CG versus MINRES on Positive Definite Systems

Michael Saunders
SOL and ICME, Stanford University
Joint work with David Fong

Recent Advances on Optimization
In honor of Philippe Toint
Toulouse, France, July 24–26, 2013

Meeting for Philippe

First thought:

The trust-region subproblem

$\|x_k\| \nearrow$ for CG

The trust-region subproblem

$$\min\ g^T p + \frac{1}{2} p^T H p \ \text{ st } \ \|p\|_M \leq \Delta$$

Apply PCG to $Hp = -g$
Exit if $d^T H d < 0$
or $\|p\|_M > \Delta$

The trust-region subproblem

$$\min\ g^T p + \frac{1}{2} p^T H p \ \text{ st } \ \|p\|_M \le \Delta$$

Apply PCG to $Hp = -g$
Exit if $d^T H d < 0$
or $\|p\|_M > \Delta$

Focus on $Ax = b$, $A \succ 0$, $M = I$
Backward errors
Stopping early

Part I: CG and MINRES

Iterative algorithms for $Ax = b$, $A = A^T$
based on the Lanczos process

Krylov-subspace methods: $x_k = V_k y_k$

Lanczos process (summary)

$$\beta_1 v_1 = b \qquad AV_k = V_{k+1} H_k$$

$$V_k = \begin{pmatrix} v_1 & v_2 & \dots & v_k \end{pmatrix}$$

$$T_k = \begin{pmatrix} \alpha_1 & \beta_2 & & \\ \beta_2 & \alpha_2 & \ddots & \\ & \ddots & \ddots & \beta_k \\ & & \beta_k & \alpha_k \end{pmatrix}$$

$$H_k = \begin{pmatrix} & T_k & \\ 0 \; \dots \; 0 & & x \end{pmatrix}$$

Lanczos process (summary)

$$\beta_1 v_1 = b \qquad AV_k = V_{k+1} H_k$$

$$V_k = \begin{pmatrix} v_1 & v_2 & \dots & v_k \end{pmatrix}$$

$$T_k = \begin{pmatrix} \alpha_1 & \beta_2 & & \\ \beta_2 & \alpha_2 & \ddots & \\ & \ddots & \ddots & \beta_k \\ & & \beta_k & \alpha_k \end{pmatrix}$$

$$H_k = \begin{pmatrix} & T_k & \\ 0 \dots 0 & & x \end{pmatrix}$$

$$\begin{aligned} r_k &= b - Ax_k \\ &= \beta_1 v_1 - AV_k y_k \\ &= V_{k+1}(\beta_1 e_1 - H_k y_k), \end{aligned}$$

Aim: $\beta_1 e_1 \approx H_k y_k$

Lanczos process (summary)

$$\beta_1 v_1 = b \qquad AV_k = V_{k+1} H_k$$

$$V_k = \begin{pmatrix} v_1 & v_2 & \dots & v_k \end{pmatrix}$$

$$T_k = \begin{pmatrix} \alpha_1 & \beta_2 & & \\ \beta_2 & \alpha_2 & \ddots & \\ & \ddots & \ddots & \beta_k \\ & & \beta_k & \alpha_k \end{pmatrix}$$

$$H_k = \begin{pmatrix} & T_k & \\ 0 \ \dots & 0 & x \end{pmatrix}$$

$$\begin{aligned} r_k &= b - Ax_k \\ &= \beta_1 v_1 - AV_k y_k \\ &= V_{k+1}(\beta_1 e_1 - H_k y_k), \end{aligned}$$

Aim: $\beta_1 e_1 \approx H_k y_k$

Two subproblems

CG	$T_k y_k = \beta_1 e_1$	$x_k = V_k y_k$
MINRES	$\min \lVert H_k y_k - \beta_1 e_1 \rVert$	$x_k = V_k y_k$

Common practice

$$Ax = b, \quad A = A^T$$

Common practice

$$Ax = b, \quad A = A^T$$

A positive definite	$\Rightarrow$	Use CG
A indefinite	$\Rightarrow$	Use MINRES

Common practice

$$Ax = b, \quad A = A^T$$

A positive definite $\Rightarrow$ Use CG

A indefinite $\Rightarrow$ Use MINRES

Experiment: CG vs MINRES on $A \succ 0$

Common practice

$$Ax = b, \quad A = A^T$$

A positive definite	$\Rightarrow$	Use CG
A indefinite	$\Rightarrow$	Use MINRES

Experiment: CG vs MINRES on $A \succ 0$

- Hestenes and Stiefel (1952) proposed both CG and CR for $A \succ 0$ and proved many properties
- CR $\equiv$ MINRES when $A \succ 0$
  They both minimize $\|r_k\| = \|b - Ax_k\|$ in the Krylov subspace

Theoretical properties for $Ax = b$, $A \succ 0$

		CG	CR (MINRES)
$\|x^* - x_k\|$	$\searrow$	HS 1952	HS 1952
$\|x^* - x_k\|_A$	$\searrow$	HS 1952	HS 1952
$\|x_k\|$	$\nearrow$	Steihaug 1983	Fong 2011

Theoretical properties for $Ax = b$, $A \succ 0$

		CG	CR (MINRES)
$\|x^* - x_k\|$	$\searrow$	HS 1952	HS 1952
$\|x^* - x_k\|_A$	$\searrow$	HS 1952	HS 1952
$\|x_k\|$	$\nearrow$	Steihaug 1983	Fong 2011

		CR (MINRES)
$\|r_k\|$	$\searrow$	HS 1952
$\|r_k\| / \|x_k\|$	$\searrow$	Fong 2011

Backward error for square systems $Ax = b$

Stopping tolerances α, β

x_k is an acceptable solution iff there exist E, f st

$$(A+E)x_k = b+f \qquad \frac{\|E\|}{\|A\|} \le \alpha \qquad \frac{\|f\|}{\|b\|} \le \beta$$

Backward error for square systems $Ax = b$

Stopping tolerances α, β

x_k is an acceptable solution iff there exist E, f st

$$(A+E)x_k = b+f \qquad \frac{\|E\|}{\|A\|} \le \alpha \qquad \frac{\|f\|}{\|b\|} \le \beta$$

Smallest perturbations E, f: (Titley-Peloquin 2010)

$$E = \frac{\alpha\|A\|}{\psi\|x_k\|} r_k x_k^T \qquad \frac{\|E\|}{\|A\|} = \alpha\frac{\|r_k\|}{\psi}$$

$$f = -\frac{\beta\|b\|}{\psi} r_k \qquad \frac{\|f\|}{\|b\|} = \beta\frac{\|r_k\|}{\psi}$$

Backward error for square systems $Ax = b$

Stopping tolerances α, β

x_k is an acceptable solution iff there exist E, f st

$$(A+E)x_k = b+f \qquad \frac{\|E\|}{\|A\|} \le \alpha \qquad \frac{\|f\|}{\|b\|} \le \beta$$

Smallest perturbations E, f: (Titley-Peloquin 2010)

$$E = \frac{\alpha\|A\|}{\psi\|x_k\|} r_k x_k^T \qquad \frac{\|E\|}{\|A\|} = \alpha\frac{\|r_k\|}{\psi}$$

$$f = -\frac{\beta\|b\|}{\psi} r_k \qquad \frac{\|f\|}{\|b\|} = \beta\frac{\|r_k\|}{\psi}$$

Backward error: $\|r_k\|/\psi$

Stopping rule: $\|r_k\| \le \psi \equiv \alpha\|A\|\|x_k\| + \beta\|b\|$

Backward error for square systems, $\beta = 0$

$$(A + E_k)x_k = b$$

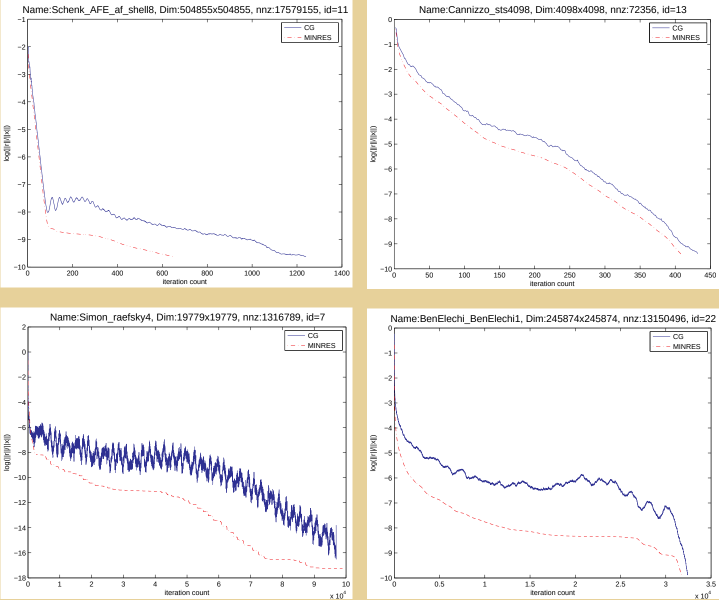
$$E_k = \frac{r_k x_k^T}{\|x_k\|^2} \qquad \|E_k\| = \frac{\|r_k\|}{\|x_k\|}$$

Data: Tim Davis's sparse matrix collection

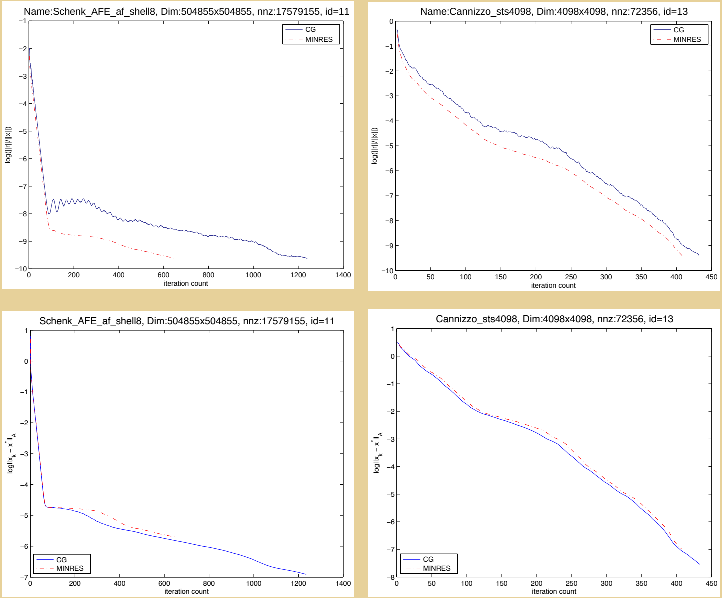
Real, symmetric posdef examples that include b

Plot $\log_{10}\|E_k\|$ for CG and MINRES

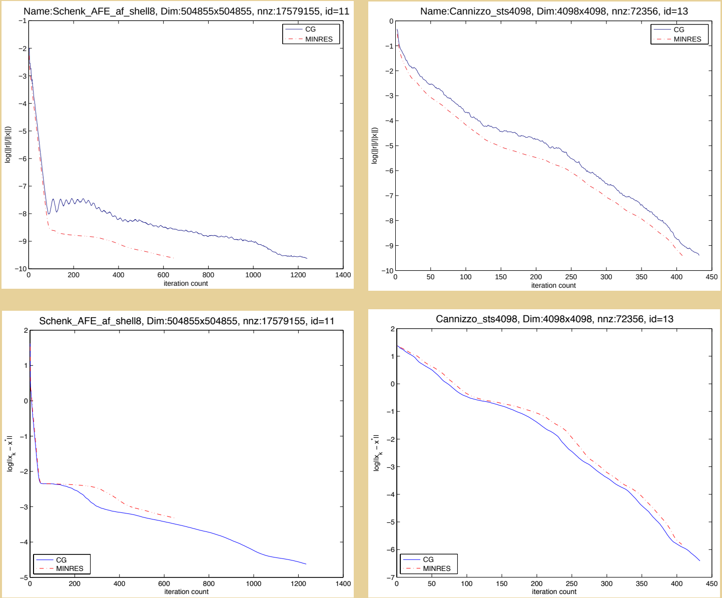
$\|r_k\|/\|x_k\|$ for $A \succ 0$ MINRES can stop sooner



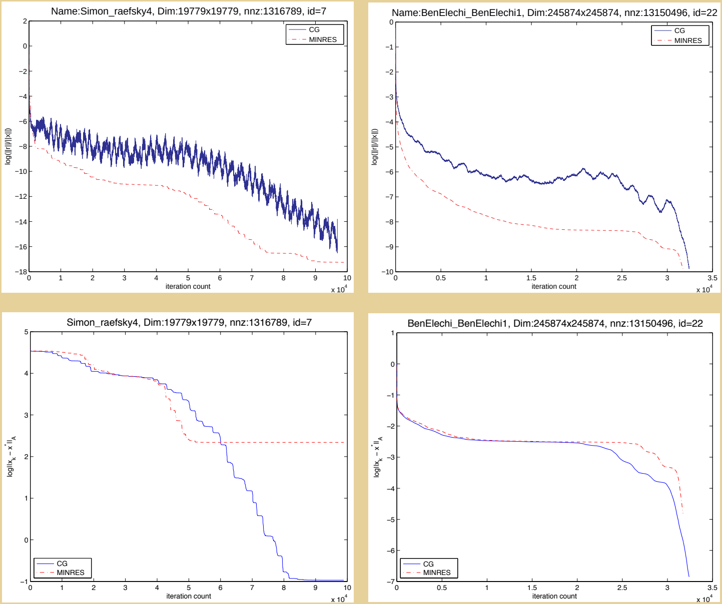
$\|r_k\|/\|x_k\|$ and $\log_{10}\|x - x_k\|_A$



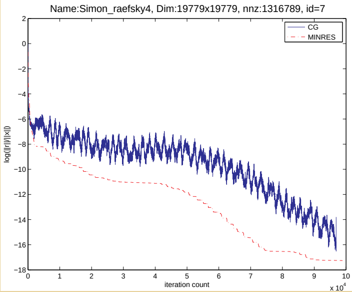
$\|r_k\|/\|x_k\|$ and $\log_{10}\|x - x_k\|$

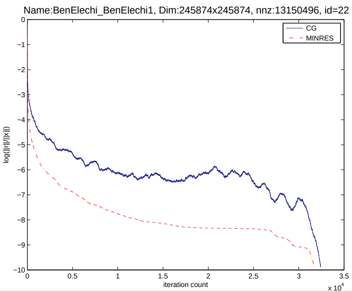


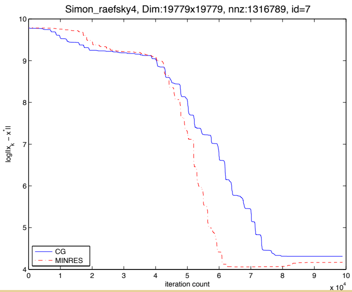
$\|r_k\| / \|x_k\|$ and $\log_{10} \|x - x_k\|_A$

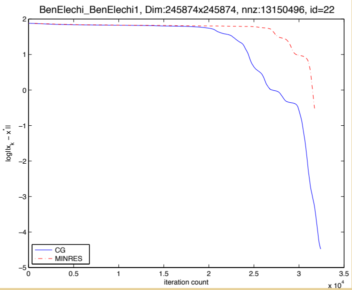


$\|r_k\|/\|x_k\|$ and $\log_{10}\|x - x_k\|$

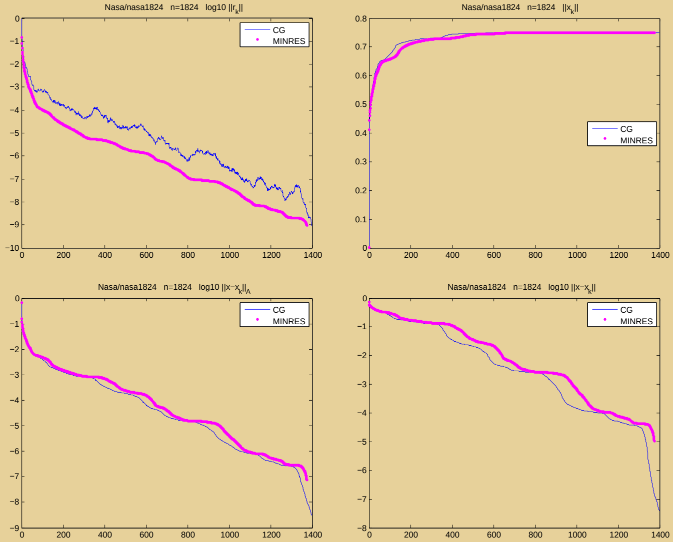




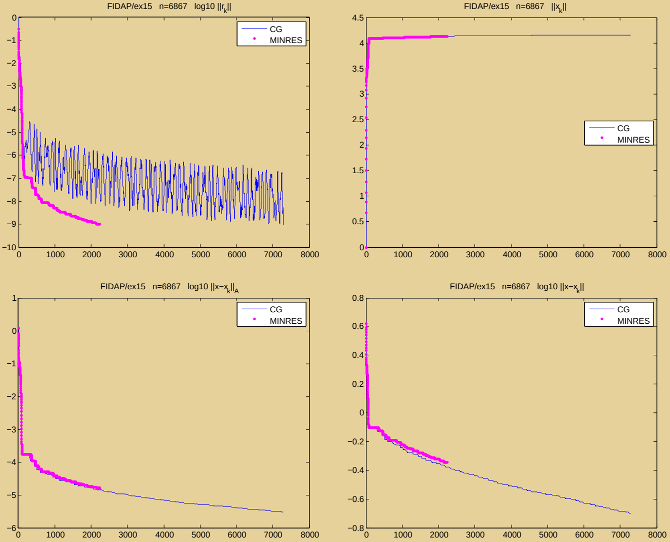




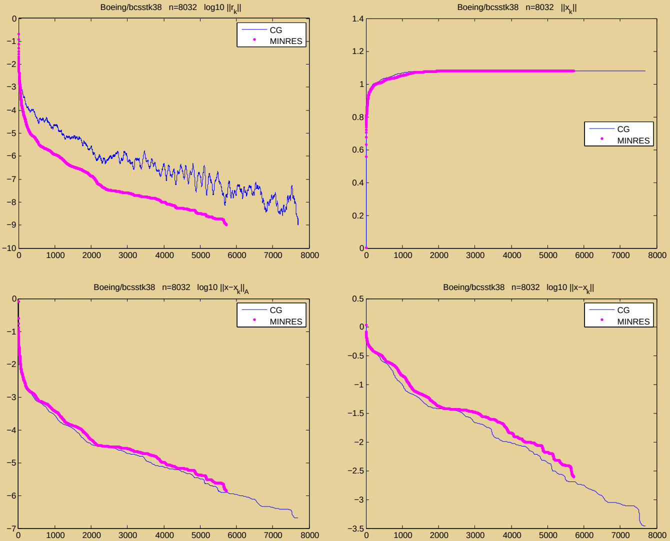
$\|r_k\| \quad \|x_k\| \quad \|x - x_k\|_A \quad \|x - x_k\| \quad \kappa(A) = 10^6$



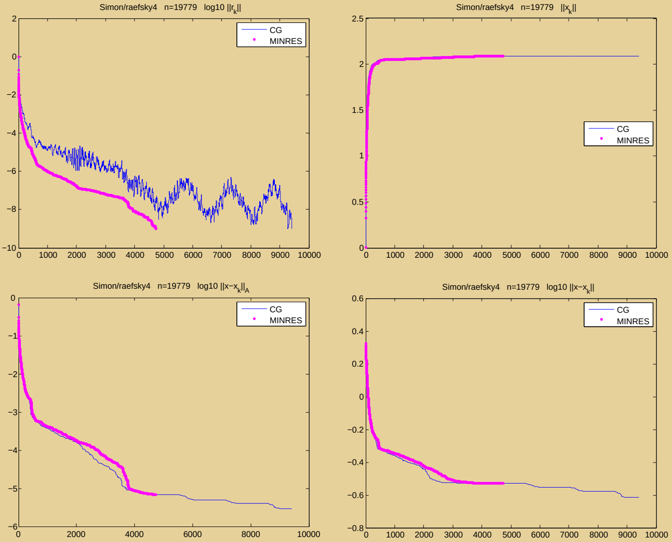
$\|r_k\| \quad \|x_k\| \quad \|x - x_k\|_A \quad \|x - x_k\| \quad \kappa(A) = 10^{11}$



$\|r_k\|$ $\|x_k\|$ $\|x - x_k\|_A$ $\|x - x_k\|$ $\kappa(A) = 10^7$



$\|r_k\| \quad \|x_k\| \quad \|x - x_k\|_A \quad \|x - x_k\| \quad \kappa(A) = 10^{12}$



Why does $\|r_k\|$ for CG lag behind MINRES?

Why does CG lag behind MINRES?

Greenbaum 1997: (Thanks David Titley-Peloquin)

$$\|r_k^C\| = \frac{\|r_k^M\|}{\sqrt{1 - \|r_k^M\|^2 / \|r_{k-1}^M\|^2}}$$

$\Rightarrow \|r_k^C\| \gg \|r_k^M\|$ if MINRES is almost stalling

Why does CG lag behind MINRES?

Greenbaum 1997: (Thanks David Titley-Peloquin)

$$\|r_k^C\| = \frac{\|r_k^M\|}{\sqrt{1 - \|r_k^M\|^2 / \|r_{k-1}^M\|^2}}$$

$\Rightarrow \|r_k^C\| \gg \|r_k^M\|$ if MINRES is almost stalling

Fong 2011: If $\alpha = 0$ in stopping rule, stop when $\|r_l\| \le \beta \|b\|$

$$\prod_{k=1}^{l} \frac{\|r_k\|}{\|r_{k-1}\|} = \frac{\|r_l\|}{\|b\|} \approx \beta$$

$\Rightarrow \|r_k^M\| / \|r_{k-1}^M\|$ closer to 1 on average if l is large
$\Rightarrow$ bigger gap on average between $\|r_k^C\|$ and $\|r_k^M\|$

Posdef systems and least squares

LS $\equiv$ PSD system

$$\min \|Ax - b\| \qquad \Rightarrow \qquad A^T Ax = A^T b$$

LS $\equiv$ PSD system

$$\min \|Ax - b\| \quad \Rightarrow \quad A^T A x = A^T b$$

Conversely, let $A = U^T U$ (Cholesky) and solve $U^T c = b$

PSD system $\equiv$ LS

$$Ax = b \quad \Rightarrow \quad U^T U x = U^T c \quad \Rightarrow \quad \min \|Ux - c\|$$

Part II: LSQR and LSMR

LSQR $\equiv$ CG on $A^TAx = A^Tb$

LSMR $\equiv$ MINRES on $A^TAx = A^Tb$

Based on Golub-Kahan bidiagonalization

Which problems do LSQR and LSMR solve?

solve $Ax = b$

Which problems do LSQR and LSMR solve?

solve $Ax = b$ $\qquad$ $\min \|x\|$ st $Ax = b$

Which problems do LSQR and LSMR solve?

solve $Ax = b$ $\quad\quad$ $\min \|x\|$ st $Ax = b$

$$\min \|Ax - b\|$$

Which problems do LSQR and LSMR solve?

solve $Ax = b$	$\min \|x\|$ st $Ax = b$
$\min \|Ax - b\|$	$\min \left\| \begin{pmatrix} A \\ \lambda I \end{pmatrix} x - \begin{pmatrix} b \\ 0 \end{pmatrix} \right\|$

Which problems do LSQR and LSMR solve?

solve $Ax = b$ $\qquad$ $\min \|x\|$ st $Ax = b$

$\min \|Ax - b\|$ $\qquad$ $\min \left\| \begin{pmatrix} A \\ \lambda I \end{pmatrix} x - \begin{pmatrix} b \\ 0 \end{pmatrix} \right\|$

- A square or rectangular ($m \times n$) and often sparse
- A can be an operator ($\Rightarrow$ allows preconditioning)
- Av, $A^T u$ plus $O(m+n)$ operations per iteration

Which problems do LSQR and LSMR solve?

solve $Ax = b$ $\qquad$ $\min \|x\|$ st $Ax = b$

$\min \|Ax - b\|$ $\qquad$ $\min \left\| \begin{pmatrix} A \\ \lambda I \end{pmatrix} x - \begin{pmatrix} b \\ 0 \end{pmatrix} \right\|$

- A square or rectangular ($m \times n$) and often sparse
- A can be an operator ($\Rightarrow$ allows preconditioning)
- Av, $A^T u$ plus $O(m+n)$ operations per iteration
- June 2013: F90 complex implementations
  Austin Benson and Victor Minden, ICME

LSQR on $Ax = b$

- $\|r_k\| \searrow 0$ by design
- $\|x_k\| \nearrow$ PS 1982
- Should have been a clue that $\|x_k\| \nearrow$ for CG!

LSQR on $Ax = b$

- $\|r_k\| \searrow 0$ by design
- $\|x_k\| \nearrow$ PS 1982
- Should have been a clue that $\|x_k\| \nearrow$ for CG!
- $\|r_k\| \leq \alpha\|A\|\|x_k\| + \beta\|b\|$ chosen by Chris Paige 1982
  We didn't know it was optimal till Titley-Peloquin 2010

LSQR on $Ax = b$

- $\|r_k\| \searrow 0$ by design
- $\|x_k\| \nearrow$ PS 1982
- Should have been a clue that $\|x_k\| \nearrow$ for CG!
- $\|r_k\| \leq \alpha\|A\|\|x_k\| + \beta\|b\|$ chosen by Chris Paige 1982
  We didn't know it was optimal till Titley-Peloquin 2010
- Realized last week:
  Backward errors are monotonic for LSQR on $Ax = b$
  (like CG and MINRES when $A \succ 0$):

$$\|r_k\|/(\alpha\|A\|\|x_k\| + \beta\|b\|) \quad \searrow$$

LSQR on $Ax = b$

- $\|r_k\| \searrow 0$ by design
- $\|x_k\| \nearrow$ PS 1982
- Should have been a clue that $\|x_k\| \nearrow$ for CG!
- $\|r_k\| \leq \alpha\|A\|\|x_k\| + \beta\|b\|$ chosen by Chris Paige 1982
  We didn't know it was optimal till Titley-Peloquin 2010
- Realized last week:
  Backward errors are monotonic for LSQR on $Ax = b$
  (like CG and MINRES when $A \succ 0$):

$$\|r_k\|/(\alpha\|A\|\|x_k\| + \beta\|b\|) \quad \searrow$$

Everything the same for LSMR!

LSQR on $Ax = b$

- $\|r_k\| \searrow 0$ by design
- $\|x_k\| \nearrow$ PS 1982
- Should have been a clue that $\|x_k\| \nearrow$ for CG!
- $\|r_k\| \leq \alpha\|A\|\|x_k\| + \beta\|b\|$ chosen by Chris Paige 1982
  We didn't know it was optimal till Titley-Peloquin 2010
- Realized last week:
  Backward errors are monotonic for LSQR on $Ax = b$
  (like CG and MINRES when $A \succ 0$):

$$\|r_k\|/(\alpha\|A\|\|x_k\| + \beta\|b\|) \quad \searrow$$

Everything the same for LSMR!

What about GMRES?

LSQR and LSMR on $\min \|Ax - b\|$

Stewart backward error

$$r_k = b - Ax_k$$

$$\frac{\|A^T r_k\|}{\|A\|\|r_k\|} \le \alpha$$

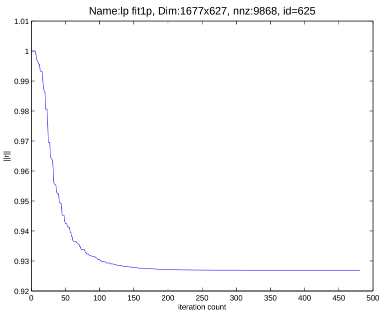
LSQR and LSMR on $\min \|Ax - b\|$

Stewart backward error

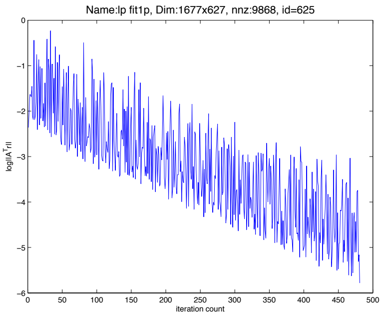
$$r_k = b - Ax_k$$

$$\frac{\|A^T r_k\|}{\|A\|\|r_k\|} \leq \alpha$$

LSQR $\|r_k\|$



LSQR $\log \|A^T r_k\|$



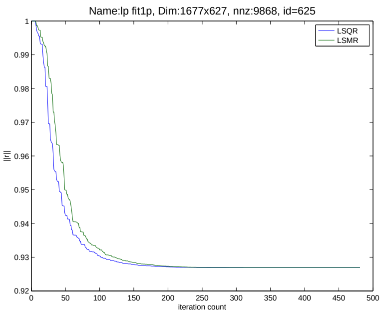
LSQR and LSMR on $\min \|Ax - b\|$

Stewart backward error

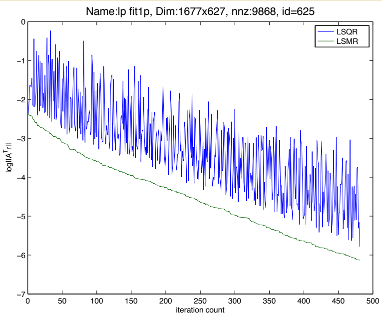
$$r_k = b - Ax_k$$

$$\frac{\|A^T r_k\|}{\|A\|\|r_k\|} \le \alpha$$

$\|r_k\|$



$\log \|A^T r_k\|$



Overdetermined systems

Test Data

- Tim Davis, University of Florida Sparse Matrix Collection
- $\mathrm{LPnetlib}$: Linear Programming Problems
- $\mathrm{A} = (\mathrm{Problem.A})'$ $\quad \mathrm{b} = \mathrm{Problem.c}$ $\quad$ (127 problems)

Overdetermined systems

Test Data

- Tim Davis, University of Florida Sparse Matrix Collection
- LPnetlib: Linear Programming Problems
- A = (Problem.A)' b = Problem.c (127 problems)

$$\text{Solve } \min \|Ax - b\|_2$$

with LSQR and LSMR

Backward error – estimates

$A^T A\hat{x} = A^T b$	$\hat{r} = b - A\hat{x}$	exact
$(A+E_i)^T(A+E_i)x = (A+E_i)^T b$	$r = b - Ax$	any x

Backward error – estimates

$$A^T A\hat{x} = A^T b \qquad \hat{r} = b - A\hat{x} \qquad \text{exact}$$

$$(A+E_i)^T(A+E_i)x = (A+E_i)^T b \qquad r = b - Ax \qquad \text{any } x$$

Two estimates given by Stewart (1975 and 1977)

$$E_1 = \frac{ex^T}{\|x\|^2} \qquad \|E_1\| = \frac{\|e\|}{\|x\|} \qquad e = \hat{r} - r$$

$$E_2 = -\frac{rr^TA}{\|r\|^2} \qquad \|E_2\| = \frac{\|A^Tr\|}{\|r\|} \qquad \text{computable}$$

Backward error – estimates

$$A^T A\hat{x} = A^T b \qquad \hat{r} = b - A\hat{x} \qquad \text{exact}$$

$$(A+E_i)^T(A+E_i)x = (A+E_i)^T b \qquad r = b - Ax \qquad \text{any } x$$

Two estimates given by Stewart (1975 and 1977)

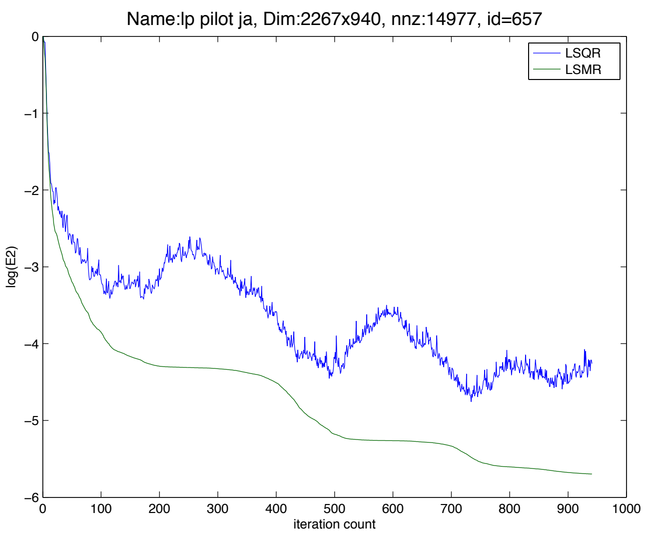
$$E_1 = \frac{ex^T}{\|x\|^2} \qquad \|E_1\| = \frac{\|e\|}{\|x\|} \qquad e = \hat{r} - r$$

$$E_2 = -\frac{rr^T A}{\|r\|^2} \qquad \|E_2\| = \frac{\|A^T r\|}{\|r\|} \qquad \text{computable}$$

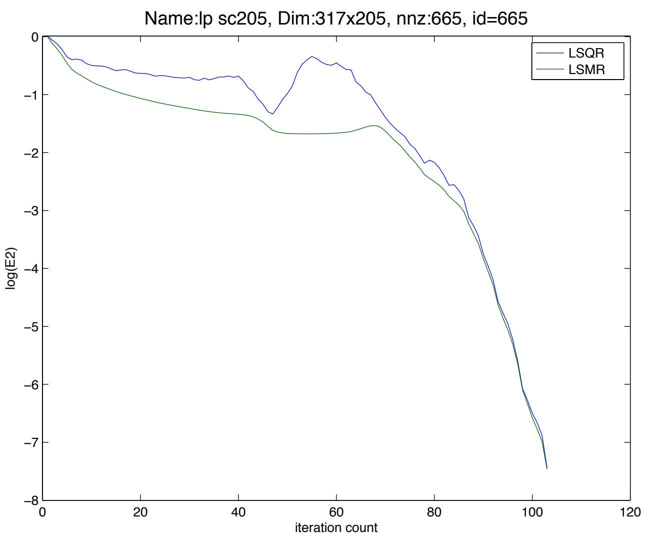
Theorem

$$\|E_2^{\text{LSMR}}\| \le \|E_2^{\text{LSQR}}\|$$

$\log_{10} \|E_2\|$ for LSQR and LSMR – typical



$\log_{10} \|E_2\|$ for LSQR and LSMR – rare



Backward error - optimal

$$\mu(x) \equiv \min_E \|E\| \quad \text{st} \quad (A+E)^T(A+E)x = (A+E)^T b$$

Exact $\mu(x)$ (Waldén, Karlson, & Sun 1995, Higham 2002)

$$C \equiv \left[A \quad \frac{\|r\|}{\|x\|}\left(I - \frac{rr^T}{\|r\|^2}\right)\right] \qquad \mu(x) = \sigma_{\min}(C)$$

Backward error - optimal

$$\mu(x) \equiv \min_E \|E\| \quad \text{st} \quad (A+E)^T(A+E)x = (A+E)^T b$$

Cheaper estimate $\tilde{\mu}(x)$ (Grcar, Saunders, & Su 2007)

$$K = \begin{pmatrix} A \\ \frac{\|r\|}{\|x\|} I \end{pmatrix} \qquad v = \begin{pmatrix} r \\ 0 \end{pmatrix}$$

$$\min_y \|Ky - v\| \qquad \tilde{\mu}(x) = \frac{\|Ky\|}{\|x\|}$$

Backward error - optimal

$$\mu(x) \equiv \min_E \|E\| \quad \text{st} \quad (A+E)^T(A+E)x = (A+E)^T b$$

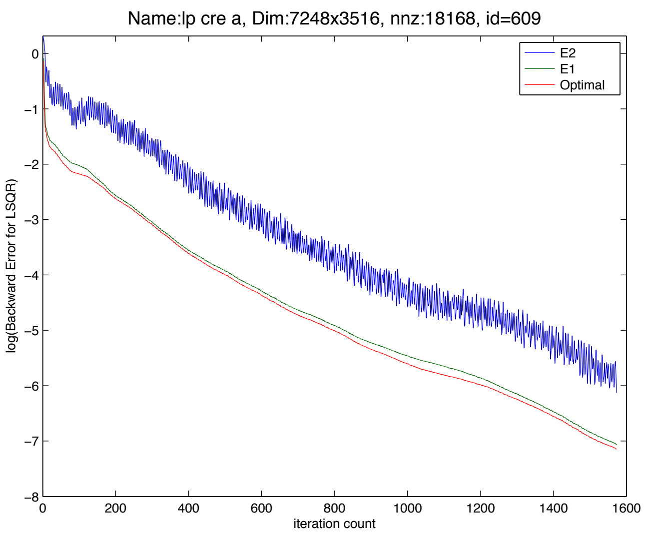
Cheaper estimate $\tilde{\mu}(x)$ (Grcar, Saunders, & Su 2007)

$$K = \begin{pmatrix} A \\ \frac{\|r\|}{\|x\|} I \end{pmatrix} \qquad v = \begin{pmatrix} r \\ 0 \end{pmatrix}$$

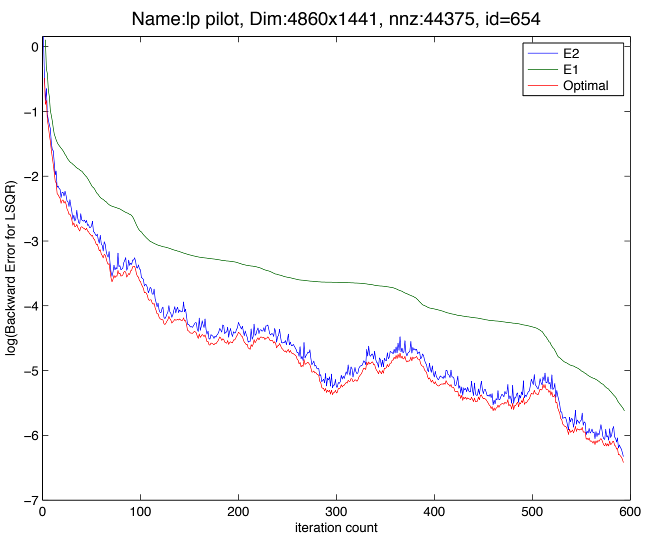
$$\min_y \|Ky - v\| \qquad \tilde{\mu}(x) = \frac{\|Ky\|}{\|x\|}$$

```
r       = b - A*x;
p       = colamd(A);
eta     = norm(r)/norm(x);
K       = [A(:,p); eta*speye(n)];
v       = [r; zeros(n,1)];
[c,R]   = qr(K,v,0);
mutilde = norm(c)/norm(x);
```

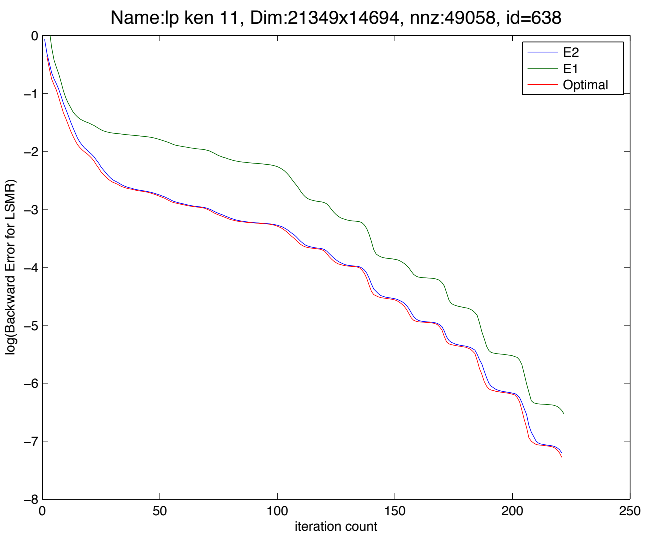
Backward errors for LSQR – typical



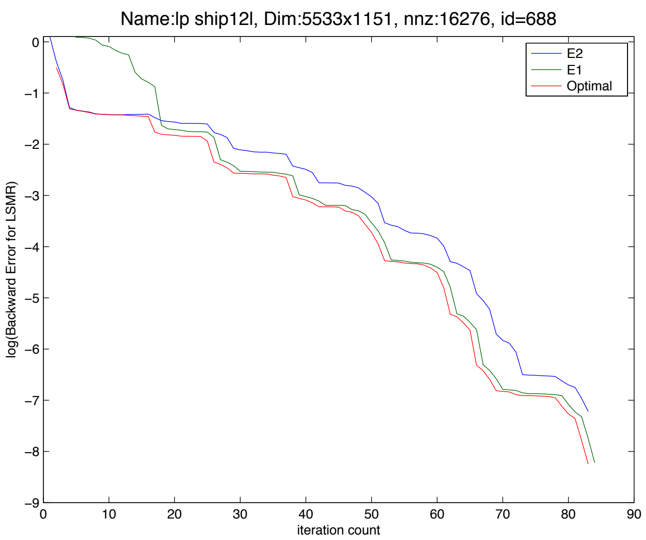
Backward errors for LSQR – rare



Backward errors for LSMR – typical



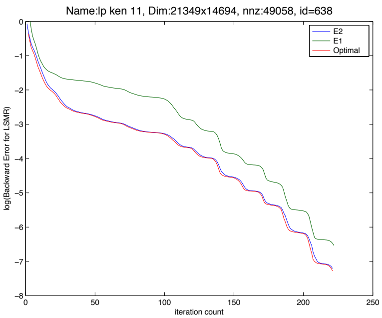
Backward errors for LSMR – rare



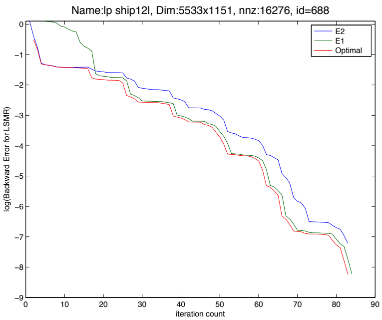
For LSMR

$\|E_2\| \approx$ optimal BE almost always

Typical: $\|E_2\| \approx \tilde{\mu}(x)$



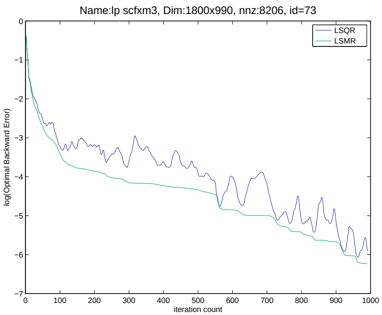
Rare: $\|E_1\| \approx \tilde{\mu}(x)$



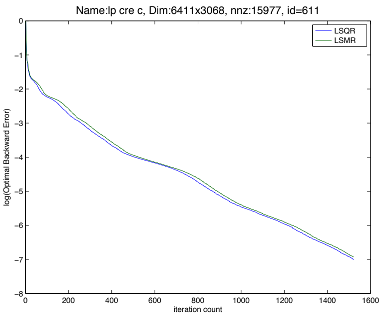
Optimal backward errors $\tilde{\mu}(x)$

Seem monotonic for LSMR Usually not for LSQR

Typical for LSQR and LSMR



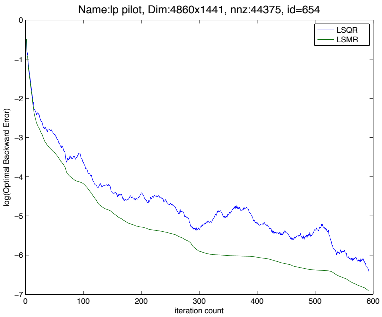
Rare LSQR, typical LSMR



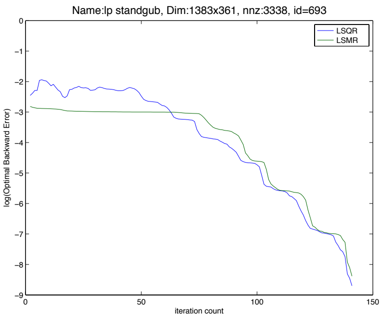
Optimal backward errors

$\tilde{\mu}(x^{\mathrm{LSMR}}) \leq \tilde{\mu}(x^{\mathrm{LSQR}})$ almost always

Typical



Rare



Summary

Theoretical properties for $Ax = b$

CG and MINRES, $A \succ 0$

$$\|x^* - x_k\|,\ \|x^* - x_k\|_A \quad \searrow$$

$$\|x_k\| \quad \nearrow$$

Theoretical properties for $Ax = b$

CG and MINRES, $A \succ 0$

$$\|x^* - x_k\|,\ \|x^* - x_k\|_A \quad \searrow$$

$$\|x_k\| \quad \nearrow$$

MINRES, $A \succ 0$

$$\|r_k\| \quad \searrow$$

$$\|r_k\|/(\alpha\|A\|\|x_k\| + \beta\|b\|) \quad \searrow$$

Theoretical properties for $Ax = b$

CG and MINRES, $A \succ 0$

$$\|x^* - x_k\|, \ \|x^* - x_k\|_A \quad \searrow$$
$$\|x_k\| \quad \nearrow$$

MINRES, $A \succ 0$

$$\|r_k\| \quad \searrow$$
$$\|r_k\|/(\alpha\|A\|\|x_k\| + \beta\|b\|) \quad \searrow$$

LSQR and LSMR, any A

$$\|r_k\| \quad \searrow$$
$$\|r_k\|/(\alpha\|A\|\|x_k\| + \beta\|b\|) \quad \searrow$$

Theoretical properties for $Ax = b$

CG and MINRES, $A \succ 0$

$$\|x^* - x_k\|,\ \|x^* - x_k\|_A \quad \searrow$$
$$\|x_k\| \quad \nearrow$$

MINRES, $A \succ 0$

$$\|r_k\| \quad \searrow$$
$$\|r_k\|/(\alpha\|A\|\|x_k\| + \beta\|b\|) \quad \searrow$$

LSQR and LSMR, any A

$$\|r_k\| \quad \searrow$$
$$\|r_k\|/(\alpha\|A\|\|x_k\| + \beta\|b\|) \quad \searrow$$

Monotonic backward errors
$\Rightarrow$ safe to stop early

Theoretical properties for $\min \|Ax - b\|$

LSQR and LSMR

$$\begin{array}{rl} \|x^* - x_k\|,\ \|r^* - r_k\| & \searrow \\ \|r_k\| & \searrow \\ \|x_k\| & \nearrow \\ x_k \to \text{min-length } x^* & \text{if } \operatorname{rank}(A) < n \end{array}$$

Theoretical properties for $\min \|Ax - b\|$

LSQR and LSMR

$$\begin{array}{rl} \|x^* - x_k\|,\ \|r^* - r_k\| & \searrow \\ \|r_k\| & \searrow \\ \|x_k\| & \nearrow \\ x_k \to \text{min-length } x^* & \text{if rank}(A) < n \end{array}$$

LSMR

$$\begin{array}{rl} \|A^T r_k\| & \searrow \\ \|A^T r_k\| / \|r_k\| & \searrow \text{ almost always} \\ & \approx \text{ optimal BE almost always} \\ & \leq \ (\|A^T r_k\| / \|r_k\|)^{\text{LSQR}} \end{array}$$

Theoretical properties for $\min \|Ax - b\|$

LSQR and LSMR

$$\begin{array}{rl} \|x^* - x_k\|,\ \|r^* - r_k\| & \searrow \\ \|r_k\| & \searrow \\ \|x_k\| & \nearrow \\ x_k \rightarrow \text{min-length } x^* & \text{if } \operatorname{rank}(A) < n \end{array}$$

LSMR

$$\begin{array}{rl} \|A^T r_k\| & \searrow \\ \|A^T r_k\| / \|r_k\| & \searrow \ \text{almost always} \\ & \approx \ \text{optimal BE almost always} \\ & \leq \ (\|A^T r_k\| / \|r_k\|)^{\text{LSQR}} \end{array}$$

For LSMR, optimal backward errors **seem** monotonic

$\Rightarrow$ safe to stop early

Final thoughts

We learn from history

Final thoughts

We learn from history
that we don't learn from history
– G. W. F. Hegel

Final thoughts

We learn from history
that we don't learn from history
– G. W. F. Hegel

If you're not fired with enthusiasm,

Final thoughts

We learn from history
that we don’t learn from history
– G. W. F. Hegel

If you’re not fired with enthusiasm,
you’ll be fired with enthusiasm
– Vince Lombardi

Final thoughts

We learn from history
that we don’t learn from history
– G. W. F. Hegel

If you’re not fired with enthusiasm,
you’ll be fired with enthusiasm
– Vince Lombardi

Special thanks

Magnus Hestenes & Eduard Stiefel
Gene Golub, William Kahan, Chris Paige
David Fong, David Titley-Peloquin
Andy, Nick, Philippe & Cerfacs friends
ONR (Office of Naval Research)