# On error estimation in CGLS and LSQR

Petr Tichý

Charles University, Prague

based on joint work with

Jan Papež, Gérard Meurant, Zdeněk Strakoš

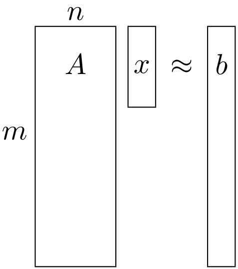
NLACIRM24

dedicated to Michela Redivo-Zaglia and Hassane Sadok

September 16-20, 2024, CIRM Luminy

# Least-squares problems

## equivalent formulations




$$\min_{x\in\mathbb{R}^n} \| \, b - Ax \, \|$$

$$A^T A \, x \;=\; A^T b \,, \qquad A \, x \;=\; b_{|\mathcal{R}(A)}$$

$A$ has full column rank (for simplicity)

# Solving system of normal equations

## using conjugate gradients (CG)

$$\min_{x \in \mathbb{R}^n} \| b - Ax \| \quad \Leftrightarrow \quad A^T A\, x = A^T b.$$

Let $x_0 = 0$, CG constructs $x_k \in \mathcal{K}_k(A^T A, A^T b)$ which minimize

$$\begin{aligned} \| x - x_k \|_{A^T A}^2 &= \| \overbrace{(b - Ax_k)}^{r_k} - \overbrace{(b - Ax)}^{r} \|^2 \\ &= \| b - Ax_k \|^2 - \| r \|^2, \end{aligned}$$

$r \perp \mathcal{R}(A)$ and $r_k - r \in \mathcal{R}(A)$.

algorithms ?

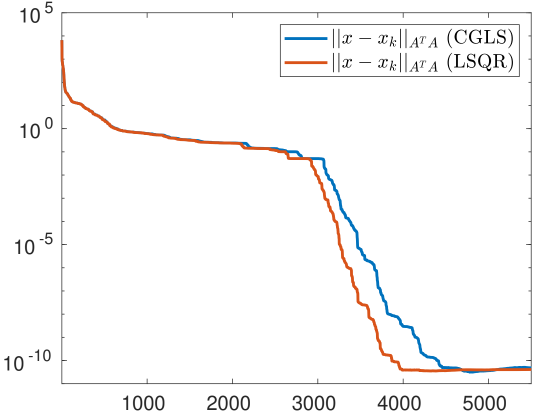
stopping ?

# CG for normal equations

## CGLS and LSQR algorithms

CGLS [Hestenes & Stiefel, 1952]

LSQR [Paige & Saunders, 1982]



# Stopping criteria

## Normwise relative backward error

- Given $\varepsilon$, ideally stop when $x_k$ solves

$$(A+E)^T(A+E)\,x_k = (A+E)^T(b+f) \qquad (\star)$$

and $$\min_{E,f,\xi}\left\{\xi : (\star) \text{ holds with } \frac{\|E\|}{\|A\|} \le \xi,\ \frac{\|f\|}{\|b\|} \le \xi\right\} \le \varepsilon$$

- Sufficient conditions (too conservative) [Paige & Saunders, 1982]

$$\begin{aligned} \|r_k\| &\le \varepsilon\,(\,\|A\|\,\|x_k\| + \|b\|\,)\,, \\ \|A^T r_k\| &\le \varepsilon\,\|A\|\,\|r_k\|\,. \end{aligned}$$

- Sufficient condition (asymptotically tight)

$$\frac{\|\,x - x_k\,\|_{A^T A}}{\|A\|\,\|x_k\| + \|b\|} \le \varepsilon\,.$$

[Chang & Paige & Titley-Peloquin, 2009]

Recall

$$\| x - x_k \|_{A^TA}^2 = \| r_k \|^2 - \| r \|^2.$$

Stop if

$$\frac{\| r_k \|^2 - \| r \|^2}{\| r \|^2} \le \varepsilon$$

i.e.

$$\| x - x_k \|_{A^TA}^2 \le \frac{\varepsilon}{1+\varepsilon} \|r_k\|^2 .$$

[Papež & Tichý, 2024]

How to estimate $\| x - x_k \|_{A^T A}$?

The conjugate gradient (CG) algorithm is almost always the iterative method of choice for solving linear systems with symmetric positive definite matrices. This book

- describes and analyzes techniques based on Gauss quadrature rules to cheaply compute bounds on norms of the error and that can be used to derive reliable stopping criteria;
- shows how to compute estimates of the smallest and largest eigenvalues during CG iterations; and
- illustrates algorithms using many numerical experiments; these algorithms also can be easily incorporated into existing CG codes.

*Error Norm Estimation in the Conjugate Gradient Algorithm* is intended for those in academia and industry who use the conjugate gradient algorithm, including the many branches of science and engineering in which symmetric linear systems have to be solved.

**Gérard Meurant** is retired from the French Atomic Energy Commission (CEA), where he worked in applied mathematics from 1970 to 2008. He was research director at the time of his retirement. He is the author of more than 60 papers on numerical linear algebra and six books, including two books co-authored with Gene H. Golub.

**Petr Tichý** is an associate professor at the Faculty of Mathematics and Physics at Charles University in Prague, Czech Republic. He is the author of more than 27 journal publications and one textbook. His research covers a variety of topics in numerical linear algebra, optimization, approximation of functions, and round-off error analysis of algorithms.

For more information about SIAM books, journals, conferences, memberships, or activities, contact:

siam

Society for Industrial and Applied Mathematics
3600 Market Street, 6th Floor
Philadelphia, PA 19104-2688 USA
+1-215-382-9800
*siam@siam.org* • *www.siam.org*

SL06

ISBN: 978-1-61197-785-1

90000

9 781611 977851

Error Norm Estimation in the Conjugate Gradient Algorithm

Gérard Meurant • Petr Tichý

siam

SL06

siam Spotlights

# Error Norm Estimation in the Conjugate Gradient Algorithm

Gérard Meurant

Petr Tichý

# Error norm estimation in CG

$A$ is symmetric and positive definite, $Ax = b$

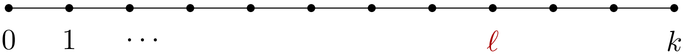
**input** $A$, $b$, $x_0$, $\tau$
$r_0 = b - Ax_0$, $p_0 = r_0$
**for** $k = 1, \ldots$ **do**
$\quad\vdots$
$\quad x_k = x_{k-1} + \gamma_{k-1} p_{k-1}$
$\quad r_k = r_{k-1} - \gamma_{k-1} A p_{k-1}$
$\quad\vdots$
$\quad \Delta_{k-1} = \gamma_{k-1} \|r_{k-1}\|^2$
$\quad (\ell, \mathtt{EST}) = \mathtt{adaptive}(\{\Delta_j\}_{j=0}^{k-1}, \tau)$
**end for**

Heuristically,

$$\frac{\|x - x_\ell\|_A^2 - \mathtt{EST}}{\|x - x_\ell\|_A^2} \leq \tau$$

$$\mathtt{EST} = \sum_{j=\ell}^{k-1} \Delta_j$$

$$\frac{\mathtt{EST}}{1 - \tau}$$




[Meurant & Papež & Tichý, 2021], [Papež & Tichý, 2024]

# Error norm estimation in CGLS

for solving least-squares problems

**input** $A$, $b$, $x_0$, $\tau$
$r_0 = b - Ax_0$
$s_0 = p_0 = A^T r_0$
**for** $k = 1, 2, \ldots$ **do**
  $q_{k-1} = Ap_{k-1}$
  $\gamma_{k-1} = \|s_{k-1}\|^2 / \|q_{k-1}\|^2$
  $x_k = x_{k-1} + \gamma_{k-1} p_{k-1}$
  $r_k = r_{k-1} - \gamma_{k-1} q_{k-1}$
  $s_k = A^T r_k$
  $\delta_k = \|s_k\|^2 / \|s_{k-1}\|^2$
  $p_k = s_k + \delta_k p_{k-1}$
  $\Delta_{k-1}^{\mathrm{CGLS}} = \gamma_{k-1} \, \|s_{k-1}\|^2$
  $(\ell, \mathtt{EST}) = \mathtt{adaptive}(\{\Delta_j^{\mathrm{CGLS}}\}_{j=0}^{k-1}, \tau)$
**end for**

[Papež & Tichý, 2024]

# Error norm estimation in LSQR

**input** $A$, $b$, $x_0 = 0$, $\tau$

$\beta_1 u_1 = b$, $\alpha_1 v_1 = A^T u_1$

$w_1 = v_1$, $\bar{\phi}_1 = \beta_1$, $\bar{\rho}_1 = \alpha_1$

**for** $k = 1, 2, \ldots$ **do**

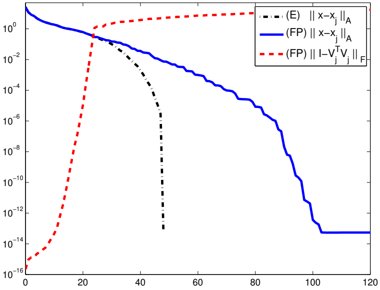
- $\beta_{k+1} u_{k+1} = A v_k - \alpha_k u_k$
- $\alpha_{k+1} v_{k+1} = A^T u_{k+1} - \beta_{k+1} v_k$
- $\rho_k = (\bar{\rho}_k^2 + \beta_{k+1}^2)^{1/2}$
- $c_k = \bar{\rho}_k / \rho_k$, $s_k = \beta_{k+1} / \rho_k$
- $\theta_{k+1} = s_k \alpha_{k+1}$, $\bar{\rho}_{k+1} = -c_k \alpha_{k+1}$
- $\bar{\phi}_{k+1} = s_k \bar{\phi}_k$, $\phi_k = c_k \bar{\phi}_k$
- $x_k = x_{k-1} + (\phi_k / \rho_k) w_k$
- $w_{k+1} = v_{k+1} + (\theta_{k+1} / \rho_k) w_k$
- $\Delta_{k-1}^{\mathrm{LSQR}} = \phi_k^2$
- $(\ell, \texttt{EST}) = \texttt{adaptive}(\{\Delta_j^{\mathrm{LSQR}}\}_{j=0}^{k-1}, \tau)$

**end for**

[Papež & Tichý, 2024]

# Finite precision computations

## and the estimates in CG




$$\| x - x_{k-1} \|_A^2 = \Delta_{k-1} + \| x - x_k \|_A^2 + \texttt{ERR}$$

and

$$\texttt{ERR} \ll \Delta_{k-1} \quad \Leftrightarrow \quad \frac{|r_k^T p_{k-1}|}{\|r_{k-1}\|^2} \ll 1\,.$$

[Strakoš & Tichý, 2002]

# Finite precision computations

## and the estimates in CGLS and LSQR

- In finite precision LSQR it holds that

$$\| x - x_{k-1} \|_{A^TA}^2 - \| x - x_k \|_{A^TA}^2 = \phi_k^2 \left(1 + \eta_{k-1} - \eta_k\right)$$

where

$$\eta_k \equiv \frac{\theta_{k+1}}{\rho_k} v_{k+1}^T w_k \,.$$

- Estimates are reliable, if **local orthogonality** is preserved

$$\frac{|s_k^T p_{k-1}|}{\|s_{k-1}\|^2} \overset{\text{CGLS}}{\ll} 1, \qquad \frac{\theta_{k+1}}{\rho_k} |v_{k+1}^T w_k| \overset{\text{LSQR}}{\ll} 1$$

where $s_k = A^T A x_k - A^T b$.

[Papež & Tichý, 2024]

# Preconditioning

## and the $A^T A$-norm of the error

- $L$ nonsingular, modify the problem

$$\min_{z \in \mathbb{R}^n} \| b - \underbrace{A L^{-T}}_{\hat{A}} \underbrace{L^T z}_{\hat{z}} \| \,.$$

- Solve the corresponding system of normal equations

$$\underbrace{L^{-1} A^T}_{\hat{A}^T} \underbrace{A L^{-T}}_{\hat{A}} \underbrace{L^T x}_{\hat{x}} = \underbrace{L^{-1} A^T}_{\hat{A}^T} b.$$

  $L \rightarrow$ a **split preconditioner** for $A^T A$.

- It holds that

$$\| \hat{x} - \hat{x}_k \|^2_{\hat{A}^T \hat{A}} = \| x - x_k \|^2_{A^T A}$$

  and our techniques can be used in PCGLS and PLSQR.

# Note on least-norm problems

Let $A \in \mathbb{R}^{m\times n}$, $b \in \mathbb{R}^m$, $b \in \mathcal{R}(A)$, consider

$$x = \arg\min_{z\in\mathbb{R}^n} \|z\| \quad \text{s.t.} \quad Az = b\,.$$

Solve $AA^T y = b$ using CG and compute $x_k = A^T y_k$. Then

$$\|\, x - x_k \,\|$$

is mimimized over $x_0 + \mathcal{K}_k(A^T A, A^T r_0)$.

CGNE [Craig, 1955; Faddeeva 1963]

CRAIG [Paige & Saunders, 1982]

results on error estimation in [Papež & Tichý, 2024]

# Numerical experiments

# Test problems

## SuiteSparse Matrix collection

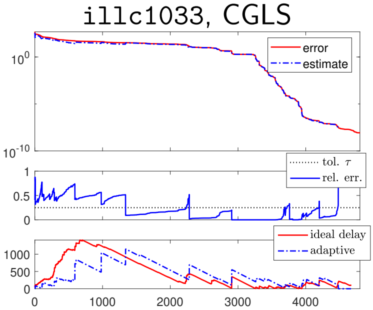
| problem | $m$ | $n$ | $b$ | precond |
|---|---|---|---|---|
| illc1033 | 1033 | 320 | ✓ | no |
| well1850 | 1850 | 712 | ✓ | no |
| illc1850 | 1850 | 712 | ✓ | ✓ |
| sls | 1 748 122 | 62 729 | rand | ✓ |

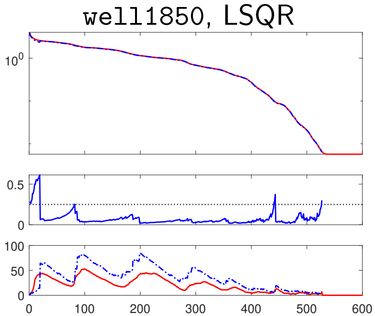
- $b$ comes together with the matrix or randomly generated.
- $L$ constructed using the incomplete Cholesky of $A^T A$ without explicitly forming it $\rightarrow$ MATLAB interface of `HSL_MI35`.

[HSL library], [Scott & Tůma, 2014]

# Problems without preconditioning

`illc1033` and `well1850`

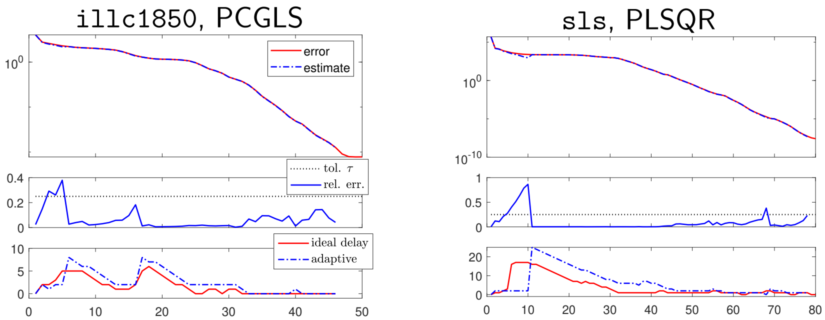



$$\frac{\|\, x - x_\ell \,\|^2_{A^TA} - \mathtt{EST}}{\|\, x - x_\ell \,\|^2_{A^TA}} \leq \tau = 0.25$$

# Problems with preconditioning

`illc1850` and `sls`




$$\frac{\| x - x_\ell \|^2_{A^TA} - \mathtt{EST}}{\| x - x_\ell \|^2_{A^TA}} \leq \tau = 0.25$$

# Conclusions

- The techniques for **error estimation** developed for CG can also be used in **CGLS** and **LSQR**.

- The estimates are **cheap**, numerically reliable, and work with preconditioning.

- The **heuristic strategy** for estimating the quantity of interest with a tolerance $\tau$ has shown to be **robust** and **reliable**.

- In the final stage, the suggested $\ell$ is usually **almost optimal**. Important for stopping the iterations.

# Related papers

J. Papež, P. Tichý,

[Estimating error norms in CG-like algorithms for least-squares and least-norm problems, Numer. Algorithms 97, pp. 1-28, 2024.]

https://github.com/JanPapez/CGlike-methods-with-error-estimate

- G. Meurant, P. Tichý, [Error Norm Estimation in the Conjugate Gradient Algorithm, SIAM Spotlights, Philadelphia, PA, 2024, x+127 p.]
- G. Meurant, J. Papež, P. Tichý, [Accurate error estimation in CG, Numer. Algorithms 88, 2021, pp. 1337-1359.]
- X.W. Chang, C.C. Paige, D. Titley-Peloquin, [Stopping criteria for the iterative solution of linear least squares problems. SIAM J. Matrix Anal. Appl. 31, 2009, pp. 831-852.]
- Z. Strakoš, P. Tichý, [On error estimation in CG and why it works in FP computations, Electron. Trans. Numer. Anal. 13, 2002, pp. 56–80.]

**Thank you for your attention!**