varphi_k \rangle_{BI} = (\varphi_j(A^T)s_0)^T \varphi_k(A) r_0 = w_{j+1}^T v_{k+1}$$

generate a basis for

$$K_n(A, r_0) = \mathrm{span}\{v_1, v_2, \ldots, v_n\}$$

and for

$$K_n(A^T, s_0) = \mathrm{span}\{w_1, w_2, \ldots, w_n\}$$

(s_n shadow residual)

Computing a Biorthogonal Basis (2/2)

Nonsymmetric Lanczos process ("monic" version)

Choose s_0,$r_0 \neq 0$ (arbitrary)

Set

$$v_0 = w_0 = 0, \quad w_1 = s_0, \quad v_1 = r_0$$

Iterate: $j = 1, 2, \ldots, n-1$

$$\begin{aligned}
\alpha_j &= \frac{\langle t\varphi_{j-1}, \varphi_{j-1} \rangle_{BI}}{\langle \varphi_{j-1}, \varphi_{j-1} \rangle_{BI}} = \frac{w_j^T A v_j}{w_j^T v_j}, \\
\beta_j &= \frac{\langle \varphi_{j-1}, \varphi_{j-1} \rangle_{BI}}{\langle \varphi_{j-2}, \varphi_{j-2} \rangle_{BI}} = \frac{w_j^T v_j}{w_{j-1}^T v_{j-1}} \\
v_{j+1} &= \varphi_j(A) r_0 = (A - \alpha_j I_N) v_j - \beta_j v_{j-1}, \\
w_{j+1} &= \varphi_j(A^T) s_0 = (A^T - \alpha_j I_N) w_j - \beta_j w_{j-1}
\end{aligned}$$

Biconjugate Gradient (BiCG), Lanczos 52, Fletcher 74

Scheme

$$r_n = p_n^{BI}(A) r_0; \quad s_n = p_n^{BI}(A^T) s_0$$

Implementation

$$\begin{aligned}
\hat{w}_{j-1} &= \nu_j \hat{w}_{j-2} - r_{j-1}, \quad \nu_j = \frac{s_{j-1}^T r_{j-1}}{s_{j-2}^T r_{j-2}} \\
w_{j-1}^* &= \nu_j w_{j-2}^* - s_{j-1} \\
x_j &= x_{j-1} + \eta_j \hat{w}_{j-1}, \quad \eta_j = -\frac{s_{j-1}^T r_{j-1}}{(w_{j-1}^*)^T A \hat{w}_{j-1}} \\
r_j &= r_{j-1} - \eta_j A\, \hat{w}_{j-1} \\
s_j &= s_{j-1} - \eta_j A^T\, w_{j-1}^*
\end{aligned}$$

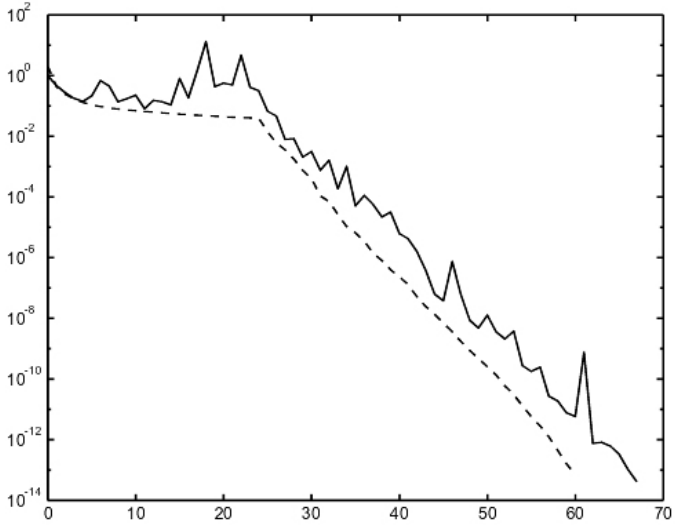
BiCG versus GMRES (1/2)

- BiCG
  - Three-term recurrence relation: storage requirement does **not** grow with j
  - **No** optimality property: convergence behaviour may be erratic (large jumps in $\|r_j\|_2$)
  - **Breakdown** can occur: $s_j^T r_j = 0$ with $s_j \neq 0, r_j \neq 0$ ($\Rightarrow$ **look-ahead strategies**)
  - Involves matrix-vector products $A^T w$ with the **transpose**

BiCG versus GMRES (2/2)

- GMRES
  - Long recurrence relation: storage requirement **does grow** with j
  - **Optimality** property: convergence behaviour is always smooth ($\|r_j\|_2$ monotonic)
  - **No** breakdown possible
  - Transpose A^T is **not** needed
  - No convergence results for restarted GMRES(m)

BiCG versus GMRES (Example)



Advection-diffusion matrix

$$A \in \mathbb{R}^{625 \times 625}$$

$\log_{10}\left(\frac{\|f - Ax_n\|_2}{\|r_0\|_2}\right)$ against n, BiCG (–) (**two** matvec's), GMRES (- -)

BiCGSTAB

Conjugate Gradient Squared (CGS), Sonneveld 89

Recall (BiCG)

$$r_n = p_n^{BI}(A) r_0$$

and (scalar needed in the implementation)

$$\nu_j = \frac{s_{j-1}^T r_{j-1}}{s_{j-2}^T r_{j-2}} = \frac{(p_j^{BI}(A^T) s_0)^T p_j^{BI}(A) r_0}{(p_{j-1}^{BI}(A^T) s_0)^T p_{j-1}^{BI}(A) r_0} = \frac{s_0^T (p_j^{BI}(A))^2 r_0}{s_0^T (p_{j-1}^{BI}(A))^2 r_0}$$

" $\Rightarrow$ " design method with

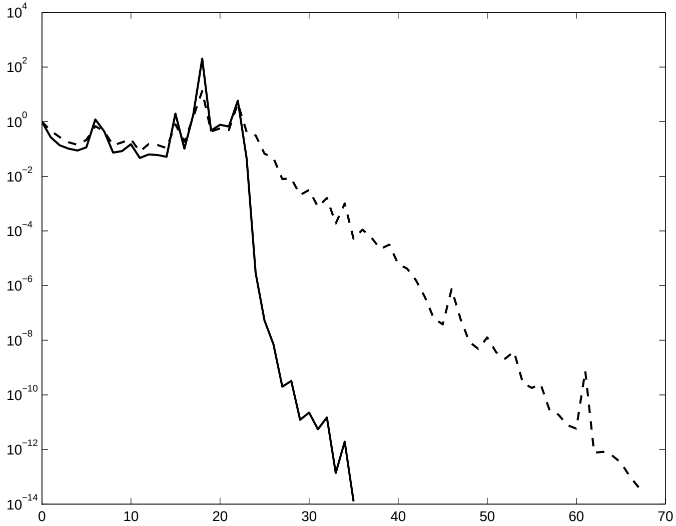
$$r_n = (\psi_n^{BI}(A))^2 r_0$$

Conjugate Gradient Squared (Properties)

Properties

- **Short** recurrences
- **No** optimality condition and rounding errors tend to be more damaging than in BiCG
- Does **not** require the transpose matrix (still **two** matvecs)
- **Breakdown** is possible
- **Starting point for BiCGSTAB**

CGS versus BiCG



Advection-diffusion matrix

$$A \in \mathbb{R}^{625\times 625}$$

$\log_{10}\left(\frac{\|f - Ax_n\|_2}{\|r_0\|_2}\right)$ against n, BiCG (– –), CGS (–)

Biconjugate Gradient Stabilized (BiCGSTAB), Van der Vorst 92

Idea: design method with

$$r_n = \psi_n^{BI}(A)\varphi_n(A)r_0$$

where

$$\varphi_n(t) = (1 - \tau_n t)\varphi_{n-1}(t)$$

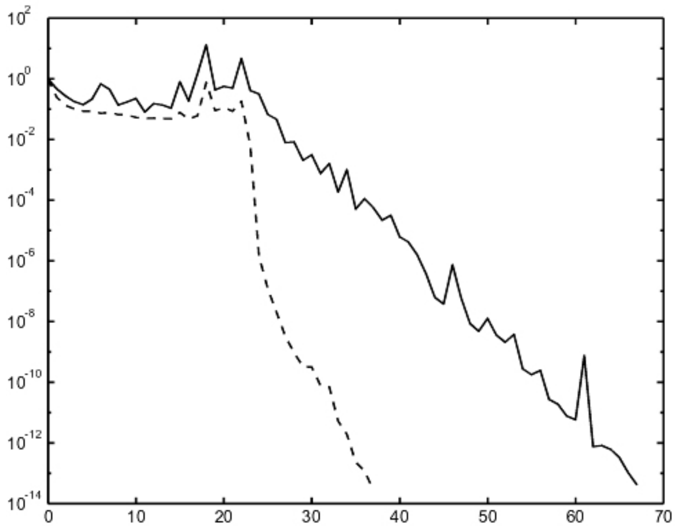
and τ_n is chosen to minimize $\|r_n\|$, with respect to τ_n
(one dimensional optimization problem)

Biconjugate Gradient Stabilized (Properties)

Properties

- **Short** recurrences
- **No** optimality condition
- **Smoother** convergence behaviour as compared to BiCG, CGS
- Does **not** require the transpose matrix (still **two** matvecs)
- **Breakdown** is possible
- May be viewed as the product of BiCG and GMRES(1)

BiCGSTAB versus BiCG

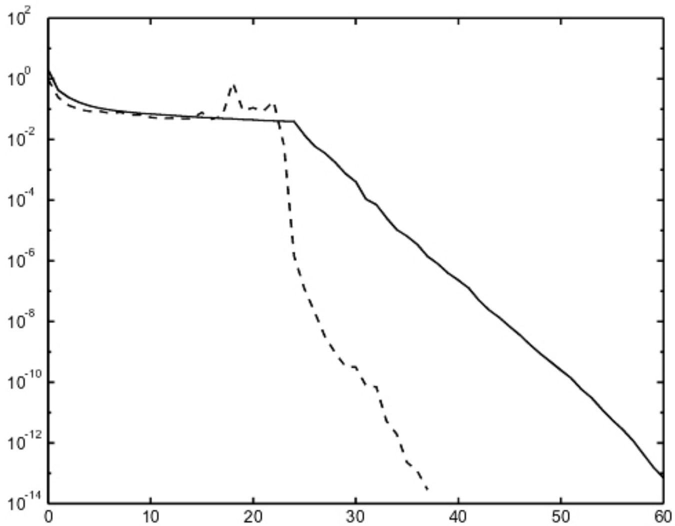


Advection-diffusion matrix

$$A \in \mathbb{R}^{625\times 625}$$

$$\log_{10}\left(\frac{\|f - Ax_n\|_2}{\|r_0\|_2}\right)$$ against n, BiCG (–), BiCGSTAB (- -)

BiCGSTAB versus GMRES



Advection-diffusion matrix

$$A \in \mathbb{R}^{625 \times 625}$$

$$\log_{10}\left(\frac{\|f - Ax_n\|_2}{\|r_0\|_2}\right),$$ GMRES (–), BiCGSTAB (- -) (**two** matvec's)

QMR

Quasi Kernel Polynomials (1/2)

Scheme

$$r_n = p_n^{QMR}(A) r_0$$

Formally orthogonal polynomials

$$\langle \varphi_j, \varphi_k \rangle_{BI} = (\varphi_j(A^T) s_0)^T \varphi_k(A) r_0 = \begin{cases} \neq 0 & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases}$$

Residual polynomials

$$p_j^{QMR}(t) = 1 - t\Phi_{j-1}(t) y^{QMR} = \Phi_j(t)(d_j - H_j^E y^{QMR})$$

where $\Phi_j(t) := [\varphi_0(t),\ \varphi_1(t), \ldots,\ \varphi_j(t)]$.

Quasi Kernel Polynomials (2/2)

Non-orthogonal matrix

$$V_j := \Phi_{j-1}(A)r_0 = [\varphi_0(A)r_0, \ldots, \varphi_{j-1}(A)r_0] = [v_1, \ldots, v_j]$$

leads to

$$\|r_j\|_2 = \|V_j(d_j - H_j^E y^{QMR})\|_2$$

Quasi kernel polynomials, quasi minimal residual

$$p_j^{QMR}(t) = 1 - t\Phi_{j-1}(t)y^{QMR}$$

with

$$\|d_j - H_j^E y^{QMR}\|_2 = \min\{\|d_j - H_j^E y\|_2 : \ y \in \mathbb{R}^j\}$$

(H_j^E is tridiagonal and has full rank)

Quasi Minimal Residual (QMR), Freund/Nachtigal 91

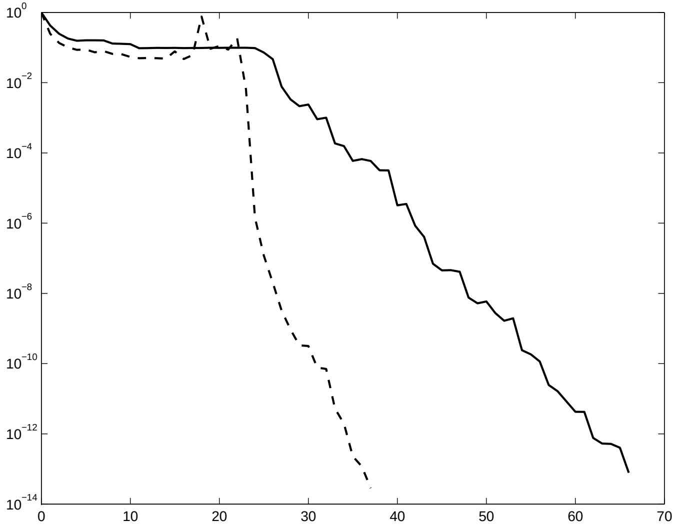
Implementation (main ingredients)

- Lanczos process: v_j and H_j^E
- Least squares problem: y_j^{QMR}
- Update: r_j, x_j

Properties

- **Short** recurrences: updates for r_j, x_j are similar to the one for MINRES
- **Quasi** optimality: regular convergence behaviour
- Requires the **transpose** matrix ($\Rightarrow$ TFQMR); in any case **two** matvecs
- **Breakdown** is possible ($\Rightarrow$ look-ahead strategies)
- Stable implementation of BiCG (see MINRES and CG)

QMR versus BiCGSTAB



Advection-diffusion matrix

$$A \in \mathbb{R}^{625\times 625}$$

$$\log_{10}\left(\frac{\|f - Ax_n\|_2}{\|r_0\|_2}\right)$$ against n, BiCGSTAB (– –), QMR (–)

list of slides

18. Prototype Algorithm
19. Minimal Residual Property (general 1/2)
20. Minimal Residual Property (general 2/2)
21. Minimal Residual Property (explicit solution)
22. Kernel Polynomials (properties)
23. Conjugate Residual Method
24. Prototype Algorithm
25. Conjugate Residual Method (implementation, Stiefel 57)
26. Conjugate Residual Method (convergence)
27. Residual Polynomials (convergence 1/2)
28. Residual Polynomials (convergence 2/2)
29. Conjugate Gradient Method (energy norm)
30. Conjugate Gradient Method (definition)
31. Conjugate Gradient Method (Kernel polynomials)
32. Prototype Algorithm
33. Conjugate Gradient Algorithm (Hestenes/Stiefel 52)
34. CG versus CR
35. CG versus CR (Residuals)
36. CG versus CR (Errors)

75. GMRES versus Restarted GMRES
76. Bilinear Form (Formally Orthogonal Polynomials 1/2)
77. Bilinear Form (Formally Orthogonal Polynomials 2/2)
78. Computing Formally Orthogonal Polynomials
79. Computing a Biorthogonal Basis (1/2)
80. Computing a Biorthogonal Basis (2/2)
81. Biconjugate Gradient (BiCG), Lanczos 52, Fletcher 74
82. BiCG versus GMRES (1/2)
83. BiCG versus GMRES (2/2)
84. BiCG versus GMRES (Example)
85. Conjugate Gradient Squared (CGS), Sonneveld 89
86. Conjugate Gradient Squared (Properties)
87. CGS versus BiCG
88. Biconjugate Gradient Stabilized (BiCGSTAB), Van der Vorst 92
89. Biconjugate Gradient Stabilized (Properties)
90. BiCGSTAB versus BiCG
91. BiCGSTAB versus GMRES
92. Quasi Kernel Polynomials (1/2)
93. Quasi Kernel Polynomials (2/2)

94. Quasi Minimal Residual (QMR), Freund/Nachtigal 91
95. QMR versus BiCGSTAB