On Error Estimation in the Conjugate Gradient Method

Zdeněk Strakoš[†,*]

Petr Tichý[†]

[†] Institute of Computer Science,
Academy of Sciences, Prague,

[*] Technical University Liberec, and
Emory University, Atlanta.

February 2002.

System of linear algebraic equations

$$Ax = b$$

$A \in \mathbb{R}^{n\times n}$ is symmetric positive definite,
$b \in \mathbb{R}^n,\ x_0\ ,\ r_0 = b - Ax_0\ .$

Conjugate Gradient Method (CG)

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

where

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

Hestenes and Stiefel (1952)

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

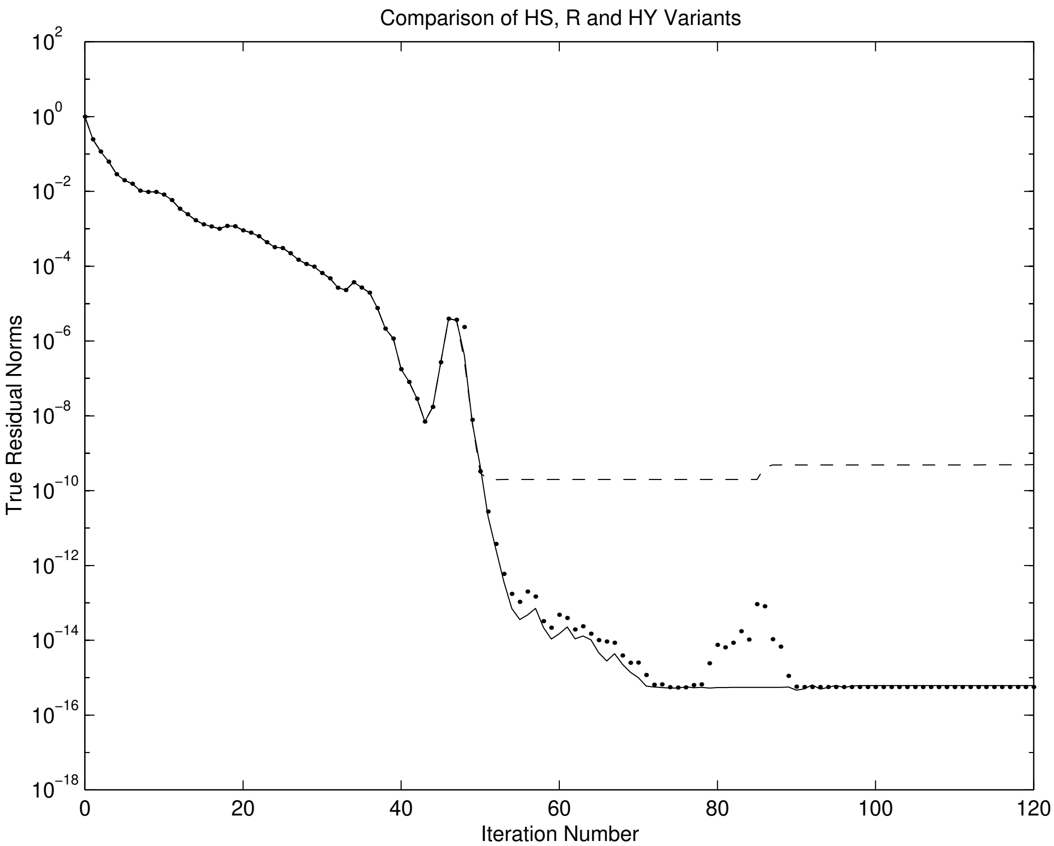
For $j = 1, 2, \ldots$

$$\begin{aligned}
&\gamma_{j-1} = (r_{j-1}, r_{j-1})/(p_{j-1}, Ap_{j-1}) \\
x_j &= x_{j-1} + \gamma_{j-1}\, p_{j-1} \\
r_j &= r_{j-1} - \gamma_{j-1}\, Ap_{j-1} \\
&\delta_j = (r_j, r_j)/(r_{j-1}, r_{j-1}) \\
p_j &= r_j + \delta_j\, p_{j-1}.
\end{aligned}$$

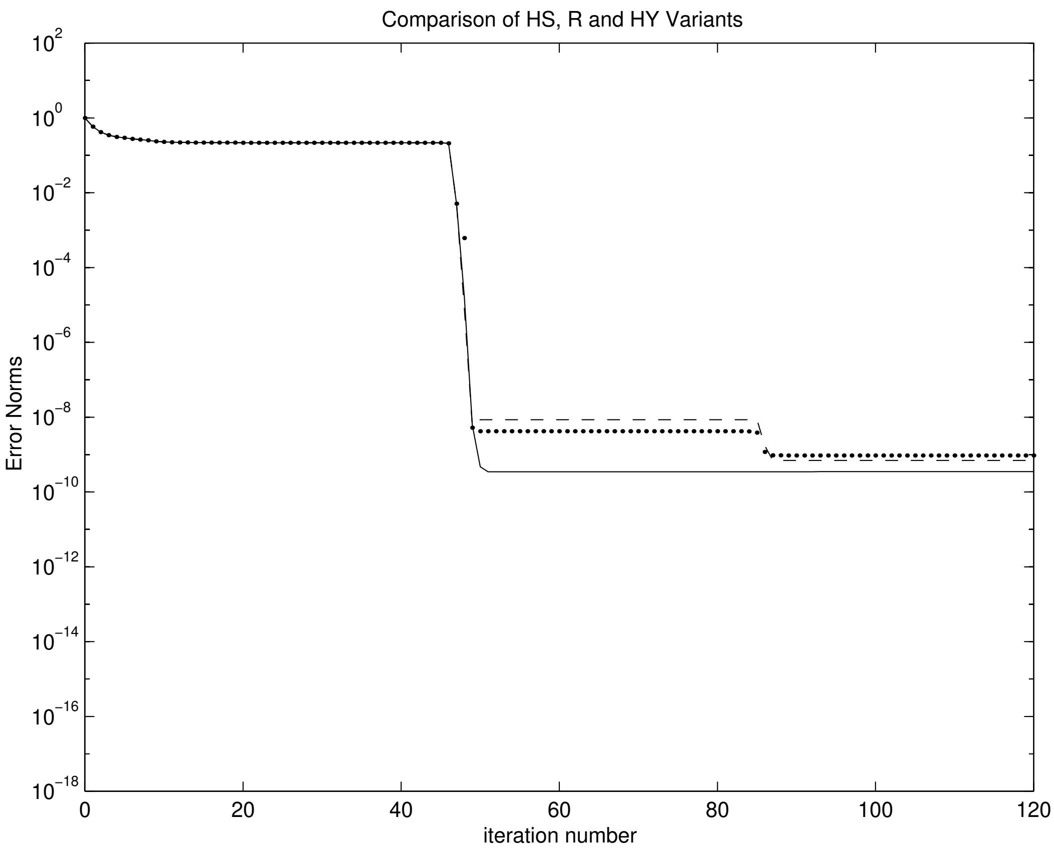
- "The cg-method is an iterative method which terminates in at most n steps if no rounding-off errors are encountered." (p. 410)

- "Normally this (residual) vector can be used as a measure of the "goodness" of the estimate x_j. However, this measure is not a reliable one because, as will be seen in section 18, it is possible to construct cases in which the *squared residual (norm)* $\|r_j\|^2$ increases at each step ..." (p. 410)

used in Gutknecht, S (2001)



• "In the cg-method the error vector $x - x_j$ is diminished in length at each step. The quantity $f(x_j) = (x - x_j, A(x - x_j))$, called the *error function* (squared A-norm or energy-norm of the error) is also diminished at each step. (p. 411)



- "The function $f(x_j)$ can be used as a measure of the "goodness" of x_j as an estimate of x." (p. 413)

$$\boxed{\frac{1}{\|A^{-1}\|}\,\|x - x_j\| \;\leq\; \|r_j\| \;\leq\; \|A\|\,\|x - x_j\|}$$

Notice that in this experiment

$$\|A\| \approx 4.8,\ \kappa(A) \approx 6 \times 10^7\,.$$

Theorem 6:1 (p. 416)

Using $f(x_j) = (x - x_j, A(x - x_j)) = \|x - x_j\|_A^2$,

$$
\begin{aligned}
f(x_{j-1}) - f(x_j) &= \gamma_{j-1}\|r_{j-1}\|^2 \\
&= \frac{(p_{j-1}, Ap_{j-1})}{\|p_{j-1}\|^2}\,\|x_{j-1} - x_j\|^2
\end{aligned}
$$

and for $0 \le j < k \le n$

$$
\boxed{f(x_j) - f(x_k) \;=\; \sum_{l=j}^{k-1} \gamma_l\,\|r_l\|^2}
$$

Theorem 6:3 (p. 416)

At each step of the cg-algorithm the error vector $y_j = x - x_j$ *is reduced in length. In fact*

$$\|y_{j-1}\|^2 - \|y_j\|^2 = \frac{\|p_{j-1}\|^2}{(p_{j-1}, Ap_{j-1})} \left(f(x_{j-1}) + f(x_j)\right).$$

- "This result establishes the cg-method as a method of successive approximations and justifies the procedure of stopping the algorithm before the final step is reached." (p. 417)

For $0 \le j < k < n$

$$\boxed{\|y_j\|^2 - \|y_k\|^2 = \sum_{l=j}^{k-1} \frac{\|p_l\|^2}{(p_l, Ap_l)} \left(f(x_l) + f(x_{l+1})\right)}$$

History of CG after 1952 is worth a historical monography. Some of the results published in the HS paper were not used until recently in S, Tichý (2001-2002).

- Golub, O'Leary (1989)
- Reid (1971) – discussion
- Paige, Saunders (1975) – SYMMLQ, MINRES
- Chebyshev bounds unfortunately often identified with convergence (thorough description in Saad, Van der Vorst (2002)), condition number and preconditioning
- despite the HS paper, residual norm used overwhelmingly even in numerical linear algebra literature
- estimation of errors – Gene Golub

Outline

1. Basic identities
    - Gauss quadrature
    - algebraic manipulations
    - equivalent formulas
2. Construction of estimates
3. Should it work in finite precision arithmetic?
4. Numerical experiments
5. Conclusion

1. Basic identities

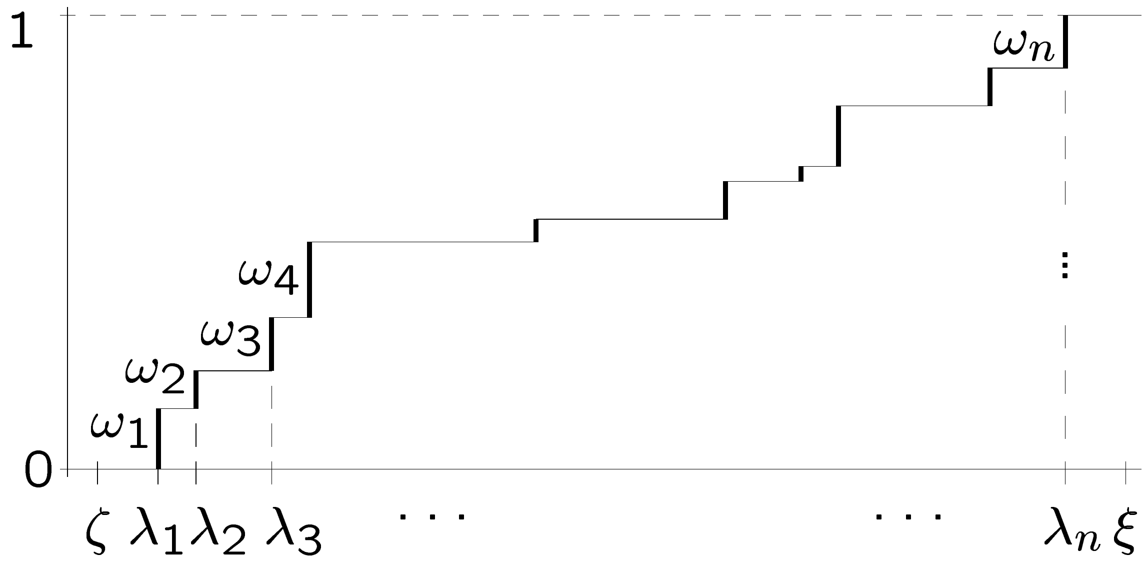
Independently of HS (1952), section 14, and of CG, Dahlquist, Eisenstat, Golub (1972):

For A and $v_1 = r_0/\|r_0\|$ consider the associated Riemann-Stieltjes integral

$$\int_{\zeta}^{\xi} f(\lambda)\, d\omega(\lambda) \;=\; \sum_{i=1}^{n} \omega_i\, f(\lambda_i)$$

with the monotonic piecewise constant distribution function $\omega(\lambda)$

- having n–points of increase $\lambda_1, \cdots, \lambda_n$ (eigenvalues of A),
- and the individual jumps $\omega_1, \cdots, \omega_n$ (squared sizes of the components of v_1 in the eigenvector basis).



$$\|x - x_0\|^2 = \|y_0\|^2 = \int_\zeta^\xi \lambda^{-2} d\omega(\lambda)$$

Gauss-Radau quadrature, Lanczos process and the algorithm by Galant for computing it (Jacobi method).

1.1. Gauss-quadrature inspiration

Dahlquist, Golub, Nash (1978):

- upper and lower bounds on $\|x - x_j\|_A$
- estimates on $\|x - x_j\|$
- all developed using moments and Gauss quadrature, and computed without determining nodes and weights
- "In order guarantee the numerical stability of the process . . ." a complete re-orthogonalization is an imperative (p. 158)
- In experiment, no difference between the simple and the re-orthogonalized processes was observed (p. 165)
  . . . explanation in Paige (1971-80)

Golub, Meurant (1994):

- an instrumental paper which summarizes theory of Gauss, Gauss-Radau and Gauss-Lobato quadrature in relation to error estimates in CG

- basis of further development of this line, Gerard Meurant (theory, experiments, references)

At any iteration step j, CG (implicitly) determines the j-point distribution function $\omega^{(j)}(\lambda)$ of the j-point Gauss quadrature,

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

In the rest we focus on the lower bound given by the Gauss quadrature.

Independently, started by a question by Golub, and by observations made by Horáček (1991) and Fischer (1992), Golub, Fischer (1994). Golub, Strakoš (1994):

- directly from the Gauss quadrature (or from the polynomial minimization)

$$(T_n^{-1})_{11} = (T_j^{-1})_{11} + \|x - x_j\|_A^2 / \|r_0\|^2$$

$$\boxed{C_n = C_j + \|x - x_j\|_A^2 / \|r_0\|^2}$$

- "reconstructing" the A-norm of the error by replacing the unknown C_n by some C_k, $k > j$, at a price of $k - j$ extra steps. $C_k - C_j$ computed explicitly ($\sqrt{\varepsilon}$ limit), the meaning misunderstood:

  "This is, of course, only a curiosity without a practical application ..." (p. 262)

- FP analysis (Part 3)

Golub, Meurant (1997):

- algorithm GQL (Gauss Quadrature and Lanczos)

- estimating $\|x - x_j\|_A$ in a stable way (an analogy of computing the difference $C_k - C_j$ directly without computing C_k and C_j).

- algorithm CGQL: estimating $\|x - x_j\|_A$ at a price of d extra steps, $k = j + d$.

Further papers:

Meurant (1997, 2001, 2001),
Calvetti et al. (1999, 2000, 2001),
Frommer and Weinberg (1999),
Arioli (2001), Arioli and Baldini (2001),
...

Warnick (2001),
Saylor and Smolarski (2001):

- scattering amplitude,
  generalizations to non-SPD problems

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

developed through moments or
Gauss quadrature.

This has inspired, together with the works by Golub and Meurant, our return to the problem.

1.2. Algebraic manipulations

We will focus on the A-norm of the error.
The simplest way:

$$x - x_0 = \sum_{l=0}^{n-1} \gamma_l \, p_l \, .$$

Using the **global** A-orthogonality,

$$\begin{aligned} \|x - x_0\|_A^2 &= \sum_{l=0}^{n-1} \gamma_l^2 \, (p_l, Ap_l) = \sum_{l=0}^{n-1} \gamma_l \, \|r_l\|^2 \\ &= \sum_{l=0}^{j-1} \gamma_l \, \|r_l\|^2 + \|x - x_j\|_A^2 \, , \end{aligned}$$

Axelsson and Kaporin (2001), p. 275.

An unspecified (most probably
local orthogonality based) way used in
Deuflhard (1993, 1994), p. 5 resp. 35,
Greenbaum (1997), p. 108.

- mathematical equivalence with the Gauss quadrature estimate seems unnoticed

- effects of rounding errors mostly not mentioned, or

  "... as soon as the superlinear convergence started, the estimate quickly converges to the A-norm of the error within 7 and more digits. This means that the orthogonality loss does not destroy the applicability of the estimate ..."

  Axelsson, Kaporin (p.278).

Using **local properties only** (orthogonality of the new residual to the previous direction vector)

$$
\begin{aligned}
\|x - x_j\|_A^2 &= \|x - x_{j+1} + x_{j+1} - x_j\|_A^2 \\
&= \|x - x_{j+1}\|_A^2 + \|x_{j+1} - x_j\|_A^2 \\
&\quad + 2(x - x_{j+1})A(x_{j+1} - x_j) \\
&= \|x - x_{j+1}\|_A^2 + \gamma_j \|r_j\|^2 \\
&\quad + 2r_{j+1}^T (x_{j+1} - x_j) \\
&= \|x - x_{j+1}\|_A^2 + \gamma_j \|r_j\|^2
\end{aligned}
$$

Here we used

$$
\begin{aligned}
b - Ax_{j+1} &= r_{j+1}\,, \\
r_{j+1}^T (x_{j+1} - x_j) &= 0\,.
\end{aligned}
$$

1.3. Equivalent formulas

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

$$\|x-x_0\|_A^2 = \|r_0\|^2\,(T_j^{-1})_{11} + \|x-x_j\|_A^2$$

$$\|x-x_0\|_A^2 = \|r_0\|^2\,C_j + \|x-x_j\|_A^2$$

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

$$\begin{aligned}\|x-x_0\|_A^2 = {} & r_0^T(x_j-x_0)+r_j^T(x_j-x_0) \\ & + \|x-x_j\|_A^2\end{aligned}$$

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

The last two formulas were derived *algebraically* using local properties only. They are equivalent to the Gauss quadrature-based ones.

2. Construction of estimates

Idea from Golub, S. (1994),
Golub, Meurant (1997):

We look for an estimate of $\|x - x_j\|_A$.
Run d extra steps. Subtracting identities for $\|x - x_{j+d}\|_A^2$ and $\|x - x_j\|_A^2$ gives

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

When $\|x - x_j\|_A^2 \gg \|x - x_{j+d}\|_A^2$, EST gives a tight (lower) estimate for $\|x - x_j\|_A$, with the inaccuracy given by $\|x - x_{j+d}\|_A$.

Mathematically equivalent estimates:

GS (1994) and GM (1997)

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

W (2000)

$$\mu^2_{j,d} \;=\; r_0^T(x_{j+d} - x_j)$$

ST (2001)

$$\vartheta^2_{j,d} \;=\; r_0^T(x_{j+d} - x_j) - r_j^T(x_j - x_0) + r_{j+d}^T(x_{j+d} - x_0)$$

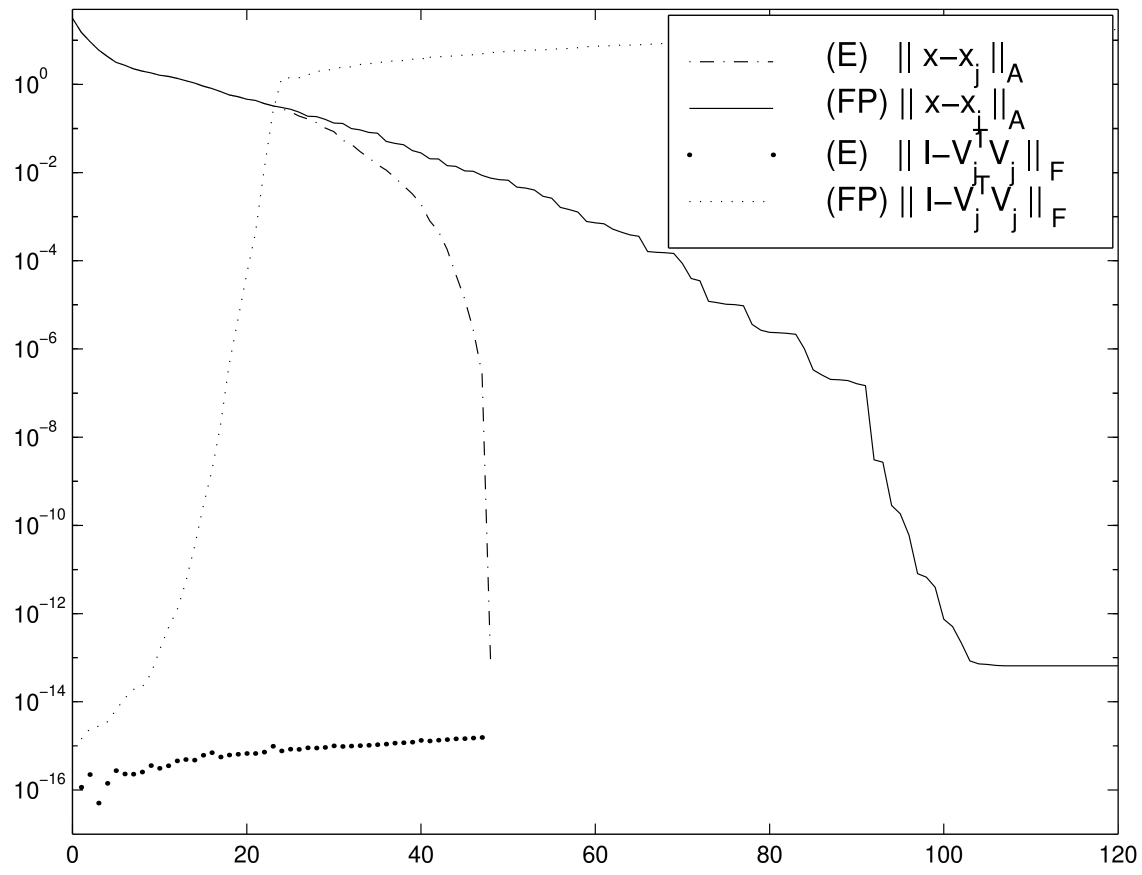
HS (1952)

$$\nu^2_{j,d} \;=\; \sum_{l=j}^{j+d-1} \gamma_l\,\|r_l\|^2$$

3. Finite precision arithmetic

Do the relations have *any relevance* for the quantities computed in FP arithmetic?

In finite precision CG orthogonality is lost and convergence is delayed!



Why should formulas assuming orthogonality (for the dash-dotted line)
give anything reasonable
when the orthogonality is lost
(for the solid line)?

- The problem *is not* in rounding errors while computing results using the formulas for η, μ, ϑ, or ν.

- The problem is in the fact, that the *input quantities* used in computing η, μ, ϑ, or ν were determined by the finite precision process, for which the orthogonality relations assumed in derivation of η, μ, ϑ, or ν do not hold.

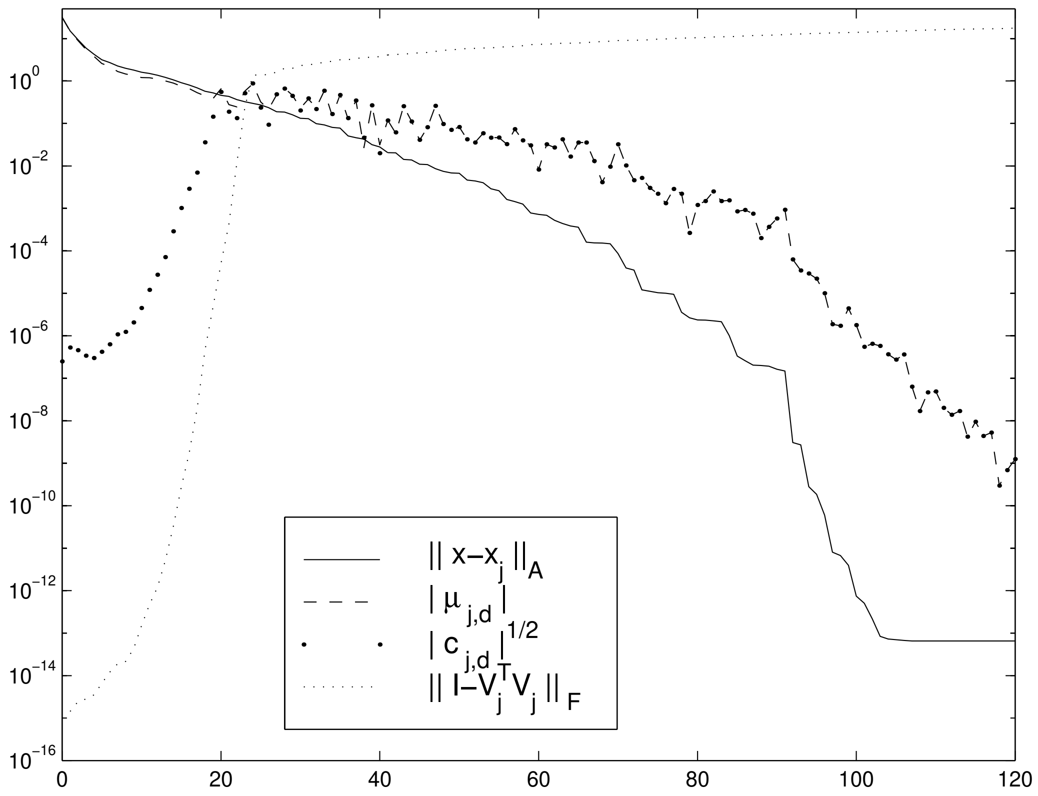
An example: μ does not work!

$$\mu_{j,d}^2 = r_0^T(x_{j+d} - x_j)$$

$$\vartheta_{j,d}^2 = \mu_{j,d}^2 - r_j^T(x_j - x_0) + r_{j+d}^T(x_{j+d} - x_0)$$

$$= \mu_{j,d}^2 + c_{j,d}$$

Without correction, $\mu_{j,d}$ fails with the loss of orthogonality.



Recall an analogy: accuracy of the computed Ritz values in the Lanczos method.

Exact arithmetic:

$$\min_i |\lambda_i - \theta_l^{(k)}| \leq \|Az_l^{(k)} - \theta_l^{(k)} z_l^{(k)}\| \leq \delta_{lk}$$

Finite precision arithmetic:

$$\begin{aligned} \min_i |\lambda_i - \theta_l^{(k)}| &\leq \|Az_l^{(k)} - \theta_l^{(k)} z_l^{(k)}\| \,/\, \|z_l^{(k)}\| \\ &\leq (\delta_{lk} + k^{1/2}\|A\|\varepsilon_1) \,/\, \|z_l^{(k)}\| \end{aligned}$$

looks similar to the previous case.
However, due to the loss of orthogonality it can happen $\|z_l^{(k)}\| \to 0$!

The quantity δ_{lk} is easy to compute with a negligible additional rounding errors. Does it tell anything about FP convergence of $\theta_l^{(k)}$?

Using the Ph.D. thesis and the subsequent papers by Chris Paige,

Theorem

$$\min_i |\lambda_i - \theta_l^{(k)}| \;\le\; \max\{2.5(\delta_{lk} + k^{1/2}\|A\|\varepsilon_1), \\ (k+1)^3\|A\|\varepsilon_2\}\,.$$

If δ_{lk} computed in a FP run of the Lanczos method is small, then $\theta_l^{(k)}$ computed in the same run must be converged to some λ_i.

Moral for our case:

Without rounding error analysis
there is no mathematical justification of usefulness of the estimates in FP computations.

Do the estimates give good information
in practical (FP) computations?

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

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

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

$\vartheta_{j,d}$ **yes** 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)

Using **local properties only** (orthogonality of the new residual to the previous direction vector)

$$
\begin{aligned}
\|x - x_j\|_A^2 &= \|x - x_{j+1} + x_{j+1} - x_j\|_A^2 \\
&= \|x - x_{j+1}\|_A^2 + \|x_{j+1} - x_j\|_A^2 \\
&\quad + 2(x - x_{j+1})A(x_{j+1} - x_j) \\
&= \|x - x_{j+1}\|_A^2 + \gamma_j \|r_j\|^2 \\
&\quad + 2r_{j+1}^T(x_{j+1} - x_j) \\
&= \|x - x_{j+1}\|_A^2 + \gamma_j \|r_j\|^2
\end{aligned}
$$

Here we used

$$
\begin{aligned}
b - Ax_{j+1} &= r_{j+1}\,, \\
r_{j+1}^T(x_{j+1} - x_j) &= 0\,.
\end{aligned}
$$

4. Numerical experiments

$A = Q \Lambda Q^T, n = 48$, where

$$\begin{aligned} Q &= \texttt{randn(n)}, \\ \Lambda &= \mathrm{diag}(\lambda_1, \ldots, \lambda_n) \end{aligned}$$

and Λ is used from S (1991),

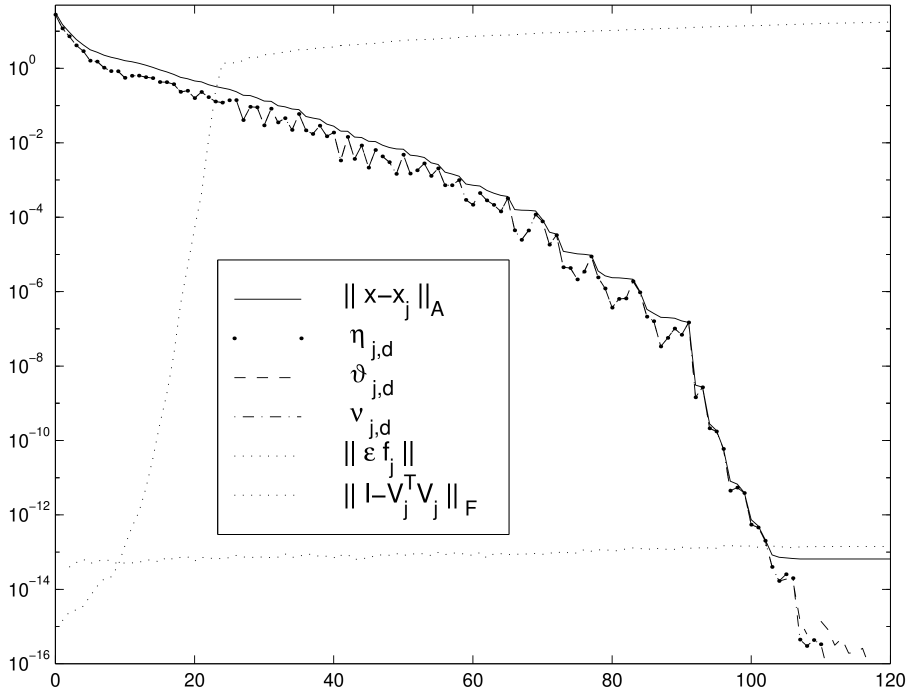
$\lambda_1 = 0.1$, $\lambda_n = 1000$, $\rho = 0.9$

$$\lambda_i = \lambda_1 + \frac{i-1}{n-1}(\lambda_n - \lambda_1)\, \rho^{n-i}\,.$$

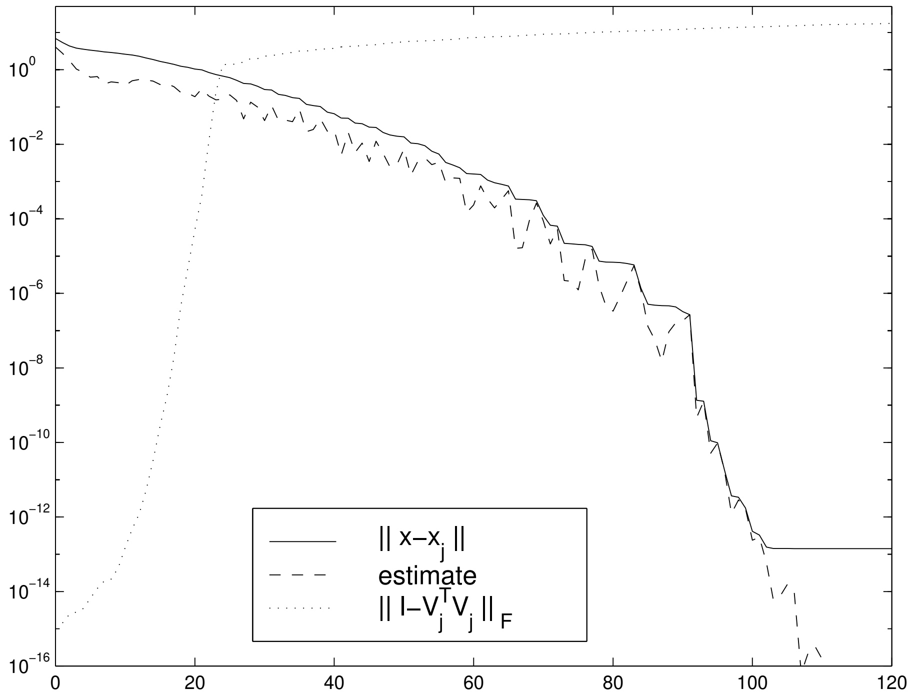
$x = (1, \ldots, 1)^T$, $b = Ax$, $x_0 = 0$.

Notation: $\|\varepsilon f_j\| \equiv \|(b - Ax_j) - r_j\|$.

$\|x - x_j\|_A,\ \|x - x_j\|,\ d = 1$

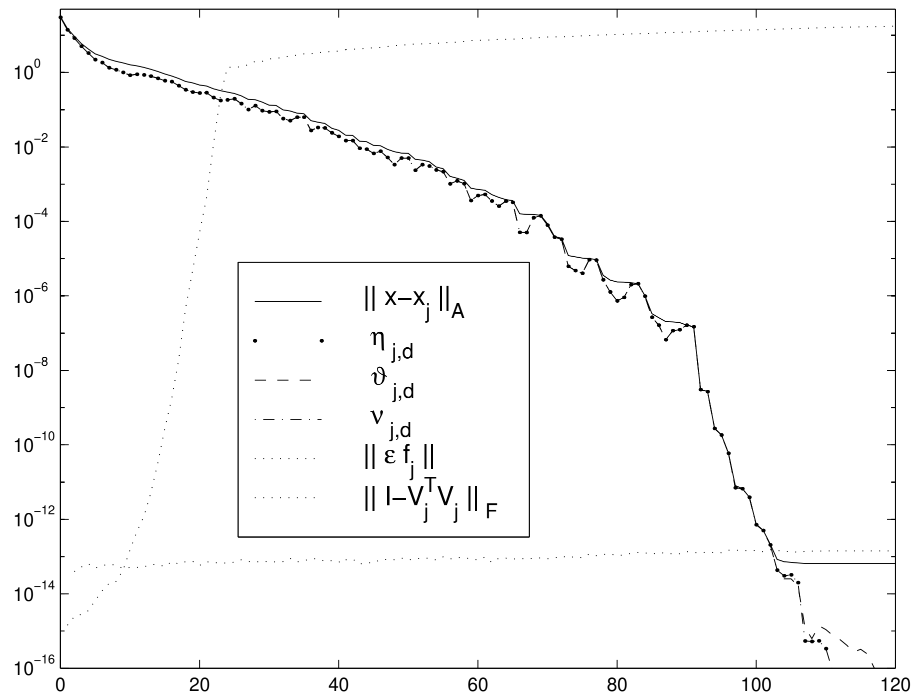


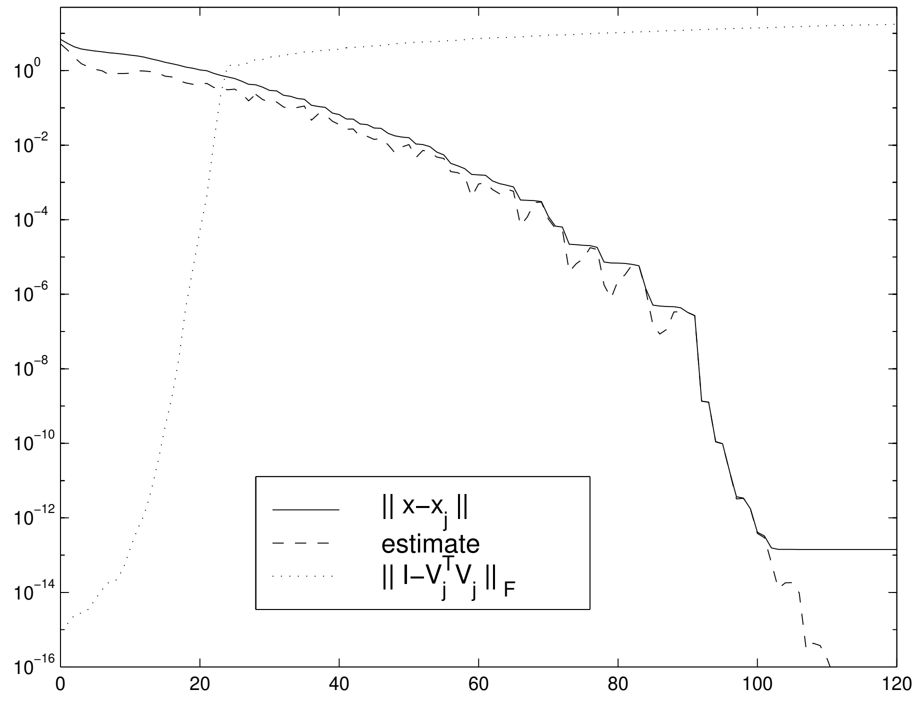
|| x−x_j ||_A
η j,d
ϑ j,d
ν j,d
|| ε f_j ||
|| I−V_j^T V_j ||_F
10^0
10^−2
10^−4
10^−6
10^−8
10^−10
10^−12
10^−14
10^−16
0
20
40
60
80
100
120



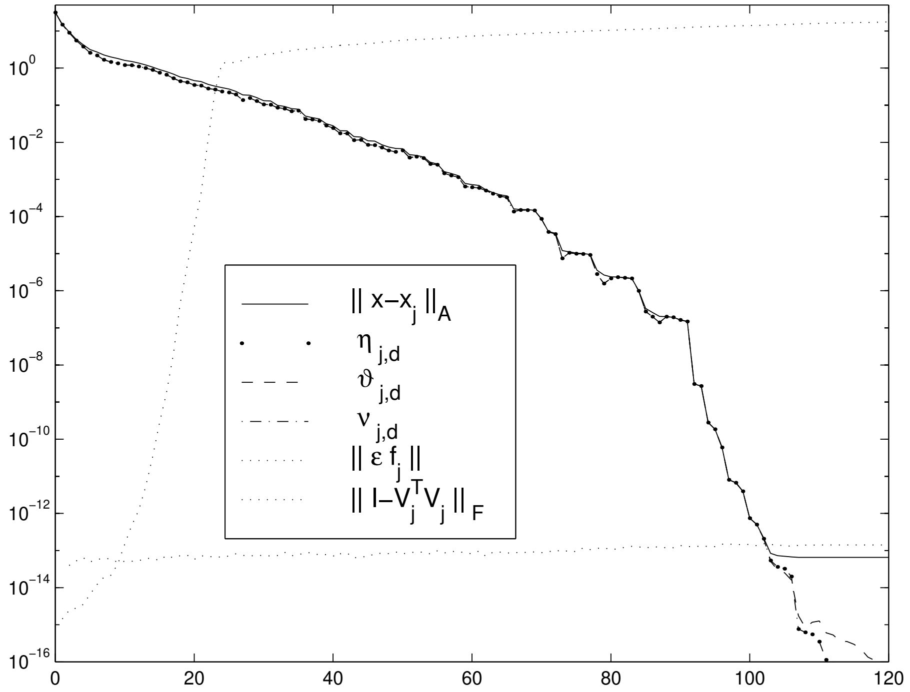
|| x−x_j ||
estimate
|| I−V_j^T V_j ||_F
10^0
10^−2
10^−4
10^−6
10^−8
10^−10
10^−12
10^−14
10^−16
0
20
40
60
80
100
120

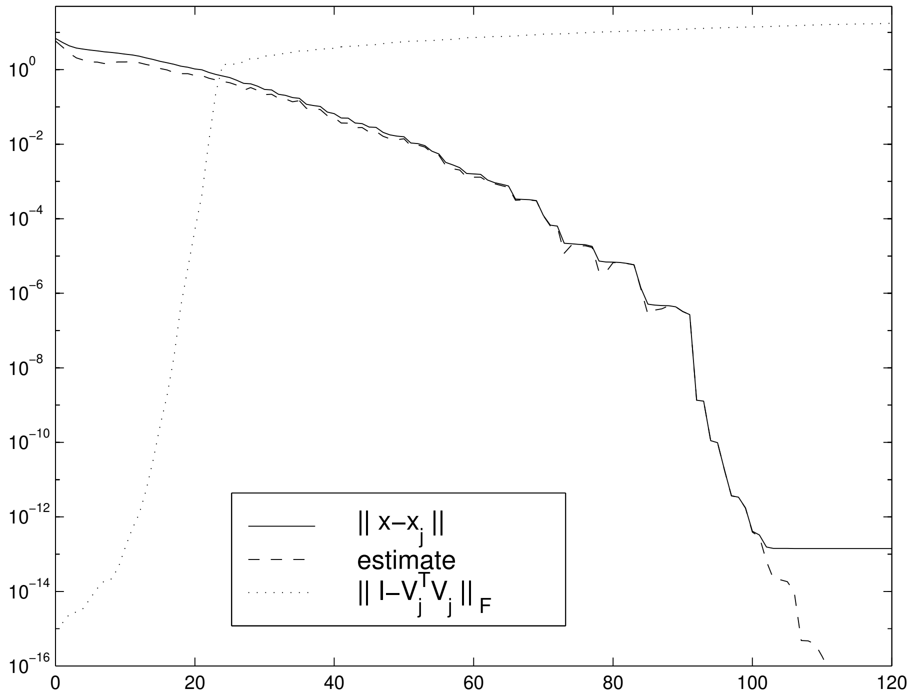
$\|x - x_j\|_A,\ \|x - x_j\|,\ d = 2$



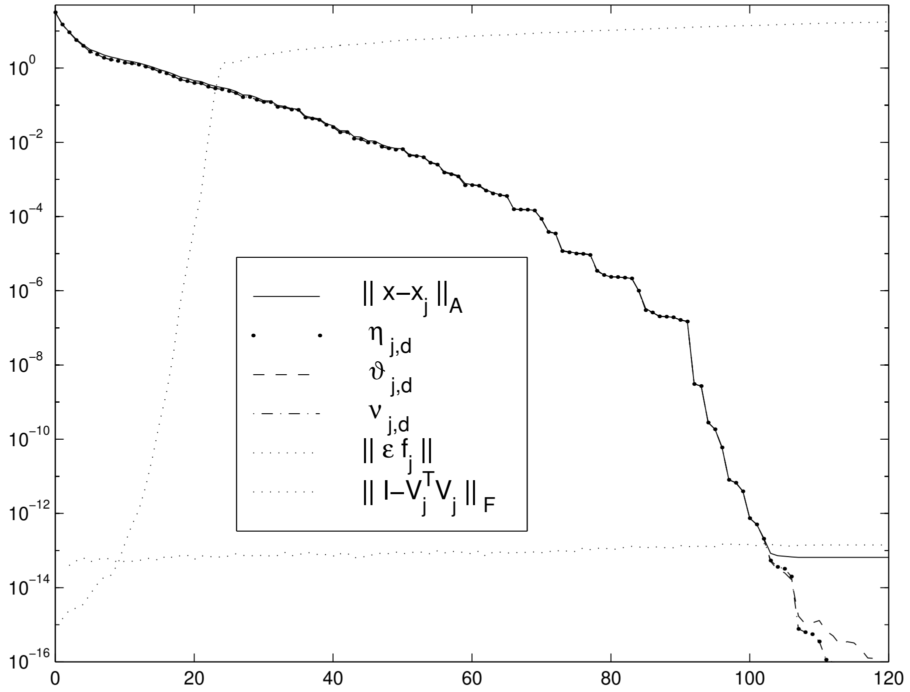


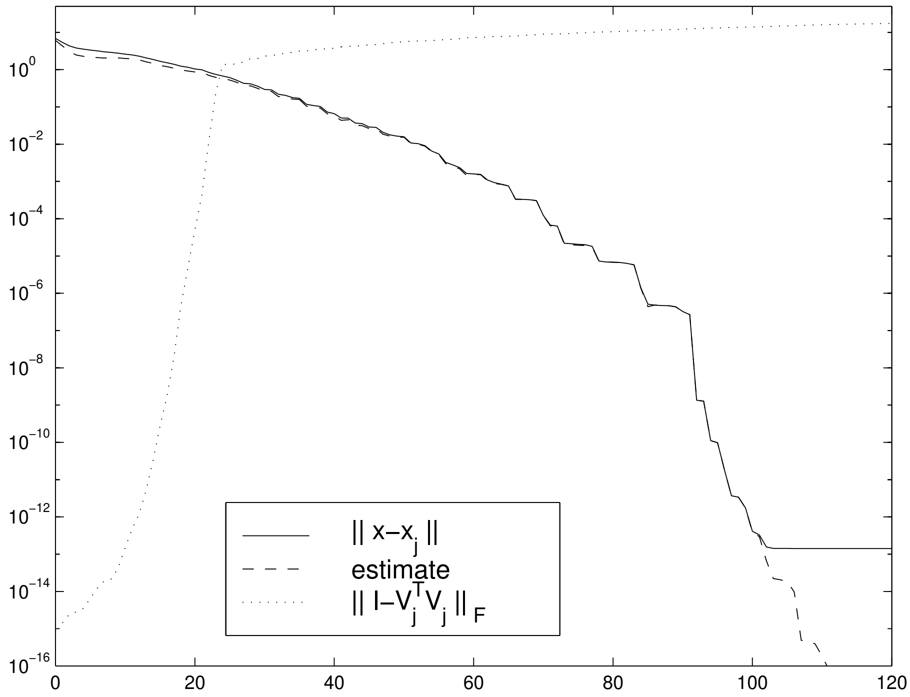
$$\|x - x_j\|_A,\ \|x - x_j\|,\ d = 4$$





$$\|x - x_j\|_A,\ \|x - x_j\|,\ d = 6$$





Estimates of $\|x - x_j\|$:

$$\|x - x_j\|^2 = \sum_{l=j}^{j+d-1} \frac{\|p_l\|^2}{(p_l, Ap_l)}(f(x_l) + f(x_{l+1})) + \|x - x_{j+d}\|^2$$

Considering $\|x - x_{j+d}\|^2 \ll \|x - x_j\|^2$ and replacing $f(x_l)$ by the estimate $\nu_{l,d}^2$ gives

$$\boxed{\|x - x_j\|^2 \geq \sum_{l=j}^{j+d-1} \frac{\|p_l\|^2}{(p_l, Ap_l)}(\nu_{l,d}^2 + \nu_{l+1,d}^2)}$$

Note: $\nu_{j+d,d}$ is needed. Various modifications possible.

Conclusions

- Existing bounds and the original HS (1952) paper.

- Lower bounds for $\|x - x_j\|_A$, $\|x - x_j\|$.

- Applications of the bounds to FP CG computations *must be justified* by a proper rounding error analysis.

- Upper bounds in FP arithmetic?

- Adaptive choice of the parameter d?

- HS (1952) should be celebrated,
  but also studied, even after 50 years!