Error Bounds for CG via SYMMLQ

Michael Saunders

MS&E and ICME, Stanford University

Joint work with Ron Estrin and Dominique Orban

Householder Symposium XX

Blacksburg, VA, June 18–23, 2017

Abstract

For SPD $Ax_* = b$, we show that the error $\|x_* - x_k^L\|_2$ can be bounded above for SYMMLQ iterates, and this leads to an upper bound on the CG error $\|x_* - x_k^C\|_2$ (assuming exact arithmetic). We follow the Gauss-Radau approach of Golub and Meurant (1997), who bound $\|x_* - x_k^C\|_A$ for CG and estimate $\|x_* - x_k^C\|_2$. For indefinite A, our SYMMLQ bound $\|x_* - x_k^L\|_2$ also becomes an estimate.

In practice we find that the bounds and estimates are remarkably tight. They suggest a cheaply implementable stopping criterion. We mention analogous error bounds for LSQR via LSLQ (the missing sister of LSQR and LSMR for $\min \|Ax - b\|_2$).

Partially supported by the

National Institute of General Medical Sciences of the National Institutes of Health (NIH)

Award U01GM102098

Lanczos for symmetric $Ax = b$

Error bounds for SYMMLQ and hence CG

Assume exact arithmetic

Check experimentally

Previous work

Error estimates for CG

Golub and Strakoš (1994)
Golub and Meurant (MMQ 1994, 1997)
Meurant (1997, 2005)
Brezinski (1999)
Frommer, Kahl, Lippert, and Rittich (2013)

Finite-precision analyses

- Strakoš and Tichý (2002)
  On error estimation in the CG method and why it works in finite precision computations
  ETNA 13
- Meurant (2006), The Lanczos and CG Algorithms: From Theory to Finite Precision Computations
  SIAM
- Greif, Paige, Titley-Peloquin, and Varah (2016)
  Numerical equivalences among Krylov subspace algorithms for skew-symmetric matrices
  SIMAX 37
- Paige (2017), Accuracy of the Lanczos process for the eigenproblem and solution of equations
  SIMAX soon (hot off the press!)

The Lanczos process for A, b

For $k = 1, 2, \ldots, \ell$
Lanczos generates $V_k = \begin{bmatrix} v_1 & v_2 & \ldots & v_k \end{bmatrix}$ and $\{\alpha_k, \beta_k > 0\}$ such that

$$\beta_1 v_1 = b \qquad AV_k = V_{k+1}\underline{T}_k \qquad \|v_k\| = 1 \qquad \beta_{\ell+1} = 0$$

$$\underline{T}_k = \begin{bmatrix} \alpha_1 & \beta_2 & & \\ \beta_2 & \alpha_2 & \ddots & \\ & \ddots & \ddots & \beta_k \\ & & \beta_k & \alpha_k \\ \hline & & & \beta_{k+1} \end{bmatrix} = \begin{bmatrix} T_k \\ \cdots \;\; \times \end{bmatrix}$$

SYMMLQ, CG, MINRES for $Ax = b$

- $x_k = V_k y_k$
- $r_k = b - Ax_k = V_{k+1}(\beta_1 e_1 - \underline{T}_k y_k)$ — r_k small if $\underline{T}_k y_k \approx \beta_1 e_1$
- 3 ways to make r_k small — 3 subproblems for choosing y_k

$$\begin{bmatrix} \alpha_1 & \beta_2 & & \\ \beta_2 & \alpha_2 & \ddots & \\ & \ddots & \ddots & \beta_k \\ \hline & & \beta_k & \alpha_k \\ \hline & & & \beta_{k+1} \end{bmatrix} y_k \approx \begin{bmatrix} \beta_1 \\ 0 \\ \vdots \\ \hline 0 \\ \hline 0 \end{bmatrix}$$

SYMMLQ: $\min \|y_k\| \;\text{ st }\; \underline{T}_{k-1}^T y_k = \beta_1 e_1$

CG: $T_k y_k = \beta_1 e_1$

MINRES: $\underline{T}_k y_k \approx \beta_1 e_1$

SYMMLQ

$$\min \|y_k\| \ \textbf{st} \ \underline{T}_{k-1}^T y_k = \beta_1 e_1 \qquad \textbf{(then } x_k^L = V_k y_k\textbf{)}$$

$$\textbf{Needs } \underline{T}_{k-1}^T = \begin{bmatrix} L_{k-1} & 0 \end{bmatrix} Q_k$$

$$x_k^L = W_{k-1} z_{k-1} = x_{k-1}^L + \zeta_{k-1} w_{k-1}$$

moves in theoretically orthogonal directions

SYMMLQ recursions

$$\underline{T}_{k-1}^T Q_k^T = \begin{bmatrix} L_{k-1} & 0 \end{bmatrix} \qquad L_{k-1} z_{k-1} = \beta_1 e_1$$

$$T_k Q_k^T = \bar{L}_k = \begin{bmatrix} & L_{k-1} & & \\ 0 & \epsilon_k & \delta_k & \bar{\gamma}_k \end{bmatrix} \qquad \bar{L}_k \bar{z}_k = \beta_1 e_1$$

$$\bar{W}_k = V_k Q_k^T = \begin{bmatrix} W_{k-1} & \bar{w}_k \end{bmatrix} \qquad \bar{z}_k = \begin{bmatrix} z_{k-1} \\ \bar{\zeta}_k \end{bmatrix}$$

SYMMLQ recursions

$$\underline{T}_{k-1}^T Q_k^T = \begin{bmatrix} L_{k-1} & 0 \end{bmatrix} \qquad L_{k-1} z_{k-1} = \beta_1 e_1$$

$$T_k Q_k^T = \bar{L}_k = \begin{bmatrix} & L_{k-1} & & \\ 0 & \epsilon_k & \delta_k & \bar{\gamma}_k \end{bmatrix} \qquad \bar{L}_k \bar{z}_k = \beta_1 e_1$$

$$\bar{W}_k = V_k Q_k^T = \begin{bmatrix} W_{k-1} & \bar{w}_k \end{bmatrix} \qquad \bar{z}_k = \begin{bmatrix} z_{k-1} \\ \bar{\zeta}_k \end{bmatrix}$$

$$x_k^L = W_{k-1} z_{k-1} = x_{k-1}^L + \zeta_{k-1} w_{k-1}$$

$$x_k^C = \bar{W}_k \bar{z}_k \qquad = x_k^L + \bar{\zeta}_k \bar{w}_k$$

W_{k-1}, $\bar{W}_k$ theoretically have orthonormal columns

SYMMLQ error bound

$$x_k^L = W_{k-1} z_{k-1}, \qquad x_k^C = \bar{W}_k \bar{z}_k$$

W_{k-1}, $\bar{W}_k$ have theoretically orthonormal columns

$$\|x_k^L\|^2 = \|z_{k-1}\|^2 = \sum_1^{k-1} \zeta_j^2$$

$$\|x_*\|^2 = \|z_\ell\|^2 \quad = \sum_1^{\ell} \zeta_j^2$$

$$\|x_* - x_k^L\|^2 = \|x_*\|^2 - \|x_k^L\|^2$$

To bound the SYMMLQ error we need a bound on $\|x_*\|^2 = b^T A^{-2} b$

Bounding $\|x_*\|^2 = b^TA^{-2}b$

Needs Golub and Meurant

Golub and Meurant (1994, 1997)

Estimate bilinear forms $u^T f(A) v$ using Gaussian-quadrature theory

Theorem

For SPD A and suitable f, fix $\lambda_{\text{est}} \in (0, \lambda_{\min}(A))$ and choose ω_k such that

$\widetilde{T}_k = \begin{bmatrix} T_{k-1} & \beta_k e_{k-1} \\ \beta_k e_{k-1}^T & \omega_k \end{bmatrix}$, $\lambda_{\min}(\widetilde{T}_k) = \lambda_{\text{est}}$. Then $b^T f(A) b \leq \|b\|^2 e_1^T f(\widetilde{T}_k) e_1$.

$$f(\xi) = \xi^{-2} \text{ gives } \|x_*\|^2 = b^T A^{-2} b \leq \|b\|^2 e_1^T \widetilde{T}_k^{-2} e_1$$

Golub and Meurant (1994, 1997)

Estimate bilinear forms $u^T f(A) v$ using Gaussian-quadrature theory

Theorem

For SPD A and suitable f, fix $\lambda_{\text{est}} \in (0, \lambda_{\min}(A))$ and choose ω_k such that
$\widetilde{T}_k = \begin{bmatrix} T_{k-1} & \beta_k e_{k-1} \\ \beta_k e_{k-1}^T & \omega_k \end{bmatrix}$, $\lambda_{\min}(\widetilde{T}_k) = \lambda_{\text{est}}$. Then $b^T f(A) b \leq \|b\|^2 e_1^T f(\widetilde{T}_k) e_1$.

$$f(\xi) = \xi^{-2} \text{ gives } \|x_*\|^2 = b^T A^{-2} b \leq \|b\|^2 e_1^T \widetilde{T}_k^{-2} e_1$$

Theorem

$\omega_k = \lambda_{\text{est}} + \eta$, where η is last entry solution of $(T_{k-1} - \lambda_{\text{est}} I) u_{k-1} = \beta_k^2 e_{k-1}$.

QR on $(T_{k-1} - \lambda_{\text{est}} I)$ gives η, ω_k LQ on $\widetilde{T}_k$ gives $\|b\|^2 e_1^T \widetilde{T}_k^{-2} e_1$

Computing $\beta_1^2 e_1^T \widetilde{T}_k^{-2} e_1$

$\widetilde{T}_k = \widetilde{L}_k \widetilde{Q}_k$ is almost the same as $T_k = \bar{L}_k Q_k$.

- Solve $\widetilde{L}_k \widetilde{z}_k = \beta_1 e_1$ to get $\widetilde{z}_k = \begin{bmatrix} z_{k-1} \\ \widetilde{\zeta}_k \end{bmatrix}$
- $\|x_*\|^2 \leq \beta_1^2 e_1^T \widetilde{T}_k^{-2} e_1 = \|\beta_1 \widetilde{L}_k^{-1} e_1\|^2 = \|\widetilde{z}_k\|^2$
- We already solve $L_{k-1} z_{k-1} = \beta_1 e_1$ and have $\|x_k^L\|^2 = \|z_{k-1}\|^2$

Hence

$$\begin{aligned} \|x_* - x_k^L\|^2 &= \|x_*\|^2 - \|x_k^L\|^2 \\ &\leq \|\widetilde{z}_k\|^2 - \|z_{k-1}\|^2 = \widetilde{\zeta}_k^2 \end{aligned}$$

and we can bound the SYMMLQ error in $O(1)$ work per iteration:

$$\|x_* - x_k^L\| \leq \epsilon_k^L \equiv |\widetilde{\zeta}_k|$$

CG error $\leq$ SYMMLQ error

Theorem (Estrin, Orban, and S. 2017)

For positive-semidefinite consistent $Ax = b$,

$$\|x_k^L\| \leq \|x_k^C\|$$
$$\|x_* - x_k^C\| \leq \|x_* - x_k^L\|$$

Immediate consequence: $\|x_* - x_k^C\| \leq \|x_* - x_k^L\| \leq \epsilon_k^L$

Better bound: $\|x_* - x_k^C\| \leq \epsilon_k^C := \sqrt{(\epsilon_k^L)^2 - \bar{\zeta}_k^2}$ $\quad (x_k^C = x_k^L + \bar{\zeta}_k \bar{w}_k)$

Golub and Strakoš (1994)

- Store $x_k^C, \ldots, x_{k+d}^C$ for moderate values of d (sliding window approach)
- Take $x_{k+d} \approx x_*$ for some purposes

Lemma (Estrin, Orban, and S. 2017)

$$\theta_k := x_*^T x_k^C - \|x_k^C\|^2 \geq 0$$

$$\|x_* - x_k^C\| \leq \sqrt{(\epsilon_k^C)^2 - 2\theta_k} \qquad \text{(not computable)}$$

$$\theta_k^{(d)} := (x_{k+d})^T x_k^C - \|x_k^C\|^2 \leq \theta_k$$

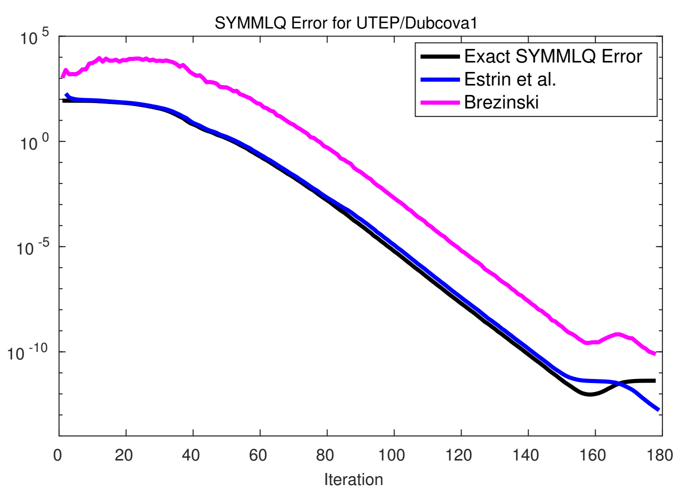
Better bound: $\|x_* - x_k^C\| \leq \sqrt{(\epsilon_k^C)^2 - 2\theta_k^{(d)}}$

Summary so far

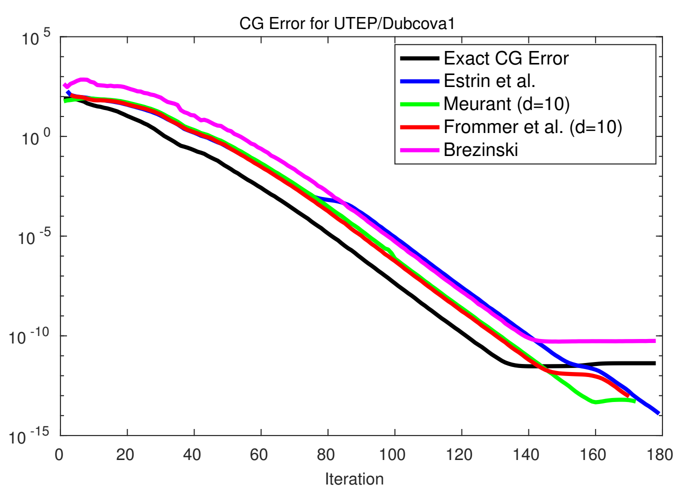
- For SPD $Ax = b$, we derived upper bounds on the SYMMLQ and CG errors, assuming exact arithmetic. The results hold if $Ax = b$ is semidefinite and consistent.
- Numerical experiments show the bounds hold until convergence, but rigorous finite-precision analysis is desirable.
- If A is indefinite, the SYMMLQ upper bound becomes an estimate. Could obtain a bound by treating $b^T A^{-2} b$ as a quadratic form in A^2, but this is expensive (2 applications of A per iteration).

Numerical results

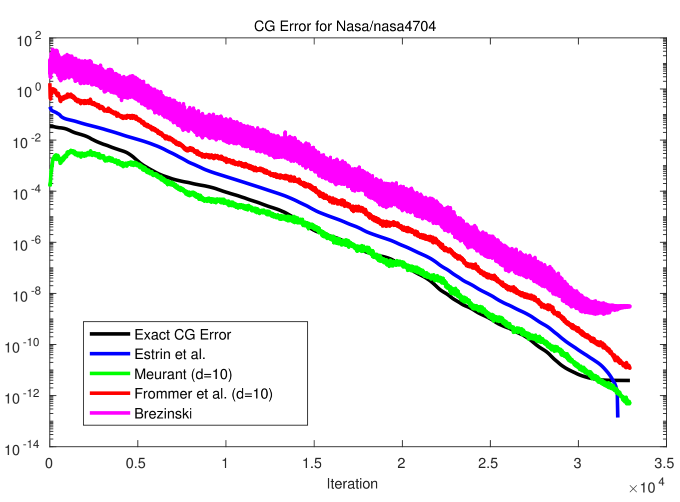
SYMMLQ error for UTEP/Dubcova1 $n = 16129$ SPD $\kappa(A) = 10^3$



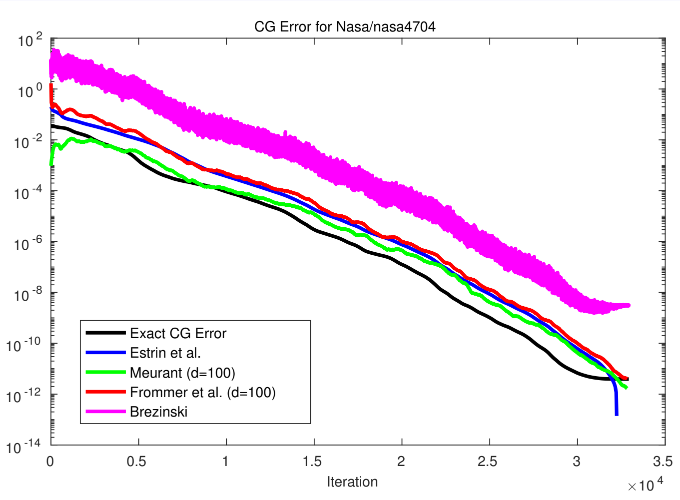
CG error for UTEP/Dubcova1 $n = 16129$ SPD $\kappa(A) = 10^3$



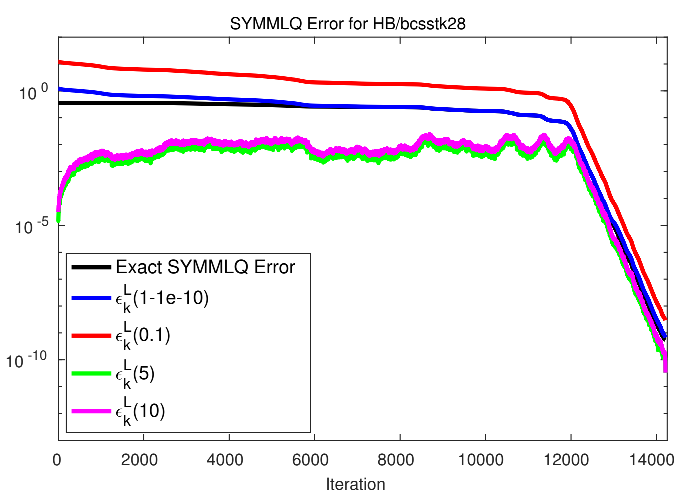
CG error for Nasa/nasa4704 $n = 4704$ SPD $\kappa(A) = 10^7$ $d = 10$



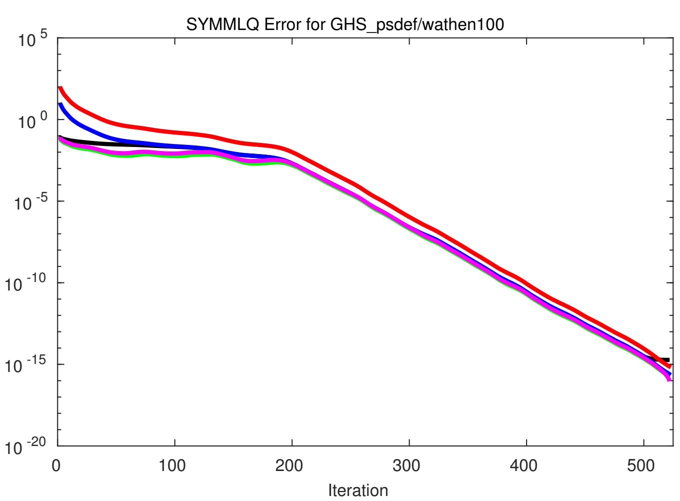
CG error for Nasa/nasa4704 $n = 4704$ SPD $\kappa(A) = 10^7$ $d = 100$



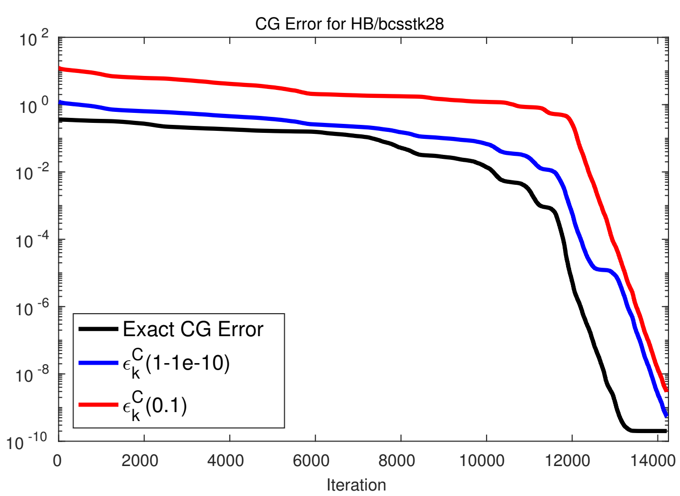
SYMMLQ error for HB/bcsstk28 $n = 4410$ SPD $\kappa(A) = 10^8$



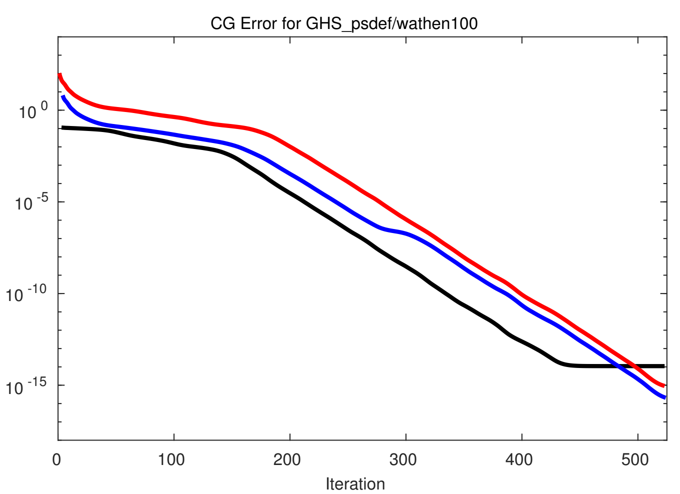
SYMMLQ error for GHS_psdef/wathen100 $n = 30401$ SPD $\kappa(A) = 10^3$



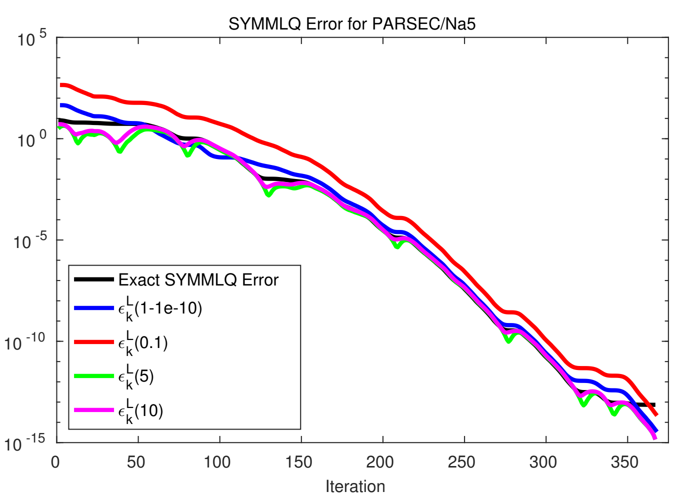
CG error for HB/bcsstk28 $n = 4410$ SPD $\kappa(A) = 10^8$



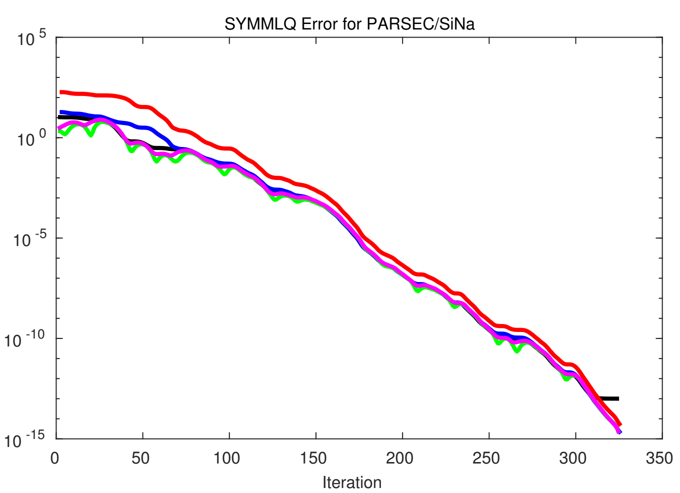
CG error for GHS_psdef/wathen100 $n = 30401$ SPD $\kappa(A) = 10^3$



SYMMLQ error for PARSEC/Na5 $n = 5822$ indef $\kappa(A) = 10^3$



SYMMLQ error for PARSEC/SiNa $n = 5743$ indef $\kappa(A) = 10^2$



Reminder: CG vs MINRES

on SPD $Ax = b$

CG vs MINRES

- D. Titley-Peloquin (2010), Backward Perturbation Analysis of Least Squares Problems, PhD thesis, McGill University

Backward errors for x_k

$$\min_{\xi, E, f} \xi \quad \text{st} \quad (A+E)x_k = b + f, \quad \frac{\|E\|}{\|A\|} \leq \alpha\xi, \quad \frac{\|f\|}{\|b\|} \leq \beta\xi$$

CG vs MINRES

- D. Titley-Peloquin (2010), Backward Perturbation Analysis of Least Squares Problems, PhD thesis, McGill University

Backward errors for x_k

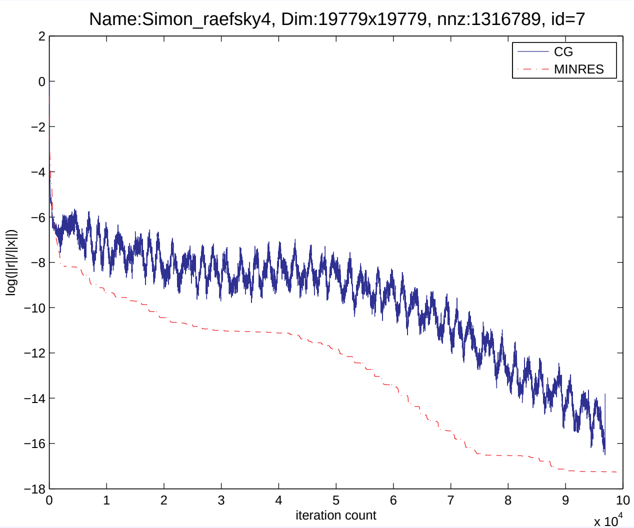
$$\min_{\xi,E,f} \xi \qquad \text{st} \qquad (A+E)x_k = b+f, \qquad \frac{\|E\|}{\|A\|} \leq \alpha\xi, \qquad \frac{\|f\|}{\|b\|} \leq \beta\xi$$

- D. C.-L. Fong and S. (2012), CG versus MINRES: An empirical comparison, SQU Journal for Science

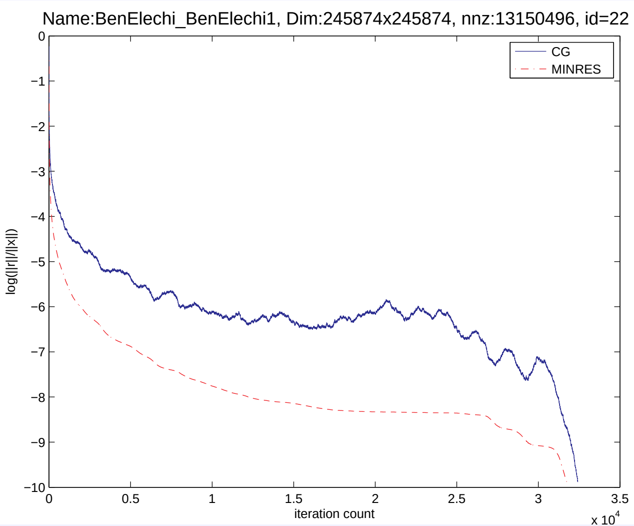
Theorem

MINRES backward errors $\|E_k\| \propto \|r_k\| \,/\, \|x_k\|$ and $\|f_k\| \propto \|r_k\|$ decrease monotonically

CG vs MINRES, $n = 19779$, backward errors $\|r_k\| \,/\, \|x_k\|$



CG vs MINRES, $n = 245874$, backward errors $\|r_k\| / \|x_k\|$



Conclusions

Conclusions

- Derived a **cheap** estimate of errors $\|x_* - x_k\|$ for SYMMLQ and CG.
- When A is SPD, the estimates are upper bounds (assuming exact arithmetic, but empirically in practice until convergence).
- Requires underestimate of smallest nonzero eigenvalue.
  - Common to Gauss-Radau quadrature based methods.
  - Depending on application (e.g. some PDEs) may be reasonable to obtain.
  - Easy for damped least-squares $(A^TA + \lambda^2 I)x = A^Tb$. Hence good for LSLQ and LSQR.
- When A is indefinite, the error bound for SYMMLQ seems a good estimate.
- Extend to LSLQ for least-squares problems.

References

- R. Estrin, D. Orban, and S.
  Euclidean-norm error bounds for SYMMLQ and CG
  SIMAX (revised May 2017)

References

- R. Estrin, D. Orban, and S.
  Euclidean-norm error bounds for SYMMLQ and CG
  SIMAX (revised May 2017)

- R. Estrin, D. Orban, and S.
  LSLQ: An iterative method for linear least-squares with an error minimization property
  SIMAX (in revision June 2017)

References

- R. Estrin, D. Orban, and S.
  Euclidean-norm error bounds for SYMMLQ and CG
  SIMAX (revised May 2017)

- R. Estrin, D. Orban, and S.
  LSLQ: An iterative method for linear least-squares with an error minimization property
  SIMAX (in revision June 2017)
  - Seismic inverse problem, PDE-constrained optimization

References

- R. Estrin, D. Orban, and S.
  Euclidean-norm error bounds for SYMMLQ and CG
  SIMAX (revised May 2017)
- R. Estrin, D. Orban, and S.
  LSLQ: An iterative method for linear least-squares with an error minimization property
  SIMAX (in revision June 2017)

  - Seismic inverse problem, PDE-constrained optimization
  - Error in gradient of penalty function is bounded by error in x

References

- R. Estrin, D. Orban, and S.
  Euclidean-norm error bounds for SYMMLQ and CG
  SIMAX (revised May 2017)

- R. Estrin, D. Orban, and S.
  LSLQ: An iterative method for linear least-squares with an error minimization property
  SIMAX (in revision June 2017)
  - Seismic inverse problem, PDE-constrained optimization
  - Error in gradient of penalty function is bounded by error in x
  - Monotonic decrease in error $\Rightarrow$ more accurate gradient

Special thanks

- Chris Paige

Special thanks

- Chris Paige
- Gene Golub

Special thanks

- Chris Paige
- Gene Golub
- Gene Golub, Gérard Meurant, Zdeněk Strakoš

Special thanks

- Chris Paige
- Gene Golub
- Gene Golub, Gérard Meurant, Zdeněk Strakoš
- Ron Estrin, Dominique Orban

Special thanks

- Chris Paige
- Gene Golub
- Gene Golub, Gérard Meurant, Zdeněk Strakoš
- Ron Estrin, Dominique Orban
- The Householder XX committees

Special thanks

- Chris Paige
- Gene Golub
- Gene Golub, Gérard Meurant, Zdeněk Strakoš
- Ron Estrin, Dominique Orban
- The Householder XX committees
- Julianne, Tia, June, Leah!

Special thanks

- Chris Paige
- Gene Golub
- Gene Golub, Gérard Meurant, Zdeněk Strakoš
- Ron Estrin, Dominique Orban
- The Householder XX committees
- Julianne, Tia, June, Leah!
- Yuja Wang, youtube

Special thanks

- Chris Paige
- Gene Golub
- Gene Golub, Gérard Meurant, Zdeněk Strakoš
- Ron Estrin, Dominique Orban
- The Householder XX committees
- Julianne, Tia, June, Leah!
- Yuja Wang, youtube
- Late-night talk shows (come back Jay Leno!)