On Estimating the A-norm of the Error in CG and PCG

Zdeněk Strakoš*

Petr Tichý*†

*Institute of Computer Science, Academy of Sciences, Prague.

†Technical University of Berlin.

Monopoli, Bari, September 2003

Consider a system of linear algebraic equations

$$Ax = b$$

where $A \in \mathbb{R}^{n \times n}$ is symmetric positive definite, $b \in \mathbb{R}^n$.

given $x_0 \in \mathbb{R}^n$, $r_0 = b - Ax_0$

Conjugate Gradient Method (CG)

$$\boxed{\begin{aligned} x_j &\in x_0 + \mathcal{K}_j(A, r_0) \\ \|x - x_j\|_A &= \min_{u \in x_0 + \mathcal{K}_j(A, r_0)} \|x - u\|_A, \end{aligned}}$$

where

$$\begin{aligned} \mathcal{K}_j(A, r_0) &\equiv \text{span}\,\{r_0, \cdots, A^{j-1} r_0\} \\ \|x - x_j\|_A &\equiv ((x - x_j), A(x - x_j))^{\frac{1}{2}} \end{aligned}$$

Conjugate Gradient Method (CG)
Hestenes and Stiefel (1952)

given x_0, $r_0 = b - Ax_0$, $p_0 = r_0$,

for $j = 0, 1, 2, \ldots$

$$\gamma_j = \frac{(r_j, r_j)}{(p_j, Ap_j)}$$

$$x_{j+1} = x_j + \gamma_j \, p_j$$

$$r_{j+1} = r_j - \gamma_j \, Ap_j$$

$$\delta_{j+1} = \frac{(r_{j+1}, r_{j+1})}{(r_j, r_j)}$$

$$p_{j+1} = r_{j+1} + \delta_{j+1} \, p_j.$$

How to measure quality of approximation?

... it depends on what problem we solve.

- **using residual information,**
  – normwise backward error,
  – relative residual norm.

  Hestenes and Stiefel: "Using of the residual vector r_j as a measure of the "goodness" of the estimate x_j is not reliable" (HS 1952 p. 410)

- **using error estimates,**
  – estimate of the A-norm of the error,
  – estimate of the Euclidean norm of the error.

  Hestenes and Stiefel: "The function $(x-x_j, A(x-x_j))$ can be used as a measure of the "goodness" of x_j as an estimate of x." (HS 1952 p. 413)

On Estimating the A-norm of the Error

Outline

1. Basic identity
2. Construction of estimate in CG
3. Estimates in finite precision arithmetic
4. Estimate of the A-norm of the error in PCG
5. Numerical experiments
6. Conclusions

1. Basic identity

At any iteration step j, CG (implicitly) determines weights and nodes of the j-point Gauss quadrature

$$\int_{\zeta}^{\xi} f(\lambda)\, d\omega(\lambda) = \sum_{i=1}^{j} \omega_i^{(j)} f(\theta_i^{(j)}) + R_j(f).$$

The Gauss quadrature for $f(\lambda) \equiv \lambda^{-1}$ gives

$$C_n \;=\; C_j \;+\; \frac{\|x - x_j\|_A^2}{\|r_0\|^2}, \qquad C_n = \frac{\|x - x_0\|_A^2}{\|r_0\|^2}\,.$$

C_n, C_j ... continued fractions corresponding to $\omega(\lambda)$ and $\omega^{(j)}(\lambda)$
Golub: DGN-1978, GF-1993, GM-1994, GS-1994, GM-1997, ...

Equivalent formulas (multiplied by $\|r_0\|^2$)

In general (ST-2002):

$$\|x - x_0\|_A^2 \;=\; \|r_0\|^2 \,\text{Gauss}\, Q(j) + \|x - x_j\|_A^2.$$

W-2000:

$$r_0^T(x - x_0) \;=\; r_0^T(x_j - x_0) + \|x - x_j\|_A^2\,.$$

HS-1952, ST-2002:

$$\|x - x_0\|_A^2 \;=\; \sum_{i=0}^{j-1} \gamma_i \|r_i\|^2 + \|x - x_j\|_A^2.$$

This formula derived *purely algebraically* is equivalent to the Gauss quadrature formula!

2. Construction of estimate

Idea: Consider, for example,

$$\|x - x_j\|_A^2 \;=\; \|r_0\|^2 \left[C_n - C_j\right] \,.$$

Run d extra steps. Subtracting identity for $\|x - x_{j+d}\|_A^2$ gives

$$\|x - x_j\|_A^2 \;=\; \|r_0\|^2 \left[C_{j+d} - C_j\right] \;+\; \|x - x_{j+d}\|_A^2 \,.$$

When $\|x - x_j\|_A^2 \gg \|x - x_{j+d}\|_A^2$, we have a tight (lower) estimate.

GS-1994 and GM-1997

Mathematically equivalent estimates:

$$\|x - x_j\|_A^2 = EST^2 + \|x - x_{j+d}\|_A^2$$

GS-1994 and GM-1997

$$\eta_{j,d} \;=\; \|r_0\|^2 \,[C_{j+d} - C_j]$$

W-2000

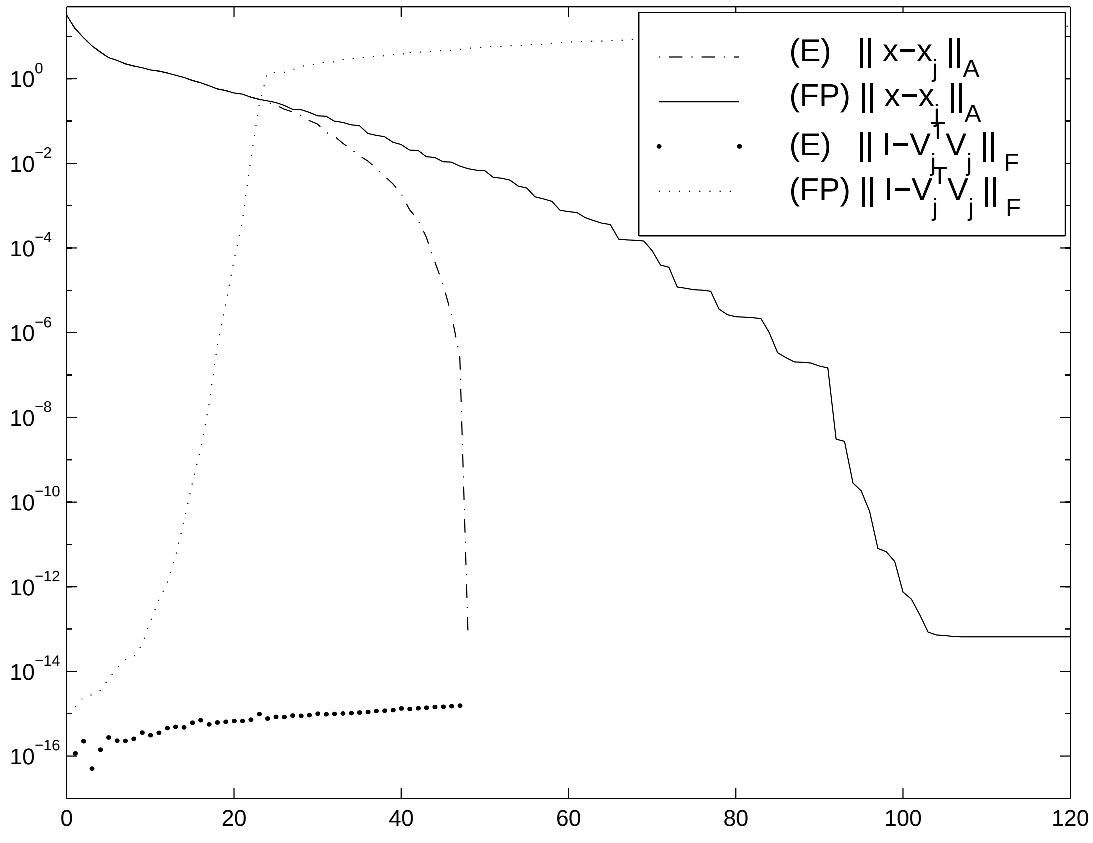
$$\mu_{j,d} \;=\; r_0^T(x_{j+d} - x_j)$$

HS-1952

$$\nu_{j,d} \;=\; \sum_{i=j}^{j+d-1} \gamma_i \|r_i\|^2$$

3. Estimates in finite precision arithmetic

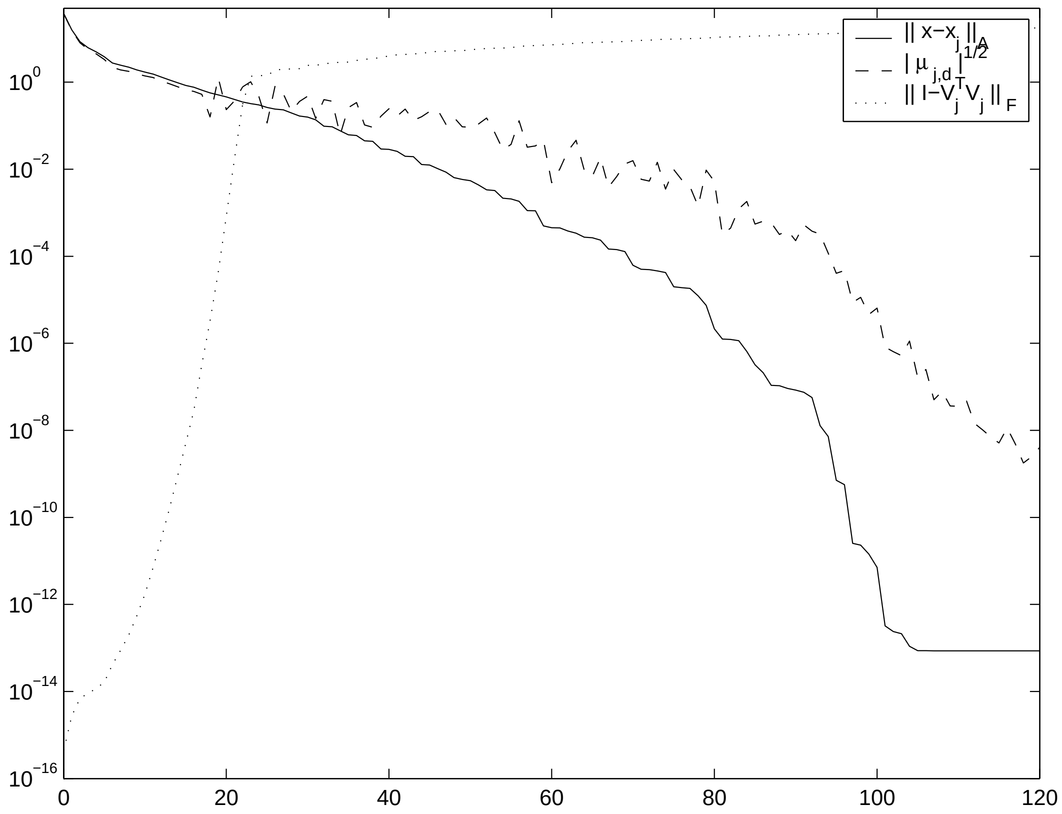
orthogonality is lost, convergence is delayed!



- The problem consists in validity of the whole identity

$$\|x - x_j\|_A^2 = EST^2 + \|x - x_{j+d}\|_A^2$$

An example:
$\mu_{j,d} = r_0^T (x_{j+d} - x_j)$
does not work!



Do the estimates give good information in practical computations?

$\eta_{j,d}$ **yes** GS-1994, GM-1997 (alg. **CGQL**)

Based on Greenbaum (1989), Greenbaum, S (1992), $\sqrt{\varepsilon}$ **limit**.

$\mu_{j,d}$ **no** ST-2002

$\nu_{j,d}$ **yes** ST-2002

Based on detailed analysis of preserving the local orthogonality in the HS algorithm, with using Paige (1971-80), Greenbaum (1989, 97)

4. Estimate in PCG

The cg-iterates are thought of beeing applied to

$$\widehat{A}\widehat{x} = \widehat{b}$$

giving CG for $\widehat{A}\widehat{x} = \widehat{b}$. We consider symmetric preconditioning

$$\widehat{A} = L^{-1}AL^{-T}, \qquad \widehat{b} = L^{-1}b.$$

Definition

$$M \equiv LL^T, \quad \gamma_j \equiv \widehat{\gamma}_j, \quad \delta_j \equiv \widehat{\delta}_j,$$

$$x_j \equiv L^{-T}\widehat{x}_j, \quad r_j \equiv L\,\widehat{r}_j, \quad p_j \equiv L^{-T}\widehat{p}_j, \quad s_j \equiv M^{-1}r_j :$$

Preconditioned Conjugate Gradient Method (PCG)

given x_0, $r_0 = b - Ax_0$, $s_0 = M^{-1} r_0$, $p_0 = s_0$,

for $j = 0, 1, 2, \ldots$

$$\gamma_j = \frac{(r_j, s_j)}{(p_j, Ap_j)}$$

$$\begin{aligned} x_{j+1} &= x_j + \gamma_j\, p_j \\ r_{j+1} &= r_j - \gamma_j\, Ap_j \\ s_{j+1} &= M^{-1} r_{j+1} \end{aligned}$$

$$\delta_{j+1} = \frac{(r_{j+1}, s_{j+1})}{(r_j, s_j)}$$

$$p_{j+1} = s_{j+1} + \delta_{j+1}\, p_j.$$

How to estimate the A-norm of the error in PCG?

Hestenes-Stiefel estimate for $\widehat{A}\widehat{x} = \widehat{b}$:

$$\|\widehat{x} - \widehat{x}_j\|_{\widehat{A}}^2 \;=\; \sum_{i=j}^{j+d-1} \widehat{\gamma}_i \;\|\widehat{r}_i\|^2 \;+\; \|\widehat{x} - \widehat{x}_{j+d}\|_{\widehat{A}}^2 .$$

Substitution $\widehat{A} = L^{-1}AL^{-T}$, $\widehat{x}_j = L^T x_j$, $\widehat{\gamma}_i = \gamma_i$, $\widehat{r}_i = L^{-1} r_i$ gives

$$\|\widehat{x} - \widehat{x}_j\|_{\widehat{A}}^2 \;=\; (L^T x - L^T x_j) L^{-1} A L^{-T} (L^T x - L^T x_j) \;=\; \|x - x_j\|_A^2$$

and

$$\boxed{\|x - x_j\|_A^2 \;=\; \sum_{i=j}^{j+d-1} \gamma_i \;(r_i, s_i) \;+\; \|x - x_{j+d}\|_A^2 \,.}$$

Numerical stability of the estimate in PCG.

ST-2003: rounding error analysis.

Analysis based on:

- rounding error analysis from ST-2002,
- solving of

$$Ms_{j+1} = r_{j+1}$$

enjoys perfect normwise backward stability (Hi-1996).

result:

Until $\|x - x_j\|_A$ reaches a level close to $\varepsilon \, \|x - x_0\|_A$, the estimate $\nu_{j,d}$ **must work**.

Estimate of the relative A-norm of the error $\frac{\|x-x_j\|_A}{\|x-x_0\|_A}$.

Since

$$\begin{aligned} \|x - x_j\|_A^2 &= \nu_{j,d} \quad\quad + \|x - x_{j+d}\|_A^2, \\ \|x - x_0\|_A^2 &= \nu_{0,j+d} + \|x - x_{j+d}\|_A^2, \end{aligned}$$

the square root of

$$\varrho_{j,d} \equiv \frac{\nu_{j,d}}{\nu_{0,j+d}} = \frac{\|x - x_j\|_A^2 - \|x - x_{j+d}\|_A^2}{\|x - x_0\|_A^2 - \|x - x_{j+d}\|_A^2} \leq \frac{\|x - x_j\|_A^2}{\|x - x_0\|_A^2}$$

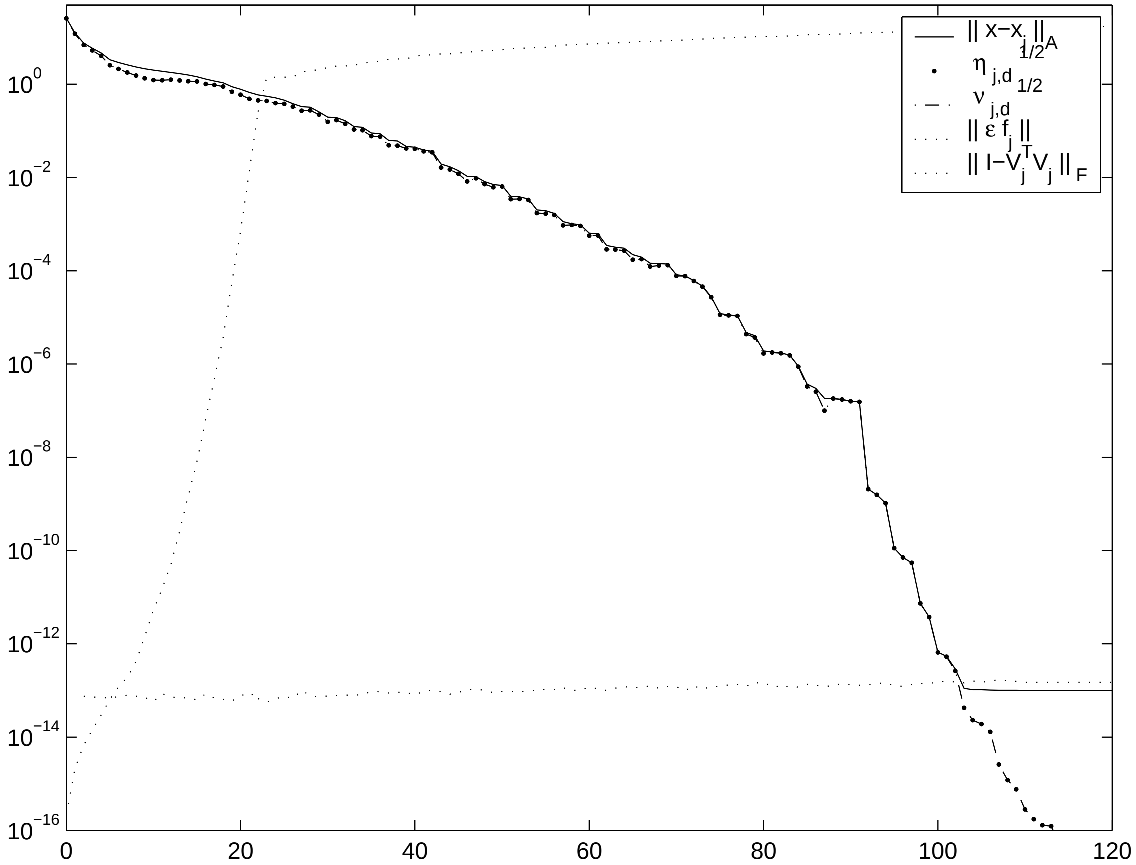
is a lower bound for the jth relative A-norm of the error.

5. Numerical Experiments

CG

Example from Strakoš-1991

$n = 48$, $d = 4$,
$\kappa(A) = 10000$
$\lambda_{max} = 1000$

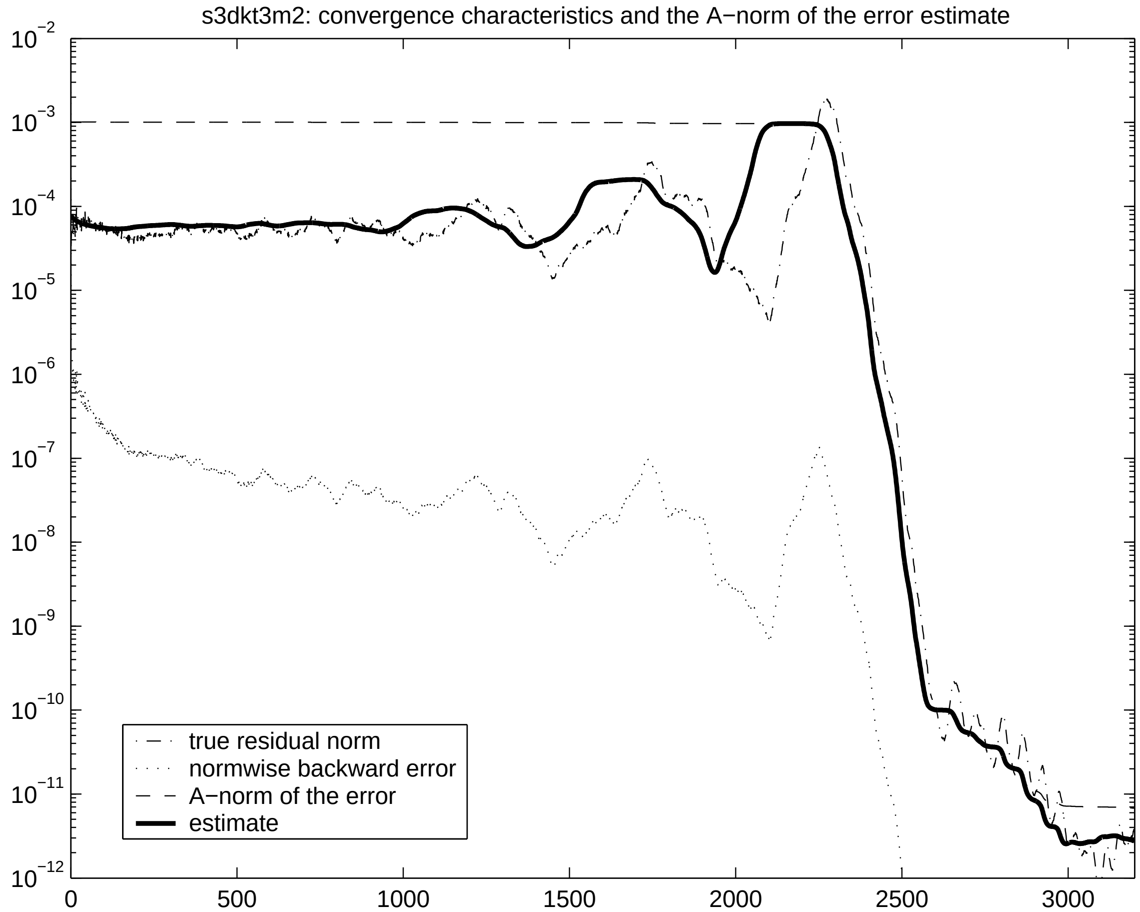


R. Kouhia: Cylindrical shell

PCG

Matrix Market
Cylshell
matrix S3DKT3M2

$\kappa(A) = 3.62e + 11$
$n = 90499$, $d = 200$
$L =$ cholinc(A,0)

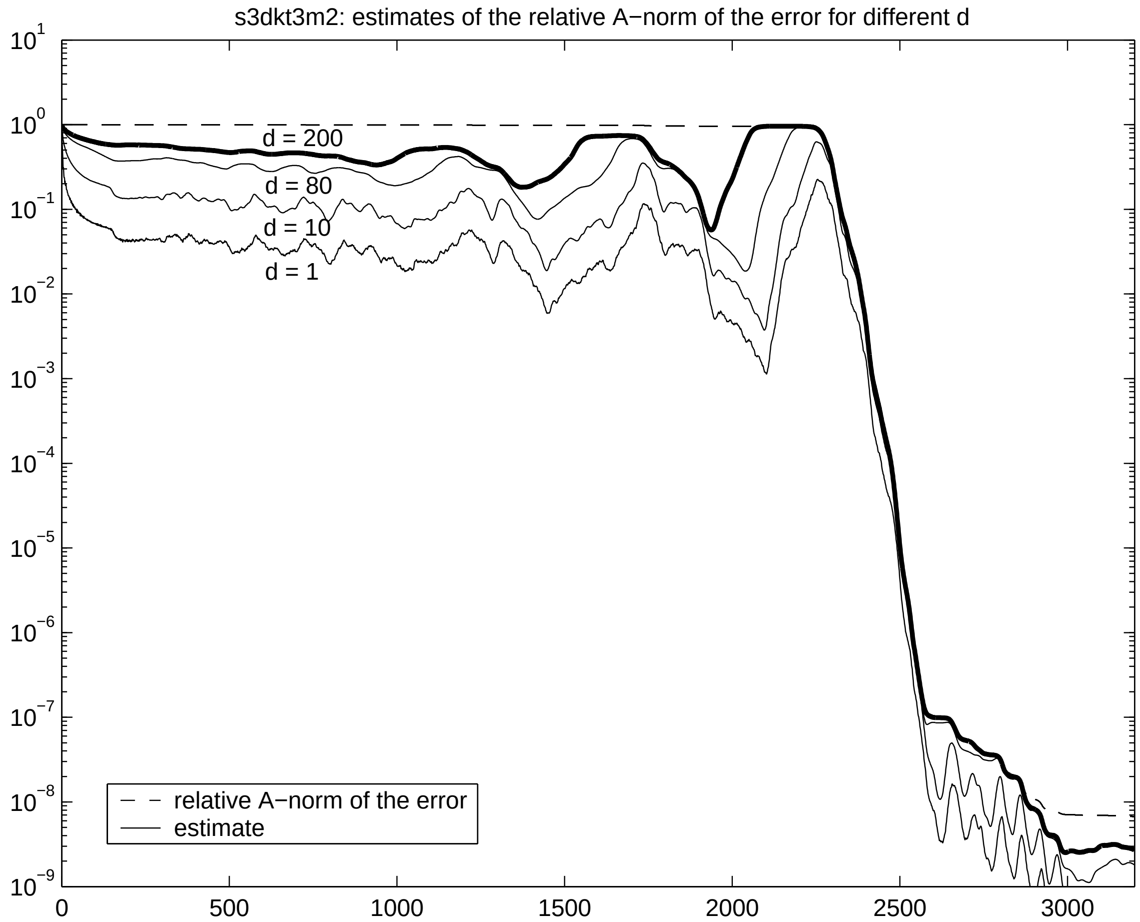


R. Kouhia: Cylindrical shell

PCG

source:
Matrix Market
Collection Cylshell
matrix S3DKT3M2

$\kappa(A) = 3.62e + 11$
$n = 90499$,
$L =$ cholinc(A,0)

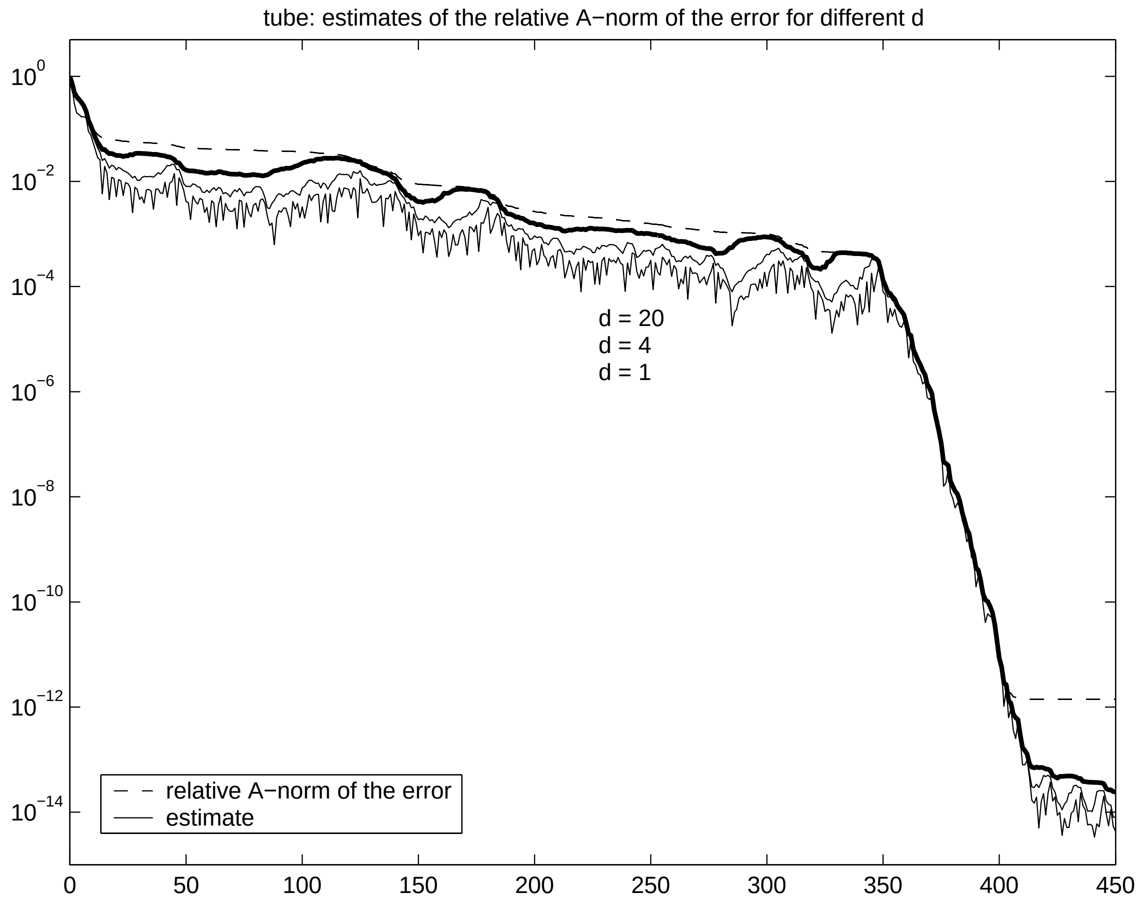


R. Kouhia: Cylindrical shell

PCG

source:
R. Kouhia homepage
matrix tube1-2

$n = 21498$,
$L =$ cholinc(A,1e-5)



6. Conclusions

• Up-to-now used bounds (based on Gauss quadrature) are **mathematically equivalent** to the formulas given (but somehow hidden) in the original Hestenes and Stiefel paper.

• Hestenes and Stiefel estimate is very simple, it can be computed almost for free and it is **numerically stable**.

• We suggest the estimates $\nu_{j,d}^{1/2}$ and $\varrho_{j,d}^{1/2}$ to be incorporated into **software realizations** of the PCG method.

• The estimates are tight if the A-norm of the error **reasonably decreases**.

Open problem: The adaptive choice of the parameter d.

More details can be found at

`http://www.cs.cas.cz/~strakos`
`http://www.cs.cas.cz/~tichy`

papers

• Strakoš, Z. and Tichý, P., *On Error Estimation in the Conjugate Gradient Method and Why It Works In Finite Precision Computations*, Electron. Trans. Numer. Anal. (ETNA), Volume 13, pp. 56-80, 2002.

• Strakoš, Z. and Tichý, P., *Simple estimation of the A-norm of the error in the Preconditioned Conjugate Gradient Method*, submitted to Comput. Methods Appl. Mech. Engrg., 2003.

technical report

• Strakoš, Z. and Tichý, P., *Simple estimation of the A-norm of the error in the Preconditioned Conjugate Gradient Method*, Technical Report No. 892, ICS AS CR, Prague, Czech Republic, 2003.