School of Engineering

CG and MINRES: An empirical comparison

Prequel to LSQR and LSMR: Two least-squares solvers

David Fong and Michael Saunders
Institute for Computational and Mathematical Engineering
Stanford University

5th International Conference on
High Performance Scientific Computing
March 5–9, 2012 — Hanoi, Vietnam

Abstract

For iterative solution of symmetric systems $Ax = b$, the conjugate gradient method (CG) is commonly used when A is positive definite, while the minimal residual method (MINRES) is typically reserved for indefinite systems. We investigate the sequence of solutions generated by each method and suggest that even if A is positive definite, MINRES may be preferable to CG if iterations are to be terminated early.

The classic symmetric positive-definite system comes from the full-rank least-squares (LS) problem $\min \|Ax - b\|$. Specialization of CG and MINRES to the associated normal equation $A^TAx = A^Tb$ leads to LSQR and LSMR respectively. We include numerical comparisons of these two LS solvers because they motivated this retrospective study of CG versus MINRES.

1. CG and MINRES
   - The Lanczos Process
   - Properties
   - Backward Errors
2. LSQR and LSMR
3. LSMR Derivation
   - Golub-Kahan bidiagonalization
   - Properties
4. LSMR Experiments
   - Backward Errors
5. Summary

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)

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

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

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

$$H_k = \begin{pmatrix} T_k \\ \beta_{k+1} e_k^T \end{pmatrix}$$

Lanczos process (summary)

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

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

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

$$H_k = \begin{pmatrix} T_k \\ \beta_{k+1} e_k^T \end{pmatrix}$$

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

Aim: $\beta_1 e_1 \approx H_k y_k$

Lanczos process (summary)

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

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

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

$$H_k = \begin{pmatrix} T_k \\ \beta_{k+1} e_k^T \end{pmatrix}$$

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

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	$\min \|H_k y_k - \beta_1 e_1\|$	$x_k = V_k y_k$

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 2012

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 2012

		CR (MINRES)
$\|r_k\|$	$\searrow$	HS 1952
$\|r_k\|/\|x_k\|$	$\searrow$	Fong 2012

Backward error for square systems $Ax = b$

An approximate solution x_k is acceptable iff $\exists\, E, f$ st

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

Backward error for square systems $Ax = b$

An approximate solution x_k is acceptable iff $\exists\, E, f$ 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 $Ax = b$

An approximate solution x_k is acceptable iff $\exists\, E, f$ 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)

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

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

Stopping rule:

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

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

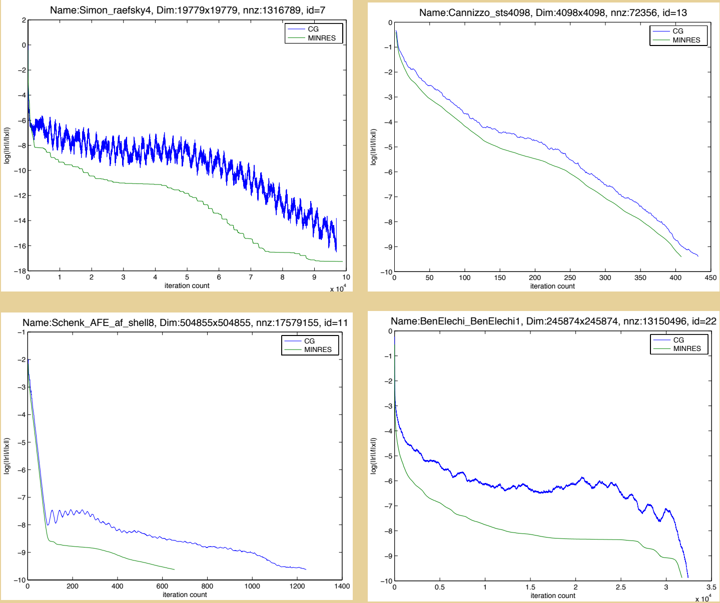
$$(A + E^{(k)})x_k = b$$

$$E^{(k)} = \frac{r_k x_k^T}{\|x_k\|^2} \qquad \|E^{(k)}\| = \frac{\|r_k\|}{\|x_k\|}$$

Data: Tim Davis's sparse matrix collection
Real, symmetric posdef examples that include b

Plot $\log_{10} \|E^{(k)}\|$ for CG and MINRES

Backward Error of CG vs MINRES on $A \succ 0$



Part II: LSQR and LSMR

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

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

What problems do LSQR and LSMR solve?

solve $Ax = b$

What problems do LSQR and LSMR solve?

solve $Ax = b$ $\qquad$ $\min \|Ax - b\|_2$

What problems do LSQR and LSMR solve?

solve $Ax = b$

$$\min \|Ax - b\|_2$$

$$\begin{array}{cc} \min & \|x\| \\ \text{st} & Ax = b \end{array}$$

What problems do LSQR and LSMR solve?

solve $Ax = b$

$$\min \|Ax - b\|_2$$

$$\begin{array}{cc} \min & \|x\| \\ \text{st} & Ax = b \end{array}$$

$$\min \left\| \begin{pmatrix} A \\ \lambda I \end{pmatrix} x - \begin{pmatrix} b \\ 0 \end{pmatrix} \right\|_2$$

What problems do LSQR and LSMR solve?

solve $Ax = b$

$$\min \|Ax - b\|_2$$

$$\begin{array}{cc} \min & \|x\| \\ \text{st} & Ax = b \end{array}$$

$$\min \left\| \begin{pmatrix} A \\ \lambda I \end{pmatrix} x - \begin{pmatrix} b \\ 0 \end{pmatrix} \right\|_2$$

Properties

- A is 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

Why invent another algorithm?

Reason one

CG vs MINRES

Reason two

Monotone convergence of residuals

$$\|r_k\| \quad \text{and} \quad \|A^T r_k\| \ \searrow$$

$\min \|Ax - b\|$

Measure of convergence

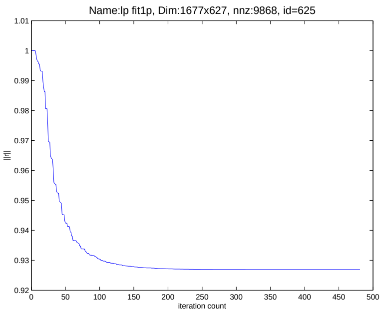
- $r_k = b - Ax_k$
- $\|r_k\| \to \|\hat{r}\|$, $\|A^T r_k\| \to 0$

$\min \|Ax - b\|$

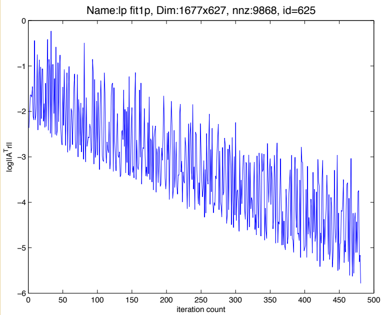
Measure of convergence

- $r_k = b - Ax_k$
- $\|r_k\| \to \|\hat{r}\|$, $\|A^T r_k\| \to 0$

LSQR $\|r_k\|$



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



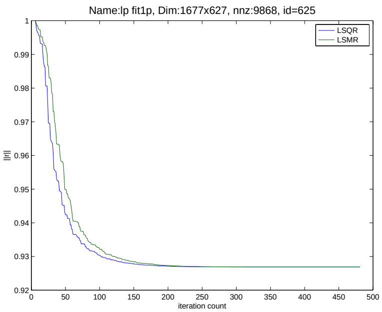
$\min \|Ax - b\|$

Measure of convergence

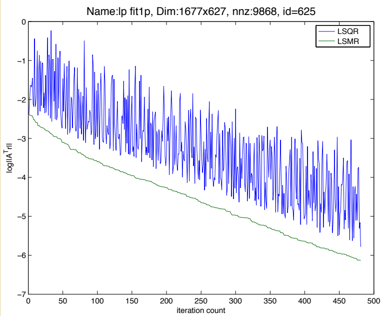
- $r_k = b - Ax_k$
- $\|r_k\| \to \|\hat{r}\|$, $\|A^T r_k\| \to 0$

— LSQR
— LSMR

$\|r_k\|$



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



LSMR Derivation

Golub-Kahan bidiagonalization

Given A $(m \times n)$ and b $(m \times 1)$

Direct bidiagonalization

$$U^T \begin{pmatrix} b & A \end{pmatrix} \begin{pmatrix} 1 & \\ & V \end{pmatrix} = \begin{pmatrix} \times & \times & & \\ & \times & \times & \\ & & \times & \times \\ & & & \times \end{pmatrix} \quad \Rightarrow \quad \begin{pmatrix} b & AV \end{pmatrix} = U \begin{pmatrix} \beta_1 e_1 & B \end{pmatrix}$$

Golub-Kahan bidiagonalization

Given A $(m \times n)$ and b $(m \times 1)$

Direct bidiagonalization

$$U^T \begin{pmatrix} b & A \end{pmatrix} \begin{pmatrix} 1 & \\ & V \end{pmatrix} = \begin{pmatrix} \times & \times & & \\ & \times & \times & \\ & & \times & \times \\ & & & \times \end{pmatrix} \quad \Rightarrow \quad \begin{pmatrix} b & AV \end{pmatrix} = U \begin{pmatrix} \beta_1 e_1 & B \end{pmatrix}$$

Iterative bidiagonalization $\mathrm{Bidiag}(A, b)$

Half a page in the 1965 Golub-Kahan SVD paper

Golub-Kahan bidiagonalization (2)

$$\begin{aligned} b &= U_{k+1}(\beta_1 e_1) \\ A V_k &= U_{k+1} B_k \\ A^T U_k &= V_k B_k^T \begin{pmatrix} I_k \\ 0 \end{pmatrix} \end{aligned}$$

where

$$B_k = \begin{pmatrix} \alpha_1 & & & \\ \beta_2 & \alpha_2 & & \\ & \ddots & \ddots & \\ & & \beta_k & \alpha_k \\ & & & \beta_{k+1} \end{pmatrix} \qquad \begin{aligned} U_k &= \begin{pmatrix} u_1 & \cdots & u_k \end{pmatrix} \\ V_k &= \begin{pmatrix} v_1 & \cdots & v_k \end{pmatrix} \end{aligned}$$

Golub-Kahan bidiagonalization (3)

V_k spans the Krylov subspace:

$$\text{span}\{v_1, \ldots, v_k\} = \text{span}\{A^T b,\ (A^T A) A^T b,\ \ldots,\ (A^T A)^{k-1} A^T b\}$$

Golub-Kahan bidiagonalization (3)

V_k spans the Krylov subspace:

$$\text{span}\{v_1, \ldots, v_k\} = \text{span}\{A^T b,\ (A^T A) A^T b,\ \ldots,\ (A^T A)^{k-1} A^T b\}$$

Define $x_k = V_k y_k$

Subproblem to solve

$$\min_{y_k} \|r_k\| = \min_{y_k} \|\beta_1 e_1 - B_k y_k\| \qquad \text{(LSQR)}$$

$$\min_{y_k} \|A^T r_k\| = \min_{y_k} \left\| \bar{\beta}_1 e_1 - \begin{pmatrix} B_k^T B_k \\ \bar{\beta}_{k+1} e_k^T \end{pmatrix} y_k \right\| \qquad \text{(LSMR)}$$

where $r_k = b - A x_k,\ \bar{\beta}_k = \alpha_k \beta_k$

Computational and storage requirement

	Storage		Work	
	m	n	m	n
MINRES on $A^TAx = A^Tb$	Av_1	x, v_1, v_2, w_1, w_2		8
LSQR	Av, u	x, v, w	3	5
LSMR	Av, u	$x, v, h, \bar{h}$	3	6

where h_k, $\bar{h}_k$ are scalar multiples of w_k, $\bar{w}_k$

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

		LSQR	LSMR
$\|x^* - x_k\|$	$\searrow$	HS 1952	HS 1952
$\|r^* - r_k\|$	$\searrow$	HS 1952	HS 1952
$\|x_k\|$	$\nearrow$	Steihaug 1983	Fong 2012
$\|r_k\|$	$\searrow$	PS 1982	Fong 2012

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

		LSQR	LSMR
$\|x^* - x_k\|$	$\searrow$	HS 1952	HS 1952
$\|r^* - r_k\|$	$\searrow$	HS 1952	HS 1952
$\|x_k\|$	$\nearrow$	Steihaug 1983	Fong 2012
$\|r_k\|$	$\searrow$	PS 1982	Fong 2012

		LSQR	LSMR
$\|A^T r_k\|$	$\searrow$		FS 2011
$\|A^T r_k\| / \|r_k\|$	$\searrow$		**mostly**

$$\|A^T r_k\| / \|r_k\| \quad \geq \quad \|A^T r_k\| / \|r_k\|$$

LSMR Experiments

Overdetermined systems

Test Data

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

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

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

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)

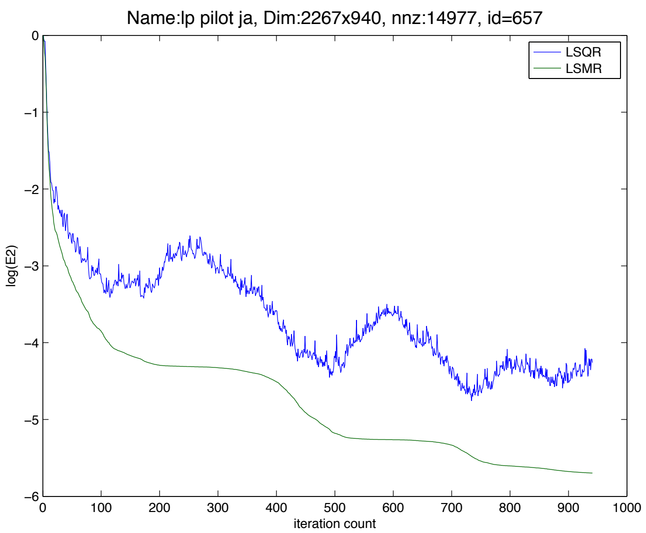
$$E_1 = \frac{ex^T}{\|x\|^2} \qquad \|E_1\| = \frac{\|e\|}{\|x\|} \qquad e = \hat{r} - r$$

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

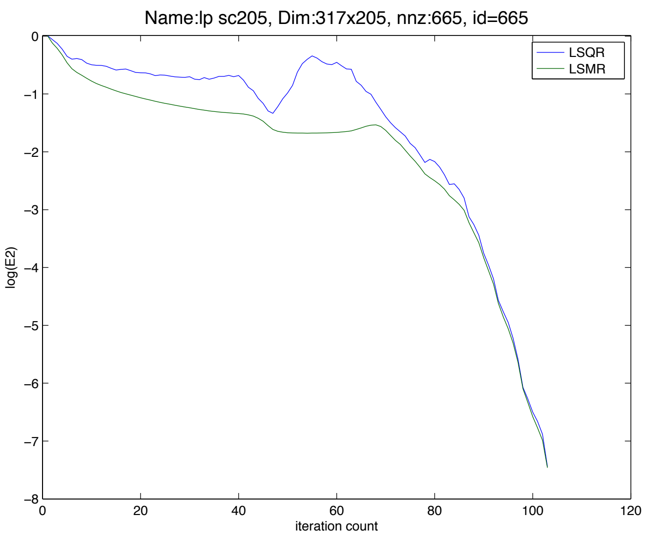
Theorem

$$\|E_2^{\mathrm{LSMR}}\| \leq \|E_2^{\mathrm{LSQR}}\|$$

$\log_{10} \|E_2\|$ 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)

$$C \equiv \begin{bmatrix} A & \frac{\|r\|}{\|x\|}\left(I - \frac{rr^T}{\|r\|^2}\right) \end{bmatrix} \qquad \mu(x) = \sigma_{\min}(C)$$

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

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

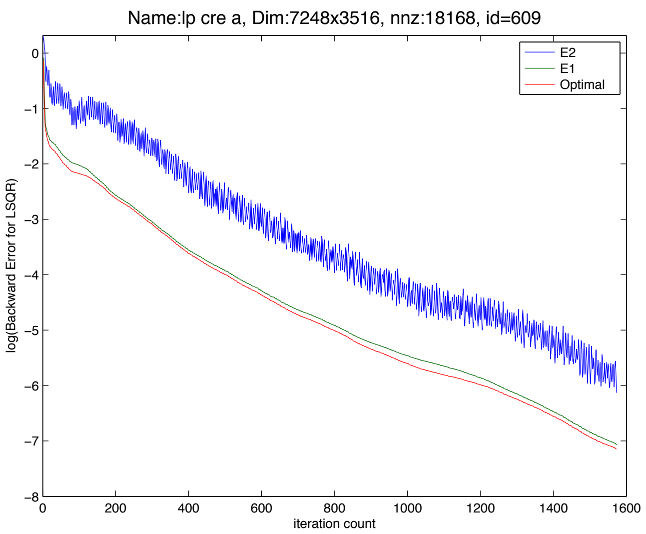
Cheaper estimate $\tilde{\mu}(x)$ (Grcar, Saunders, & Su 2007)

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

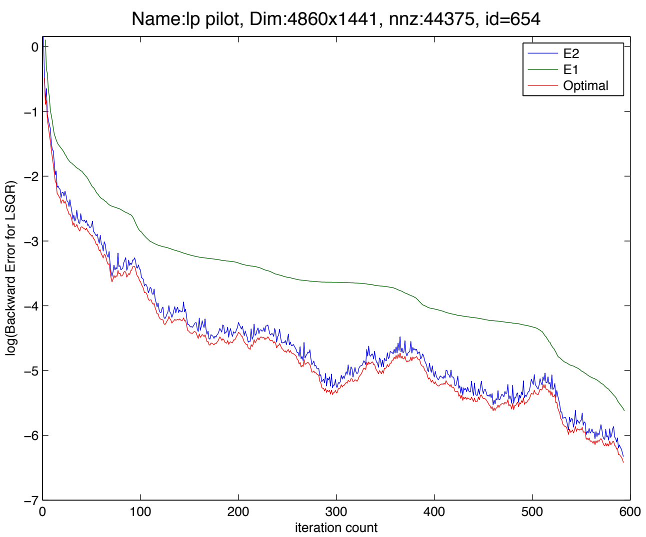
$$\min_y \|Ky - v\| \qquad \tilde{\mu}(x) = \frac{\|Ky\|}{\|x\|}$$

```
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);
```

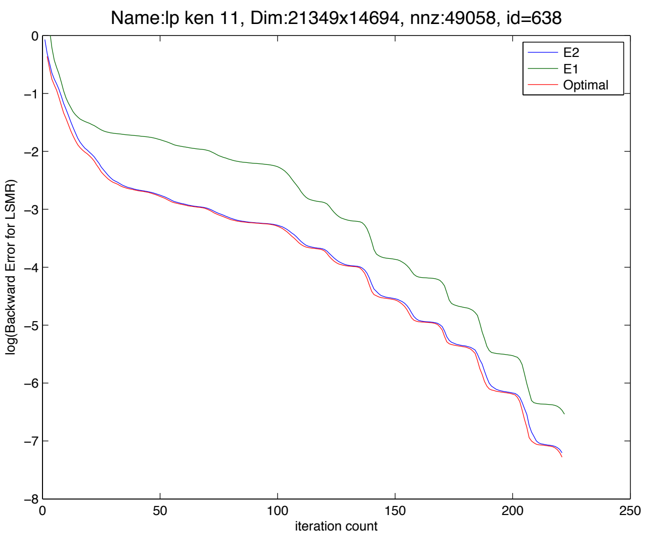
Backward errors for LSQR – typical



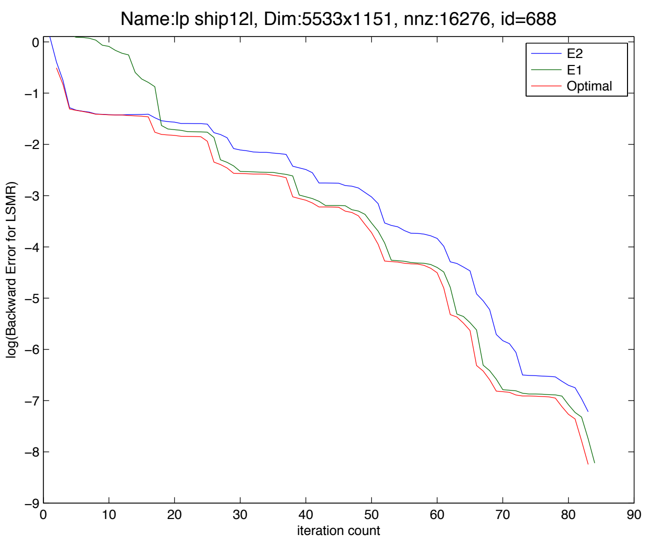
Backward errors for LSQR – rare



Backward errors for LSMR – typical



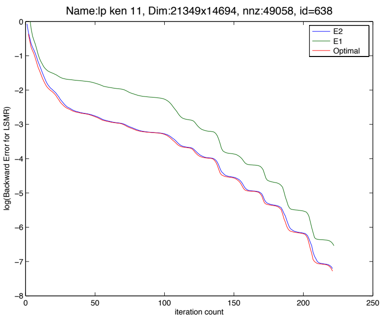
Backward errors for LSMR – rare



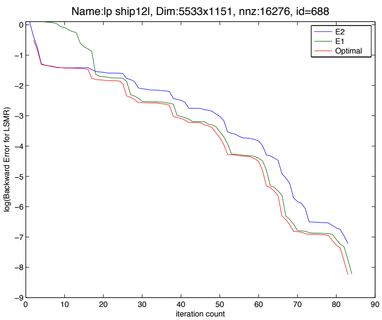
For LSMR

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

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



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

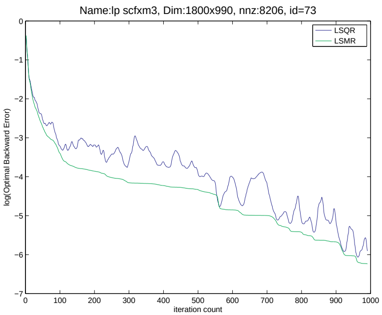


Optimal backward errors $\tilde{\mu}(x)$

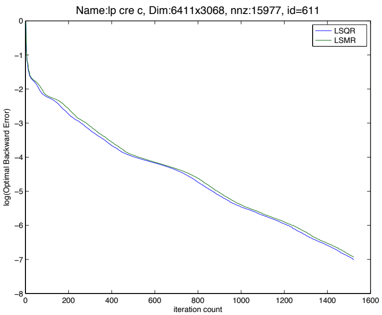
Seem monotonic for LSMR

Usually not for LSQR

Typical for LSQR and LSMR



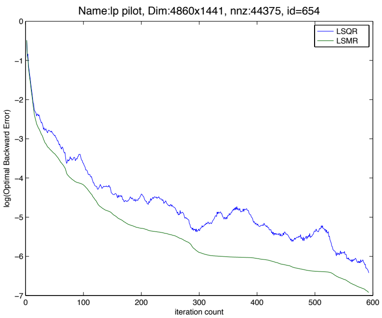
Rare LSQR, typical LSMR



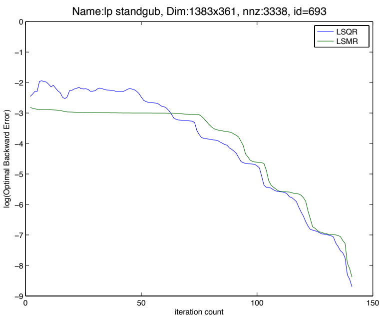
Optimal backward errors

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

Typical



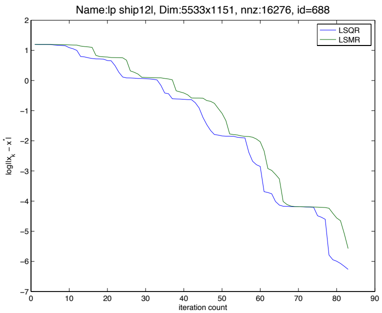
Rare

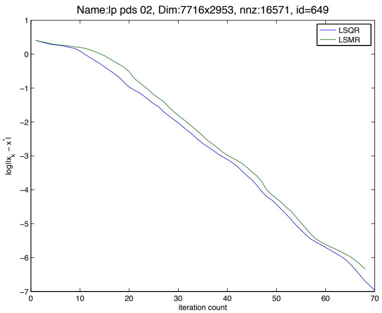


Errors in x_k

- $\|x^{\text{LSQR}} - x^*\| \leq \|x^{\text{LSMR}} - x^*\|$ seems true

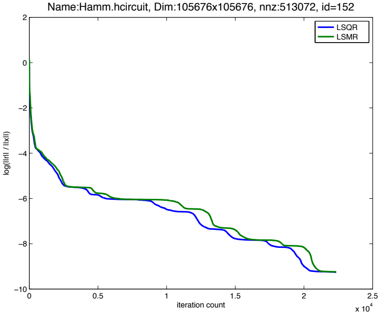
$\|x_k - x^*\|$ for LSMR and LSQR

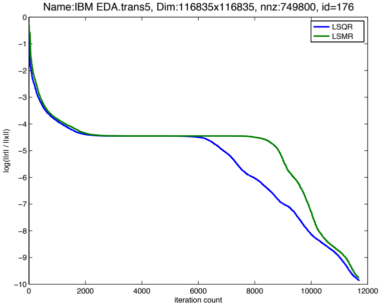




Square consistent systems

- $Ax = b$
- Backward error: $\frac{\|r_k\|}{\|x_k\|}$
- LSQR slightly faster than LSMR in most cases



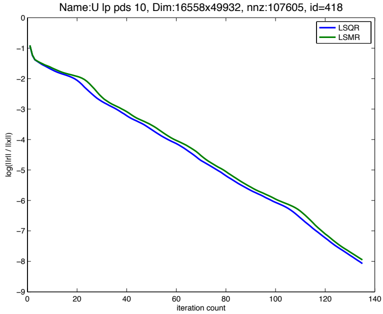


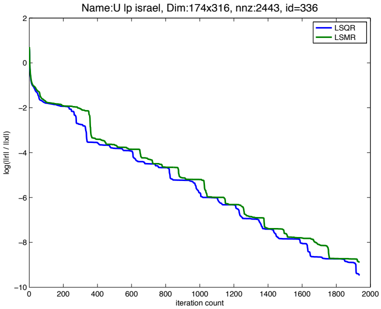
Underdetermined systems

Infinitely many solutions	Unique solution
$Ax = b$	$\min \|\|x\|\| \text{ st } Ax = b$

Theorem

LSQR and LSMR both return the minimum-norm solution





Summary

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

CG and MINRES

$$\begin{array}{ll} \|x^* - x_k\| & \searrow \\ \|x^* - x_k\|_A & \searrow \\ \|x_k\| & \nearrow \end{array}$$

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

CG and MINRES

$$\begin{array}{ll} \|x^* - x_k\| & \searrow \\ \|x^* - x_k\|_A & \searrow \\ \|x_k\| & \nearrow \end{array}$$

MINRES

$$\begin{array}{ll} \|r_k\| & \searrow \\ \|r_k\|/\|x_k\| & \searrow \\ \|r_k\|/(\alpha\|A\|\|x_k\| + \beta\|b\|) & \searrow \end{array}$$

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

CG and MINRES

$$\begin{array}{ll} \|x^* - x_k\| & \searrow \\ \|x^* - x_k\|_A & \searrow \\ \|x_k\| & \nearrow \end{array}$$

MINRES

$$\begin{array}{ll} \|r_k\| & \searrow \\ \|r_k\|/\|x_k\| & \searrow \\ \|r_k\|/(\alpha\|A\|\|x_k\| + \beta\|b\|) & \searrow \end{array}$$

For MINRES, backward errors are monotonic

$\Rightarrow$ safe to stop early

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

LSQR and LSMR

$\|x^* - x_k\| \searrow$

$\|r^* - r_k\| \searrow$

$\|r_k\| \searrow$

$\|x_k\| \nearrow$

$x_k \rightarrow$ min-length x^* if $\mathrm{rank}(A) < n$

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

LSQR and LSMR

$$\|x^* - x_k\| \searrow$$
$$\|r^* - r_k\| \searrow$$
$$\|r_k\| \searrow$$
$$\|x_k\| \nearrow$$

$x_k \rightarrow$ min-length x^* if $\mathrm{rank}(A) < n$

LSMR

$$\|A^T r_k\| \searrow$$

$\|A^T r_k\| / \|r_k\| \searrow$ almost always

$\approx$ optimal BE almost always

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

LSQR and LSMR

$$\|x^* - x_k\| \searrow$$
$$\|r^* - r_k\| \searrow$$
$$\|r_k\| \searrow$$
$$\|x_k\| \nearrow$$

$x_k \rightarrow$ min-length x^* if $\text{rank}(A) < n$

LSMR

$$\|A^T r_k\| \searrow$$
$\|A^T r_k\| / \|r_k\| \searrow$ almost always

$\approx$ optimal BE almost always

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

$\Rightarrow$ safe to stop early

References:

- LSMR: An iterative algorithm for sparse least-squares problems
  David Fong and Michael Saunders, SISC 2011
- CG versus MINRES: An empirical comparison
  David Fong and Michael Saunders, SQU Journal for Science 2012

Kindest thanks:

Georg Bock and colleagues
Phan Thanh An and colleagues