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
1 Meeting for Philippe

2 CG and MINRES
   The Lanczos Process
   Properties
   Backward Errors

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

4 Posdef systems and Least squares

5 LSQR and LSMR
   Backward Errors

6 Summary
Meeting for Philippe

First thought:
The trust-region subproblem

$$\|x_k\| \uparrow$$ for CG
The trust-region subproblem

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

Apply PCG to \( Hp = -g \)
Exit if \( d^T Hd < 0 \)
\[\text{or} \quad \|p\|_M > \Delta\]
The trust-region subproblem

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

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

Focus on \[Ax = b, \quad A \succ 0, \quad M = I\]
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)

\[ \beta_1 v_1 = b \quad AV_k = V_{k+1} H_k \]

\[ V_k = \begin{pmatrix} v_1 & v_2 & \ldots & v_k \end{pmatrix} \]

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

\[ H_k = \begin{pmatrix} T_k \\ 0 & \ldots & 0 & x \end{pmatrix} \]
Lanczos process (summary)

\[ \beta_1 v_1 = b \quad AV_k = V_{k+1} H_k \]

\[ V_k = \begin{pmatrix} v_1 & v_2 & \ldots & v_k \end{pmatrix} \]

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

\[ H_k = \begin{pmatrix} T_k \\ 0 & \ldots & 0 & x \end{pmatrix} \]

\[ r_k = b - Ax_k \]
\[ = \beta_1 v_1 - AV_k y_k \]
\[ = V_{k+1} (\beta_1 e_1 - H_k y_k), \]

Aim: \[ \beta_1 e_1 \approx H_k y_k \]
Lanczos process (summary)

\[ \beta_1 v_1 = b \quad AV_k = V_{k+1}H_k \]

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

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

\[ H_k = \begin{pmatrix} T_k & \vdots \\ \vdots & \ddots \\ \vdots & & \ddots & \beta_k \\ 0 & \cdots & 0 & x \end{pmatrix} \]

\[ r_k = b - Ax_k \]

\[ = \beta_1 v_1 - AV_k y_k \]

\[ = V_{k+1}(\beta_1 e_1 - H_k y_k), \]

Aim: \[ \beta_1 e_1 \approx H_k y_k \]

Two subproblems

CG \[ T_k y_k = \beta_1 e_1 \quad x_k = V_k y_k \]

MINRES \[ \min \| H_k y_k - \beta_1 e_1 \| \quad x_k = V_k y_k \]
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

• \( \text{CR} \equiv \text{MINRES} \) when \( A \succ 0 \)
  They both minimize \( \| r_k \| = \| b - Ax_k \| \) in the Krylov subspace
Theoretical properties for $Ax = b$, $A \succ 0$

<table>
<thead>
<tr>
<th></th>
<th>CG</th>
<th>CR (MINRES)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$|x^* - x_k|$</td>
<td>HS 1952</td>
<td>HS 1952</td>
</tr>
<tr>
<td>$|x^* - x_k|_A$</td>
<td>HS 1952</td>
<td>HS 1952</td>
</tr>
<tr>
<td>$|x_k|$</td>
<td>Steihaug 1983</td>
<td>Fong 2011</td>
</tr>
</tbody>
</table>
Theoretical properties for $Ax = b$, $A \succ 0$

<table>
<thead>
<tr>
<th></th>
<th>CG</th>
<th>CR (MINRES)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$|x^* - x_k|$</td>
<td>HS 1952</td>
<td>HS 1952</td>
</tr>
<tr>
<td>$|x^* - x_k|_A$</td>
<td>HS 1952</td>
<td>HS 1952</td>
</tr>
<tr>
<td>$|x_k|$</td>
<td>Steihaug 1983</td>
<td>Fong 2011</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>CR (MINRES)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$|r_k|$</td>
<td>HS 1952</td>
</tr>
<tr>
<td>$|r_k|/|x_k|$</td>
<td>Fong 2011</td>
</tr>
</tbody>
</table>
Backward error for square systems $Ax = b$

Stopping tolerances $\alpha$, $\beta$

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

$$(A + E)x_k = b + f$$

$$\frac{\|E\|}{\|A\|} \leq \alpha$$

$$\frac{\|f\|}{\|b\|} \leq \beta$$
Backward error for square systems $Ax = b$

Stopping tolerances $\alpha$, $\beta$

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

$$(A + E)x_k = b + f$$

$$\frac{\|E\|}{\|A\|} \leq \alpha \quad \frac{\|f\|}{\|b\|} \leq \beta$$

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

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

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

$$\frac{\|E\|}{\|A\|} = \alpha \frac{\|r_k\|}{\psi}$$

$$\frac{\|f\|}{\|b\|} = \beta \frac{r_k}{\psi}$$
Backward error for square systems $Ax = b$

Stopping tolerances $\alpha$, $\beta$

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

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

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

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

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

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

Stopping rule: $\|r_k\| \leq \psi \equiv \alpha \|A\| \|x_k\| + \beta \|b\|$
Backward error for square systems, $\beta = 0$

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

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

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

Plot $\log_{10} \|E_k\|$ for CG and MINRES
\[ \| r_k \| / \| x_k \| \text{ for } A \succ 0 \] MINRES can stop sooner

Name:Schenk_AFE_af_shell8, Dim:504855x504855, nnz:17579155, id=11

Name:Cannizzo_sts4098, Dim:4098x4098, nnz:72356, id=13

Name:Simon_raefsky4, Dim:19779x19779, nnz:1316789, id=7

Name:BenElechi_BenElechi1, Dim:245874x245874, nnz:13150496, id=22
\[ \frac{\| r_k \|}{\| x_k \|} \text{ and } \log_{10} \| x - x_k \|_A \]
\[ \|r_k\| / \|x_k\| \text{ and } \log_{10} \|x - x_k\| \]

**Name:** Schenk_AFE_af_shell8, Dim: 504855x504855, nnz: 17579155, id=11

**Name:** Cannizzo_sts4098, Dim: 4098x4098, nnz: 72356, id=13
\[ \frac{\|r_k\|}{\|x_k\|} \text{ and } \log_{10} \|x - x_k\|_A \]
\[ \frac{\|r_k\|}{\|x_k\|} \text{ and } \log_{10} \|x - x_k\| \]
\[ \| r_k \| \quad \| x_k \| \quad \| x - x_k \|_A \quad \| x - x_k \| \quad \kappa(A) = 10^6 \]
\[ \| r_k \| \quad \| x_k \| \quad \| x - x_k \|_A \quad \| x - x_k \| \quad \kappa(A) = 10^{11} \]
\[ \| r_k \| \quad \| x_k \| \quad \| x - x_k \|_A \quad \| x - x_k \| \quad \kappa(A) = 10^7 \]

![Graphs showing convergence of iterative methods](image)
$\|r_k\| \quad \|x_k\| \quad \|x - x_k\|_A \quad \|x - x_k\| \quad \kappa(A) = 10^{12}$
Why does $\| r_k \|$ for CG lag behind MINRES?
Why does CG lag behind MINRES?

Greenbaum 1997: (Thanks David Titley-Peloquin)

\[
\| r^C_k \| = \frac{\| r^M_k \|}{\sqrt{1 - \| r^M_k \|^2 / \| r^M_{k-1} \|^2}}
\]

\[\Rightarrow \| r^C_k \| \gg \| r^M_k \| \text{ if MINRES is almost stalling}\]
Why does CG lag behind MINRES?

Greenbaum 1997: (Thanks David Titley-Peloquin)

\[ \|r^C_k\| = \frac{\|r^M_k\|}{\sqrt{1 - \|r^M_k\|^2 / \|r^M_{k-1}\|^2}} \]

\[ \Rightarrow \|r^C_k\| \gg \|r^M_k\| \text{ if MINRES is almost stalling} \]

Fong 2011: If \( \alpha = 0 \) in stopping rule, stop when \( \|r_l\| \leq \beta \|b\| \)

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

\[ \Rightarrow \|r^M_k\| / \|r^M_{k-1}\| \text{ closer to 1 on average if } l \text{ is large} \]

\[ \Rightarrow \text{ bigger gap on average between } \|r^C_k\| \text{ and } \|r^M_k\| \]
Posdef systems
and least squares
LS $\equiv$ PSD system

\[ \min \|Ax - b\| \implies A^T Ax = A^T b \]
LS ≡ PSD system

\[ \min \| Ax - b \| \quad \Rightarrow \quad A^T Ax = A^T b \]

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

PSD system ≡ LS

\[ Ax = b \quad \Rightarrow \quad U^T U x = U^T c \quad \Rightarrow \quad \min \| U x - c \| \]
Part II: LSQR and LSMR

\[
\text{LSQR} \equiv \text{CG on } A^T A x = A^T b
\]
\[
\text{LSMR} \equiv \text{MINRES on } A^T A x = A^T b
\]

Based on Golub-Kahan bidiagonalization
Which problems do LSQR and LSMR solve?

solve $Ax = b$
Which problems do LSQR and LSMR solve?

\[
\text{solve } Ax = b \quad \min \|x\| \quad \text{st } Ax = b
\]
Which problems do LSQR and LSMR solve?

\[
\text{solve } Ax = b \quad \min \| x \| \quad \text{st } Ax = b
\]

\[
\min \| Ax - b \|
\]
Which problems do LSQR and LSMR solve?

\[
\begin{align*}
solve \ Ax &= b \\
\min \ ||x|| & \text{ st } Ax = b \\
\min \ ||Ax - b|| & \text{ st } Ax = b \\
\min \ ||(A \lambda I)x - (b 0)|| &
\end{align*}
\]
Which problems do LSQR and LSMR solve?

\[
\text{solve } Ax = b \quad \min \|x\| \quad \text{st } Ax = b
\]

\[
\min \|Ax - b\| \quad \min \left\| \begin{pmatrix} A \\ \lambda I \end{pmatrix} x - \begin{pmatrix} b \\ 0 \end{pmatrix} \right\|
\]

- \( A \) square or rectangular \((m \times n)\) and often sparse
- \( A \) can be an operator \((\Rightarrow \text{ allows preconditioning})\)
- \( Av, A^Tu \) plus \( O(m + n) \) operations per iteration
Which problems do LSQR and LSMR solve?

\[
\text{solve } Ax = b \quad \text{min } \|x\| \quad \text{st } Ax = b
\]

\[
\text{min } \|Ax - b\| \quad \text{min } \left\| \begin{pmatrix} A \\ \lambda I \end{pmatrix} x - \begin{pmatrix} b \\ 0 \end{pmatrix} \right\|
\]

- \(A\) square or rectangular \((m \times n)\) and often sparse
- \(A\) can be an operator \((\Rightarrow \text{ allows preconditioning})\)
- \(Av, A^Tu\) 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 $Ax = b$

- $\|r_k\| \downarrow 0$ by design
- $\|x_k\| \uparrow$ PS 1982
- Should have been a clue that $\|x_k\| \uparrow$ 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-Peloquinn 2010
LSQR on $Ax = b$

- $\| r_k \| \downarrow 0$ by design
- $\| x_k \| \uparrow$ PS 1982
- Should have been a clue that $\| x_k \| \uparrow$ 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 $A \succ 0$):

$$\frac{\| r_k \|}{(\alpha \| A \| \| x_k \| + \beta \| b \|)} \downarrow$$
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 $A \succ 0$):

  $$\frac{\|r_k\|}{(\alpha \|A\| \|x_k\| + \beta \|b\|)} \searrow$$

Everything the same for LSMR!
LSQR on $Ax = b$

- $\|r_k\| \downarrow 0$ by design
- $\|x_k\| \uparrow$ PS 1982
- Should have been a clue that $\|x_k\| \uparrow$ 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 $A \succ 0$):

\[
\frac{\|r_k\|}{\alpha \|A\| \|x_k\| + \beta \|b\|} \downarrow
\]

Everything the same for LSMR!

What about GMRES?
LSQR and LSMR on $\min \|Ax - b\|

Stewart backward error

$r_k = b - Ax_k$

\[
\frac{\|A^T r_k\|}{\|A\|\|r_k\|} \leq \alpha
\]
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 and LSMR on \( \min \|Ax - b\| \)

Stewart backward error

\[
r_k = b - Ax_k
\]

\[
\frac{\|A^T r_k\|}{\|A\| \|r_k\|} \leq \alpha
\]

---

**\( \|r_k\| \)**

Name:lp fit1p, Dim:1677x627, nnz:9868, id=625

**\( \log \|A^T r_k\| \)**

Name:lp fit1p, Dim:1677x627, nnz:9868, id=625
Overdetermined systems

Test Data

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

Test Data

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

\[
\text{Solve } \min \|Ax - b\|_2
\]

with LSQR and LSMR
Backward error – estimates

\[ A^T A \hat{x} = A^T b \quad \hat{r} = b - A \hat{x} \quad \text{exact} \]

\[(A + E_i)^T (A + E_i) x = (A + E_i)^T b \quad r = b - Ax \quad \text{any } x\]
Backward error – estimates

\[ A^T \hat{x} = A^T b \]
\[ \hat{r} = b - A\hat{x} \quad \text{exact} \]

\[ (A + E_i)^T (A + E_i)x = (A + E_i)^T b \]
\[ r = b - Ax \quad \text{any } x \]

Two estimates given by Stewart (1975 and 1977)

\[ E_1 = \frac{ex^T}{\|x\|^2} \]
\[ \|E_1\| = \frac{\|e\|}{\|x\|} \quad e = \hat{r} - r \]

\[ E_2 = -\frac{rr^T A}{\|r\|^2} \]
\[ \|E_2\| = \frac{\|A^T r\|}{\|r\|} \quad \text{computable} \]
Backward error – estimates

\[ A^T \hat{x} = A^T b \quad \hat{r} = b - A\hat{x} \quad \text{exact} \]

\[ (A + E_i)^T (A + E_i)x = (A + E_i)^T b \quad r = b - Ax \quad \text{any } x \]

Two estimates given by Stewart (1975 and 1977)

\[ E_1 = \frac{ex^T}{\|x\|^2} \quad \|E_1\| = \frac{\|e\|}{\|x\|} \quad e = \hat{r} - r \]

\[ E_2 = -\frac{rr^TA}{\|r\|^2} \quad \|E_2\| = \frac{\|A^r\|}{\|r\|} \quad \text{computable} \]

**Theorem**

\[ \|E_2^{\text{LSMR}}\| \leq \|E_2^{\text{LSQR}}\| \]
\log_{10} \|E_2\| \text{ for LSQR and LSMR – typical}

Name: lp pilot ja, Dim: 2267x940, nnz: 14977, id=657

Iteration count vs. \log(E_2)

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

Name: lp sc205, Dim: 317x205, nnz: 665, id=665

Graph showing the comparison between LSQR and LSMR methods for a specific problem with the iteration count on the x-axis and $\log(E_2)$ on the y-axis.
Backward error - optimal

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

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

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

\[ K = \begin{pmatrix} A \\ \|r\| I \end{pmatrix} \quad v = \begin{pmatrix} r \\ 0 \end{pmatrix} \]

\[ \min_{y} \|K y - v\| \quad \tilde{\mu}(x) = \frac{\|K y\|}{\|x\|} \]
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 \( \tilde{\mu}(x) \) \hspace{1cm} (Grcar, Saunders, & Su 2007)

\[
K = \begin{pmatrix}
    A \\
    \| r \| I \\
    \| x \|
\end{pmatrix}
\quad
v = \begin{pmatrix}
    r \\
    0
\end{pmatrix}
\]

\[ \min_{y} \| Ky - v \| \quad \tilde{\mu}(x) = \frac{\| Ky \|}{\| x \|} \]

\[
r = b - A^*x; \\
p = \text{colamd}(A); \\
eta = \text{norm}(r) / \text{norm}(x); \\
K = [A(:,p); \text{eta} \cdot \text{speye}(n)]; \\
v = [r; \text{zeros}(n,1)]; \\
[c,R] = \text{qr}(K,v,0); \\
\text{mutilde} = \text{norm}(c) / \text{norm}(x); \]
Backward errors for LSQR – typical

Name: lp cre a, Dim: 7248x3516, nnz: 18168, id=609

log(Backward Error for LSQR) vs iteration count

- Blue line: E2
- Green line: E1
- Red line: Optimal
Backward errors for LSQR – rare

Name:lp pilot, Dim:4860x1441, nnz:44375, id=654
Backward errors for LSMR – typical
Backward errors for LSMR – rare

Name: lp ship12l, Dim: 5533x1151, nnz: 16276, id=688

- E2
- E1
- Optimal

log(Backward Error for LSMR) vs iteration count
For LSMR

$\| E_2 \| \approx \text{optimal BE almost always}$

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

\[ \tilde{\mu}(x^{\text{LSMR}}) \leq \tilde{\mu}(x^{\text{LSQR}}) \text{ almost always} \]
Summary
Theoretical properties for $Ax = b$

CG and MINRES, $A \succ 0$

$$\|x^* - x_k\|, \|x^* - x_k\|_A \downarrow$$

$$\|x_k\| \uparrow$$
Theoretical properties for $Ax = b$

**CG and MINRES, $A \succ 0$**

<table>
<thead>
<tr>
<th>Expression</th>
<th>Behavior</th>
</tr>
</thead>
<tbody>
<tr>
<td>$|x^* - x_k|$, $|x^* - x_k|_A$</td>
<td>$\downarrow$</td>
</tr>
<tr>
<td>$|x_k|$</td>
<td>$\uparrow$</td>
</tr>
</tbody>
</table>

**MINRES, $A \succ 0$**

<table>
<thead>
<tr>
<th>Expression</th>
<th>Behavior</th>
</tr>
</thead>
<tbody>
<tr>
<td>$|r_k|$</td>
<td>$\downarrow$</td>
</tr>
<tr>
<td>$|r_k|/ (\alpha |A| |x_k| + \beta |b|)$</td>
<td>$\downarrow$</td>
</tr>
</tbody>
</table>
Theoretical properties for $Ax = b$

<table>
<thead>
<tr>
<th>Method</th>
<th>Property</th>
<th>Condition</th>
<th>Notation</th>
</tr>
</thead>
<tbody>
<tr>
<td>CG and MINRES, $A \succ 0$</td>
<td>$|x^* - x_k|$, $|x^* - x_k|_A$</td>
<td>$\downarrow$</td>
<td>$|x_k|$ $\uparrow$</td>
</tr>
<tr>
<td>MINRES, $A \succ 0$</td>
<td>$|r_k|$</td>
<td>$\downarrow$</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$|r_k|/(\alpha |A| |x_k| + \beta |b|)$</td>
<td>$\downarrow$</td>
<td></td>
</tr>
<tr>
<td>LSQR and LSMR, any $A$</td>
<td>$|r_k|$</td>
<td>$\downarrow$</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$|r_k|/(\alpha |A| |x_k| + \beta |b|)$</td>
<td>$\downarrow$</td>
<td></td>
</tr>
</tbody>
</table>

Monotonic backward errors $\Rightarrow$ safe to stop early
Theoretical properties for \( Ax = b \)

**CG and MINRES, \( A \succ 0 \)**

\[
\| x^* - x_k \|, \quad \| x^* - x_k \|_A
\]

\[
\| x_k \| \quad \uparrow
\]

**MINRES, \( A \succ 0 \)**

\[
\| r_k \|
\]

\[
\| r_k \| / (\alpha \| A \| \| x_k \| + \beta \| b \|)
\]

\[
\downarrow
\]

**LSQR and LSMR, any \( A \)**

\[
\| r_k \|
\]

\[
\| r_k \| / (\alpha \| A \| \| x_k \| + \beta \| b \|)
\]

\[
\downarrow
\]

Monotonic backward errors

\( \Rightarrow \) safe to stop early
Theoretical properties for $\min \|Ax - b\|$ 

**LSQR and LSMR**

\[
\|x^* - x_k\|, \|r^* - r_k\| \quad \downarrow \\
\|r_k\| \quad \downarrow \\
\|x_k\| \quad \uparrow \\
\]

$x_k \to \text{min-length } x^*$ if $\text{rank}(A) < n$
Theoretical properties for $\min ||Ax - b||$

**LSQR and LSMR**

\[
\begin{align*}
||x^* - x_k||, \quad ||r^* - r_k|| & \searrow \\
||r_k|| & \searrow \\
||x_k|| & \nearrow \\
\end{align*}
\]

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

**LSMR**

\[
\begin{align*}
||A^T r_k|| & \searrow \\
||A^T r_k|| / ||r_k|| & \searrow \text{almost always} \\
& \approx \text{optimal BE almost always} \\
\leq & (||A^T r_k|| / ||r_k||)^{\text{LSQR}} \\
\end{align*}
\]
Theoretical properties for \( \min \|Ax - b\| \)

**LSQR and LSMR**

\[
\begin{align*}
\|x^* - x_k\|, \quad \|r^* - r_k\| & \searrow \\
\|r_k\| & \searrow \\
\|x_k\| & \nearrow \\
x_k \rightarrow \text{min-length } x^* \quad \text{if } \text{rank}(A) < n
\end{align*}
\]

**LSMR**

\[
\begin{align*}
\|A^T r_k\| & \searrow \\
\|A^T r_k\| / \|r_k\| & \searrow \text{almost always} \\
\approx \text{optimal BE almost always} \\
\leq (\|A^T r_k\| / \|r_k\|)^{\text{LSQR}}
\end{align*}
\]

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)